DOCUMENT RESUME

ED 316 494 SO 020 695

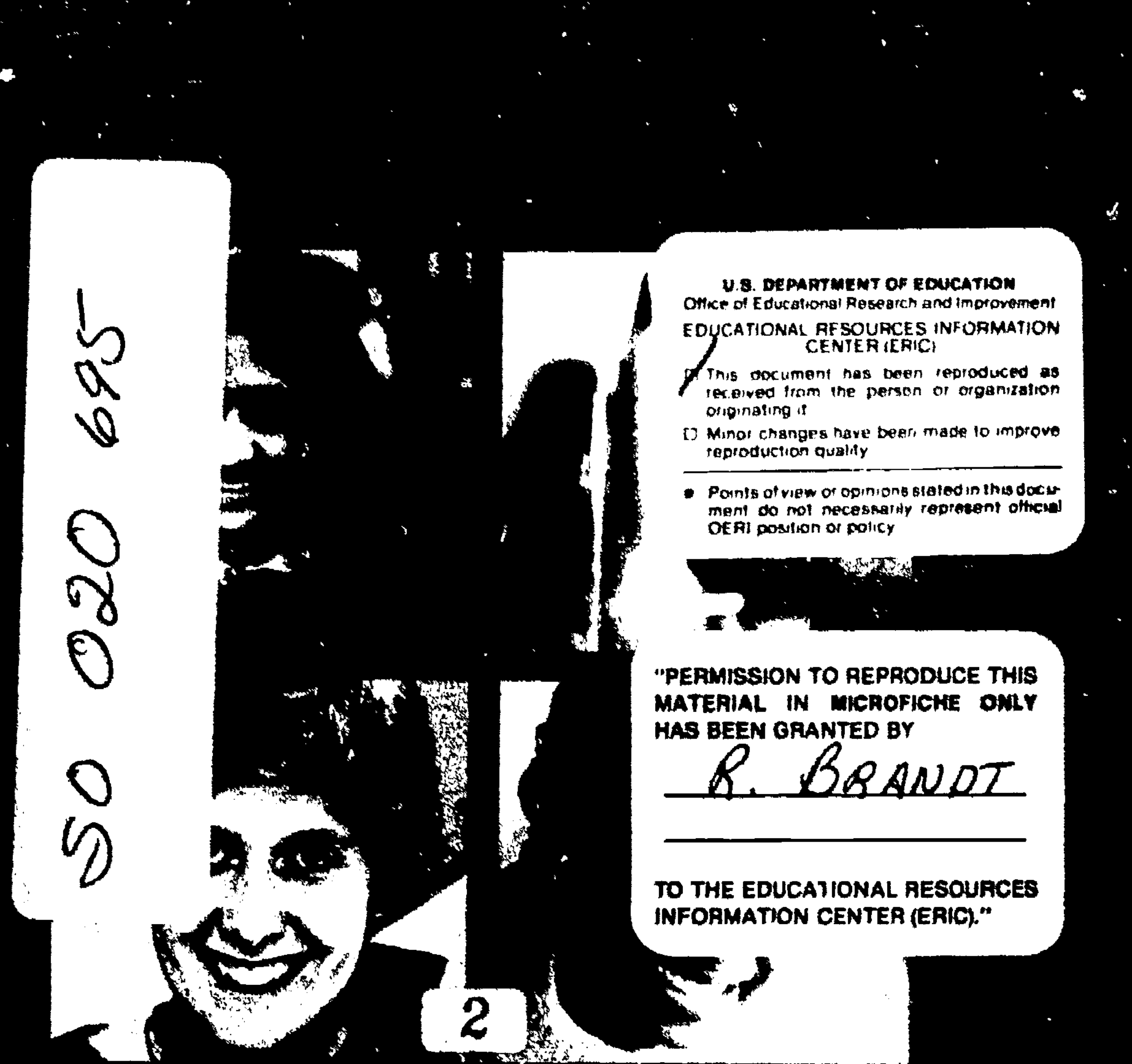
AUTHOR Purves, Alan C., Ed.
TITLE International Comparisons and Educational Reform.
INSTITUTION Association for Supervision and Curriculum Development, Alexandria, Va.
REPORT NO ISBN-0-87120-157-7
PUB DATE 89
NOTE 127p.
AVAILABLE FROM Association for Supervision and Curriculum Development, Alexandria, VA (Stock No. 611-89106, $14.95).
PUB TYPE Books (010) -- Collected Works - General (020)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Case Studies; *Comparative Analysis; Curriculum Development; *Educational Change; *International Studies; Language Arts; Mathematics; Policy Formation; Science Education; Teacher Characteristics; Teaching Methods
IDENTIFIERS International Assn Evaluation Educ Achievement

ABSTRACT

Increasingly over the past decade there have appeared in the press bar graphs comparing the performance of U.S. students with their counterparts in other countries. Usually these graphs have been accompanied by dire warnings and pointing fingers. The main source of the data on international comparisons has been the International Association for the Evaluation of Education (IEA). The articles that make up this document focus primarily on the studies conducted by the IEA. Seth Spaulding sets the stage by looking at the history and policy implications of comparative studies in education, including those of IEA. Ian Westbury shows how complex the interpretation of comparative achievement scores can be. Inger Marklund from Sweden and Zoltan Bathory from Hungary show how two systems of education have used the study results to foster curricular changes in the schools of their respective countries. Both show how the public and planners can use comparative data in a thoughtful fashion. Edward Kifer, Lorin Anderson and Neville Postlethwaite, and Alan Purves examine the various studies to see what broad implications can be drawn concerning issues related to curriculum and instruction. Herbert Walberg examines the relationship between student performance measures and other indicators of a nation's well-being in the world. Chester Finn concludes the volume with a look at comparative performance from a policymaker's point of view. (JB)

International Comparisons and



50

International Comparisons and Educational Reform

Edited by Alan C. Purves

Association for Supervision and Curriculum Development

Printed in the United States of America

Typeset by Mid-Atlantic Photo Composition, Inc.

Printed by R.R. Donnelley & Sons Company

Ronald S. Brandt, Executive Editor
Nancy Modrak, Managing Editor, ASCD Books
René Townsley, Associate Editor
Pearl G. Aldrich, Editorial Assistance

ASCD Stock No. 611-89106
ISBN 0-87120-157-7

$14.95

International comparison and educational reform / edited by Alan C. Purves.
p. cm.
Contents: Foreword / Patricia C. Conran—Introduction / Alan C. Purves—Comparing educational phenomena / Seth Spaulding—The problems of comparing curriculums across educational systems / Ian Westbury—How two educational systems learned from comparative studies / Inger Marklund—The Hungarian experience / Zoltán Báthory—What IEA studies say about curriculum and school organization / Edward Kifer—What IEA studies say about teachers and teaching / Lorin W. Anderson and T. Neville Postlethwaite—Science, mathematics, and national welfare / Herbert J. Walberg—The IEA studies and reforms in reading, writing, and literacy / Alan C. Purves—Afterword—a world of assessment, a universe of data / Chester E. Finn, Jr.
ISBN 0-87120-157-7
1. Comparative education. I. Purves, Alan C. 1931– II. Association for Supervision and Curriculum Development.
LB43.I58 1989 89-439
370.19'5—dc19 CIP

International Comparisons and Educational Reform

Foreword

The timeliness of this book is evidenced by the continuing and growing interest of educators in international education. Whether that interest is motivated by a desire to better the common good of all people, to improve the competitiveness of our economic system, to ensure the survival of our democratic government, or other reasons, it is to bring meaning to international education and global connectedness.

ASCD'S long-range plan carries a mandate to develop international education activities. A task force of ASCD members is giving direction to those activities, and in daily practice we are striving to improve students' understanding of the world in which they live. To model the urgency of the need for international education, the ASCD Executive Council has met in countries outside the United States, and the Association has increased its offerings of professional development activities in various countries. Each year, ASCD welcomes mentors and new international affiliates.

Thus, publications such as this complement the person-to-person interaction of ASCD members around the world with others interested in international education. This book moves us toward the goal of internationalization of ASCD's programs, services, and governance. To the extent that it enhances the meaning of international education, it contributes to the worthiest goal: the betterment of mankind.

Patricia Conran
ASCD President, 1989–90

Introduction

Increasingly over the past decade there have appeared in the press bar graphs comparing the performance of United States students with their counterparts in other countries. There have been other charts showing declines or improvements by students in comparison with their elders. Usually these graphs have been accompanied by dire warnings and pointing fingers. The targets have variously been schools, teachers, television, drugs, race, and parents. As far as the international comparisons are concerned, the main source of the data has been one organization, The International Association for the Evaluation of Educational Achievement (IEA), of which I am the current Chairman. A year ago I was asked by ASCD to prepare a volume that would help elucidate the various subtle issues behind such a glaring display as a bar graph that shows the students of a country like the United States performing below most of the other industrialized countries. Such is the purpose of the present volume, which focuses primarily on the IEA studies, which are the oldest and most comprehensive of the comparative studies, and which have been imitated by private organizations like the Dallas newspapers in the early 1980s and more recently by the National Assessment of Educational Progress. These studies have been complemented by more detailed studies such as those of Harold Stevenson and the various United States–Japan studies. The IEA studies serve, however, as a good focal point for discussion.

In the late 1950s, IEA started as an organization of researchers from around the world who found that they were concerned with a number of issues that could not be studied well within the confines of one school system. The reason for this is that most systems are more or less uniform with respect to such matters as class size, age of school starting, length of the school year, comprehensive secondary schooling, and the like. If one wanted to study these effects, one needed to design costly and politically risky experiments. There was, however, "natural variation" if one were to go beyond the borders of a single geographic unit. The idea of comparative studies of achievement and its antecedents and consequents was thus born.

One of the first problems the group faced was that of creating comparable measures, tests that could be used across languages and cultures. The initial experiments showed that this task was feasible, so a full-scale study of mathematics was launched in the early 1960s and was followed by the "six-subject survey" (reading, literature, science, civic education, and English and French as foreign languages) in the 1970s. During the course of the past decade, IEA has conducted a second study in mathematics and science, a study of written composition, of classroom environments, and is currently conducting studies of preprimary education and computers in education. It is launching a study of reading literacy, the first in a series of projected cyclic studies of learning in the basic school subjects, and is contemplating a study of social values and moral reasoning.

In general, IEA's methodology is one of survey research, with an emphasis on careful test construction combined with sets of questionnaires for students, teachers, and school heads, as well as "national" curriculum questionnaires and supplementary histories and interviews. Over the course of its history, IEA has used various approaches to the analysis of data and has been among the pioneers of various sorts of causal modeling and analysis.

In some systems of education, IEA has come to be seen much as people in the United States see the "wall chart" pitting state against state on the basis of SAT scores. Like SAT, IEA is a household word, both feared and respected and even cherished, depending upon the educational stripe of the viewer. It has been seen as creating an educational Olympics. At the same time, those involved in the work of IEA have always been chary of the simple comparisons; they have downplayed the simple comparisons of mean test scores for a number of reasons which are the substance of this volume. In summary, school systems are often structurally different and any comparison must be seen first in light of such structural differences as percentage of the age cohort in school, amount of prior schooling, tracking and streaming practices; second, in terms of the subtle curricular differences that IEA has labeled "opportunity to learn"; and third, in terms of real cultural differences in conceptions of the given subject matter and of schooling in general. Each of these issues is addressed in this volume.

Seth Spaulding sets the stage by looking at the history and policy implications of comparative studies in education, including those of IEA. He elucidates the various theoretical and practical issues in comparative studies and traces the sorts of interpretive problems that IEA and other comparative studies have faced in comparing the incomparable.

Turning to specific issues of what to make of a comparison, Ian Westbury shows how complex may be the interpretation of comparative achievement scores. Using the recent mathematics data as an example, he shows that the comparisons raise a number of alternatives in curriculum planning and school organization. Any simple answer may be misleading.

Despite this caveat, the IEA studies have spurred educational reforms: Inger Marklund from Sweden and Zoltan Bathory from Hungary show how two systems of education have used the results to foster curricular changes in the schools of their respective countries. Both show how the public and planners can use comparative data in a thoughtful fashion. Their description of how their systems used the data are potential guides to curriculum planners at the district and state level in the United States.

Edward Kifer, Lorin Anderson and Neville Postlethwaite, and I examine the various studies to see what broad implications might be drawn concerning issues related to curriculum and instruction. Kifer pays particular attention to issues related to the sorting of students as well as to other issues of school organization such as retentivity, selection, and the differentiation of the curriculum. These all have their effect on opportunity to learn and thus on achievement. Anderson and Postlethwaite look at the IEA results in terms of characteristics of teachers and teacher training and at classroom arrangement and teaching practices. They draw some general conclusions about the curriculum that affect the students and their teachers. In tracing the particular implications for a single area of the curriculum, language arts in the mother tongue, I examine the results of the reading, literature, and written composition studies to see what implications might be drawn for reform. The studies in reading and literature had an effect similar to that of the mathematics and science studies; they sent researchers and teacher trainers to a particular country, New Zealand, to seek to determine why the students there appeared to be far better readers than students in other countries. That experience has led to profound shifts in the language arts curriculum, particularly at the elementary level.

Herbert Walberg follows with a provocative examination of the relationship between student performance measures and other indicators of a nation's well-being in the world. His chapter sets forth the idea that educational performance can be a harbinger of economic well-being and productivity in an age of information and technology. It is this powerful link that has helped spur the sorts of studies that IEA represents. Although the other chapters suggest that gross indicators mask many important nuances, Walberg reminds us that gross indicators are used and will continue to be used by policymakers around a world that is increasingly interconnected and dependent upon a highly educated workforce.

Chester Finn concludes the volume by looking at comparative performance from a policymaker's point of view. As his chapter indicates, during his term in the United States Department of Education, Finn pushed for the development of clear and understandable educational indicators that could be used much like economic indicators to show the changes and relative position of educational systems. Such indicators represent a part of the policmaker's need to report clearly and simply to the public. He places international studies in the context of national and state-by-state comparisons. This paper makes a plea for continued comparative indicator studies, particularly among the industrialized nations. Such studies, including the IEA studies, are an important aid to the educational decision maker, whether professional policymaker or informed citizen.

ALAN C. PURVES

10

Comparing Educational Phenomena: Promises, Prospects, and Problems

SETH SPAULDING

We all learn from comparisons, and the history of comparing and borrowing ideas in education, according to historian/comparative educator Bill Brickman (1988, in an article published just before his recent death), can be traced at least to the fourth century B.C. when, in *Cyropaedia*, Xenophon analyzed "what he described as Persian education, thereby enabling the Greek reader to compare it with Spartan schooling" (p. 3). Indeed, for over two millennia since then, until relatively modern times, comparative studies in education were essentially reports of scholars, government officials, and others who described the educational institutions, systems, and programs they saw when traveling.

With the formalization of comparative education after World War II, scholars began to use more structured approaches borrowed from various social science disciplines, conceptual frameworks and methods that provided quantifiable elements to compare. The questions were, essentially, how to compare educational systems and phenomena so that we know we are looking at something truly comparable and, at the same time, discover reasonable explanations about why things happen as they do. To some scholars, such comparisons were primarily of intellectual interest, part of the drive to know more about other cultures and peoples. Others, particularly in the United Kingdom and Europe,

Seth Spaulding is Professor of International and Development Education, Department of Administrative and Policy Studies, School of Education, University of Pittsburgh.

thought policymakers could learn from the experience of other countries.[1]

Comparisons in education, whatever the motive, take place in a complex environment. Children in different countries attend schools in substantially different societies and educational systems. They attend for varying periods of time, both during the school year and during their cumulative school career. Some systems are oriented toward centralized decision making and national examinations to select those to continue beyond compulsory levels. Others are highly decentralized with little standardized achievement testing as a basis for promoting students. Some countries retain most of their children in school through the secondary level; many still do not provide a full primary school education for every child.

At the same time, the educational problems and issues between and among nations are often very similar, essentially revolving around how to provide education that is democratic, efficient, and relevant. It seems appropriate, therefore, that we try to learn from one another.

Sharing Descriptive Information Among Countries

The United Nations Educational Scientific Organization (UNESCO), established in 1946, early recognized the need for sharing of information about educational systems among countries. However, statistics collected from member states did not to report comparable phenomena. Age groups in one country were slightly different at varying levels or grades than in other countries. Some countries required more years at various educational levels than others. Some reported expenditures in categories not used by others, and, of course, curriculum emphases varied from country to country.

Over the years, UNESCO developed, in cooperation with educational authorities in member states, the International Standard Classification of Education (ISCED) that provides agreed-upon standards and definitions so that we now have a better chance of understanding what statistical data from various countries mean. ISCED is designed for

[1] See Erwin Epstein, "Currents Left and Right: Ideology in Comparative Education" in *New Approaches in Comparative Education*, edited by P.G. Altbach and G. Kelly (Chicago: University of Chicago Press, 1986), for a fascinating discussion of the various ideological and methodological currents in the field

assembling data on current educational phenomena such as enrollment, teaching staff, and finances, as well as for statistics of the "stock" of educated people obtained, for example, by a national census. It classifies "*courses, programmes*, and *fields* of education according to their educational content. The educational content of each course, programme, and field is designated according to its *level category* and its *subject matter*."[2] Further refinements make it possible to intelligently compare statistics on education from various countries, as long as they use ISCED. Fortunately, most do, or at least report data so that they can be regrouped following ISCED standards, and UNESCO's statistical yearbook is an invaluable source of information. These statistics are often used in connection with other studies, such as the rate-of-return, cost-benefit, and manpower studies often done by economists, and the cognitive achievement studies done by groups such as the International Educational Achievement Association (IEA).

The availability of better statistics, of course, has helped to describe what is happening in various countries in terms of enrollment at various educational levels, broad curriculum trends, drop-out and wastage problems, educational finance trends, and so forth, but these data have only illuminated the problems without much information about solving them. UNESCO, accordingly, encourages sharing qualitative information through documentation center work, both at UNESCO headquarters in Paris and at its International Bureau of Education in Geneva, and through a variety of conferences and studies. Perhaps most notable is the International Conference on Education, held every two years in Geneva, which attracts ministerial delegations from over 100 countries. Before the conference, each delegation responds to a questionnaire about its theme, and the results are summarized in conference documents. In addition, most countries submit a report that is microfilmed and made available to researchers and educators. Unfortunately, the Reagan administration withdrew from UNESCO in 1984. Since then, the U.S. has not participated in most UNESCO educational efforts.

[2] UNESCO, *International Standard Classification of Education*, abridged edition (ED/BIE//CONFINTED.35/Ref.8) (Paris, July 1985), 1 and 21. See also "Educational Statistics: National and International Sources and Services," in *Educational Documentation and Information* (Bulletin of the International Bureau of Education, No. 202, 1st Quarter 1977), which describes in detail the UNESCO statistics collection system.

The Unique Role of the IEA

UNESCO is the major intergovernmental organization responsible for sharing educational information worldwide; numerous other governmental organizations are involved as well. The World Bank, the Organization for Economic Cooperation and Development (OECD), the Organi· ·on of American States (OAS), and var·ous bilateral organizations s· as the U.S. Agency for International Development (AID), Swedish International Development Agency (SIDA), Canadian International Development Agency (CIDA), and the U.K. Overseas Development Authority (ODA) continue to support various kinds of educational activities in a number of countries. None of these agencies, however, has attempted to coordinate a study as complex as those undertaken by IEA. Such studies have been possible only because of a uniq·ıe process of collaboration among researchers, private and public institutions, and governments.

IEA and its work have grown from modest beginnings at a meeting of educators, psychologists, and curriculum specialists at the Hamburg UNESCO Institute for Education in the late fifties. They felt the need for more concrete information about what students around the world were, in fact, learning, especially in academic areas. Some felt that such information would be useful in identifying the possible effects of differences in educational systems; others simply had a scholarly interest in finding ways to measure and compare cognitive achievement among countries of different cultures and languages. It became, as the eminent American comparative educator George Bereday said to then IEA chairman Torsten Husén a few years later, an attempt to compare the incomparable.

Although viewed by some as a kind of cognitive Olympics, the IEA national achievement rankings have often been used by critics of educational systems and by those who urge educational reform. Clearly, more and more countries and educators in those countries are intrigued by the IEA's work, as the number of participating countries rose in recent years from 20 to some 50. Both the Soviet Union and China possibly will participate in the future.

But the IEA studies are more than a cognitive Olympics. In a very real sense, they represent the frontier of comparative research in education. The studies attempt to answer questions raised by policy groups both nationally and internationally about what learning educational systems and institutions produce and at what cost. People worldwide are searching for indicators of educational effectiveness and efficiency

A number of scholars see the IEA studies as an important complement to one-country studies such as the U.S. National Assessment of Educational Progress (NAEP).[3] As NAEP has moved in 20 years from a concern for achievement levels to a concern for variables that may be related to achievement,[4] so IEA has moved vigorously into the analysis of these variables or indicators. Unfortunately, methodological differences exist between the two approaches. Among others, NAEP examines students on test items selected by a panel outside the educational system, while IEA uses curriculum-based tests.

Indicators, in educational policy research circles, are essentially statistically demonstrated relationships between cognitive achievement and content, structure, methods, teacher characteristics, and/or organizational processes within the school and/or societal factors outside the school system. The problem, of course, is whether such relationships suggest cause and effect and, if they do, whether we can prescribe improvements based on these relationships.

Methodological Issues[5]

IEA so far has used essentially surveys to gather data, and all the problems of such a methodology are multiplied in a multination effort. The Association is largely dependent on the good will of participating institutions and researchers, and methodological issues have been resolved within each project by discussion and agreement among the researchers involved. Some of these issues are truly formidable, especi-

[3] See, for example, Richard M. Wolf, "The NAEP and International Comparisons," *Phi Delta Kappan* 69 (April 1988): 580-581.

[4] See, for example, Daniel P. and Lauren B. Resnick, "Understanding Achievement and Acting to Produce It: Some Recommendations for the NAEP," *Phi Delta Kappan* 69 (April 1988): 576-579.

[5] The following researchers and educators affiliated with IEA projects graciously responded to my request for suggestions and comments on the issues I wished to explore. Only lack of space prevents a more complete inclusion of their richly detailed responses. They are: Ricardo Charters d'Azevedo (Portugal), R.A. Garden (New Zealand), Jack Holbrook (Hong Kong), Torsten Husén (Sweden), Inger Marklund (Sweden), Panom Pongpaibool (Thailand), Neville Postlethwaite (Germany), Alan Purves (USA); Vivien M. Talisayon (Philippines), Aneli Vahapassi (Finland), Michael Wilson (Australia), and Richard Wolf (USA).

ally when attempting to decide what is relevant for testing in each country and at each grade level.

The IEA participants and informed observers alike engage in lively and spirited dialogue about both the process and the methodologies used. Keeves, for instance, notes that the early studies lacked a clear theoretical framework, and at the same time, wonders if, in certain kinds of projects, the development and use of a coherent theory are impossible. He also notes the time and resource constraints for getting research participants together and for developing common procedures and statistical techniques. He comments that the working group involved in setting the aims and objectives of the Second IEA Mathematics Study was dominated by research workers and curriculum specialists from the United States, "who apparently cannot conceive of very different approaches to the teaching of mathematics than are employed in their country" (Keeves 1980, p. 8). He also notes that the earlier studies involved a global strategy for examining the data and did not, in general, seek to explore specific relationships and to test specific hypotheses (1980, p. 19).

Marklund (1983), in turn, suggests that curriculum and research methods are constantly changing in many countries and questions the longevity of research results. He says that when searching for indicators which may have some relationship with school achievement, disaggregating the data is important so that one can see what happens in individual schools and within various population groups. Postlethwaite, involved in IEA from its beginning, is fully aware of the complexities of comparative study. He notes that IEA methodology has matured and now most projects not only collect achievement scores, but also construct and apply attitude scales and background information questionnaires. "The background questions obtain information from students, teachers and school principals—indicators which can be used to 'explain' differences in achievement outcomes."[6]

Although IEA projects began as straightforward comparative achievement studies, they soon evolved into "effective schools

[6] T. Neville Postlethwaite, "Cross National Convergence of Concepts and Measurement of Educational Achievement" (paper delivered at the annual meeting of the American Educational Research Association, New Orleans, April 1988). The current sophistication of the IEA studies is illustrated in the recently issued preliminary IEA report, written by Postlethwaite, on *Science Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988).

research." "Effective schools research" usually focuses on what happens in individual schools, and such studies to date concentrated on elementary schools in urban areas. IEA, on the other hand, deals with both elementary and secondary levels in both urban and rural settings and has reported aggregate data by country. However, current IEA and IEA-related studies are increasingly disaggregating school-level data.

Heyneman of the World Bank (an organization that stresses quantification of indicators so it can assess the impact of its investment in educational development) suggests that the availability of computers beginning in the late '60s made possible the identification, measurement, control, and analysis of "hundreds of influences on school achievement simultaneously" (1986, p. 3). The Coleman report was an earlier attempt to do this on a national scale in the U.S., and its results disturbed educators. They felt that the effects of schooling were minimized by Coleman's methodology. Bowles and Gintis point out, for instance, the dangers in constructing a model in which personal influences were forced into a regression before other variables and suggest that too many variables were colinear and insufficiently accounted for (cited in Heyneman 1986, p. 4). Heyneman cites several studies which suggest that, in many less-developed countries, little relationship exists between socioeconomic status and educational attainment, although studies in advanced countries show that they do make a difference. Clearly, methodological issues are not yet entirely resolved.

Further Complexities of Interpretation

Effective schools research attempts to establish the functional relationships "that exist within and among the four stages of educational production—inputs, process, outputs and outcomes" (Windham 1988, p. 149). Figure 1.1 lists some of the factors in what economists call the "education production process." In such a model, elaborate attempts are made to determine the relationships among the various factors. In addition, the usual assumption is that the various inputs and the ways they are put together (process) are *determinants* of the outputs and outcomes (the effects). Such models do not deal with qualitative issues involving the relevance of the *outputs* (cognitive achievement, skills, attitudes, behaviors) to such goals as employment, earnings, and social status. Thus students might achieve academically, but be complete misfits in terms of many societal and/or individual needs. Essentially, models such as Figure 1.1 largely ignore the political, economic, and cultural

Figure 1.1
Major Factors in the Education Production Process

DETERMINANTS

Inputs	*Process*
Student characteristics	Forms of instructional organization
Teacher characteristics	Alternative technologies
School characteristics	Use of teacher and student time
Instructional materials and equipment characteristics	

EFFECTS

Outputs	*Outcomes*
Cognitive achievment	Employment
Improved manual skills	Earnings
Attitudinal changes	Status
Behavioral changes	Attitudinal changes
	Behavioral changes

Source: Douglas M. Windham, "Improving the Efficiency of Educational Systems: Indicators of Educational Effectiveness and Efficiency" (paper presented for IEES Educational Efficiency Clearinghouse, Learning Systems Institute, Tallahassee, Florida, January 1988), 9

systems that contain the educational settings and may provide more powerful explanations for what is happening i[illegible] the schools than do the factors in the model.

Marlaine E. Lockheed of the World Bank notes that such a model "must be able to separate the influences of the different inputs to the educational process in order to judge their effectiveness," the costs of each must be determined, and all of this must be interpreted in terms of the way an educational system uses inputs in various settings (Lockheed 1988, pp. 12-14). Lockhee[illegible]cognizes the difficulties in undertaking such policy analyses [illegible] notes, for instance, that educated parents in developing countries [illegible]y to provide learning in the home and other resources and also to [illegible]d their children to schools with better resources and better trained teachers. In a personal letter (May 22, 1988), Jack Holbrook points out that socioeconomic level can affect achievement in Hong Kong because students from wealthier families have access to tutoring. Illiterate parents, in turn, are likely to send their children to schools with lesser resources. How can the influence of specific school inputs be separated from each other or from that of parents (Lockheed 1988, p. 12)? Furthermore, the use of inputs is difficult to determine. Ably used textbooks, for instance, may show a sizable effect on achievement, while poorly used (or badly constructed) texts may show no effect.

Several critics of the IEA methodologies take a similar position: that the IEA studies are significant but that the explanations of variance, using the limited information collected, are questionable. "Future efforts must be devoted to contextualizing the data we are generating in a meaningful way," suggest Theisen and his coauthors (1986), who list various "educational production function" studies in poor countries in which factors related to achievement conflict. A critique of Theisen's study, however, suggests that the IEA studies, plus the broad range of spin-off national and international studies, are in fact increasingly examining local phenomena,[7] though even these authors suggest that "there is no statistical approach that is even minimally acceptable in ruling out the many alternative explanations for the differences that are observed" (Schwille and Burstein 1987, p. 602). World Bank specialist Steve Heyneman (1986, p. 1), on the other hand, suggests that many educational phenomena are similar from country to country and culture to culture, and that concerns for differences should not obscure the similarities.

Policy Uses and Abuses of Comparative Studies

The comparative study of educational phenomena, it appears, is now an item on the agenda of chiefs of state. "When President Reagan and Prime Minister Nakasone met in 1983, one result of their discussions was an agreement that a cooperative undertaking by Japan and the United States to study education in each other's country would be worthwhile" (Leestma et al. 1987, p. vii). Each country formed a task force to study the education system of the other, and the U.S. report (a kind of case study) on Japanese education was published in 1987. IEA figures showing that Japanese achievement scores in mathematics are higher than in the U.S. undoubtedly motivated, in part, presidential interest in the studies.

[7] Recent examples include *A Origen Socioeconomica do Aluno e o Sucesso Escolar* (Lisbon, Portugal: Ministerio da Educacao, February 1987) and *Determinants of Effective Schools: Regional Syntheses* (Bangkok, Thailand: Office of the National Education Commission, May 1987).

In an epilogue to the report on Japanese education, the then Secretary of Education William J. Bennett lists what he considers to be implications for American education. He chides American educators who, he feels, "have tended to shun the 'lessons' of Japanese education" (Leestma et al. 1987, p. 69). He lists twelve "principles" that he gleaned from the study and which, he feels, agree with U.S. research findings: the importance of parental engagement, clarity of purpose, motivation, expectations and standards, comprehensiveness and quality, character and values, well-ordered school environments sufficient time on task, sufficient resources, good teachers' salaries, and youngsters who take responsibility for their educational achievement. He takes his usual swipe at Colleges of Education: "It may be noted that, in most cases, the Japanese do not enter the teaching profession via colleges of education, nor is it necessary to do so in order to be knowledgeable about one's field and competent to transmit one's knowledge to young people" (in Leestma et al. 1987, p. 71).

Some of Bennett's generalizations are somewhat disturbing and illustrate the questionable interpretation of comparative data. For instance, his implied notion that we could well eliminate professional education courses when preparing teachers is a clear misreading of the data in the report. Although most Japanese teachers do not attend colleges of education, per se, they are required to have substantial preparation in professional education subjects to obtain a teaching certificate. These subjects are offered by colleges and universities that prepare young people for teaching certificates even though there may not be a formal "college of education" in each institution. The minimum requirements for a preschool or elementary teaching certificate include about twice as many credits in professional education subjects as in teaching subjects, and minimum requirements for a secondary teaching certificate include from about one-quarter to one-half professional education subjects (depending on the level and class of certificate). This is hardly ammunition for Bennett's oft-repeated criticisms of professional education courses (Leestma et al. 1987, p. 16). Furthermore, a complex system of inservice teacher education centers continually update teaching staff in both content and method (Leestma et al. 1987, pp. 17-18).

The rosy picture painted by Secretary Bennett of "lessons" to be learned from Japanese education is further eroded by a mimeographed document originally prepared by the Center for Education Statistics of the Department of Education to accompany the report on Japanese edu-

20

cation, but as of mid-1988, issued only in draft form (Suter n.d.). Much of this report builds on the IEA math study and related studies in Japan and the U.S. It points out, for instance, that although math achievement scores in Japan seem higher than those in the U.S., Japanese students "are not necessarily likely to express a high interest in the subject. Japanese students are less likely than U.S. students to say that mathematics is important or fun . . . [or] . . . that they perceive the utility of learning mathematics for the future . . ." (Suter 1988, p. 11).

In addition, Bennett's lessons from the Japanese study suggest that we should learn from their approach to parental engagement and the way this contributes to motivation. The statistical analysis of Japanese and U.S. survey data, however, suggests that "Japanese students perceive less parental encouragement than do U.S. eighth-grade students" (Suter 1988, p. 11).

Bennett also lauds the Japanese system for providing more time in school and more homework, on the whole, than we do. The results of the Department's statistical study, however, suggest "that the differences in use of class time and amount of hours of homework can ot explain the differences in students' achievement between the two countries" (Suter 1988, p. 29).

Perhaps more significantly, some features of the Japanese education system from which the Secretary gleans his "lessons" are those which the Japanese themselves question. The same report, just before the Secretary's epilogue, notes that an outside review by the 24-nation OECD some years ago lauded Japan's educational accomplishments, but criticized the extent of "centralized control, standardization, conformity, institutional hierarchy, and the emphasis on university entrance examinations" (Leestma et al. 1987, p. 63).

The report also noted that the achievement of Japanese 13-year-olds in the second (1981-2) IEA mathematics study declined since the first study in 1964, and that there is an "increasing number of low achievers in primary and secondary schools; increased school violence, especially at junior high schools; and the emerging phenomenon of voluntary dropouts from senior high schools" (Leestma et al. 1987, p. 63). A reported decrease in public confidence in education led reformers to concentrate on enforced uniformity "that is believed to stifle individuality, create frustrations and contribute to disorder in schools, and on the heavy emphasis on university examinations which is believed to hinder personal and intellectual development" (Leestma et al. 1987, p. 63).

Japan appointed, in 1986, a National Council on Educational Reform that issued two critical reports suggesting that the rigidity of the system created problems such as "bullying, school violence, juvenile delinquency, and the refusal to go to school" and that the "rigid, uniform school programs, excessive controls on students, and other factors prevent sound character formation, increase pressures on children, and create frustration." Furthermore, the Council "fears that an excessive emphasis on memorization has produced many conformist children who are unable to think creatively" (Leestma et al. 1987, p. 65). The report notes that the current reform movement in Japan is moving in the opposite direction from that in the United States. Education reformers in Japan are seeking some decentralization of control, greater diversification of institutions, less uniformity and standardization of curriculum, more flexibility in teaching, and more individualization in instruction (Leestma et al., p. 67).

The sharply chiseled "principles" that Bennett derives from the Department of Education's study of Japanese education leave something to be desired. Indeed, Japanese students in the aggregate seem to do somewhat better in math than U.S. students, but at what price?[8]

The difficulties in applying "principles" from other countries are identified by R.A. Garden, Director of Research and Statistics, Department of Education in New Zealand, and IEA General Assembly representative for his country, who, in a personal letter, suggests that "what appears to be a growth industry for U.S. educational researchers, comparing the USA and Japan is likely to be fruitless [in terms of suggesting policy reforms]. Factors which influence opportunity and motivation to learn are too different in these widely differing cultures."

Ideology, Interpretation, and Borrowing

Former Secretary Bennett's interpretation of the Japanese case study is a clear demonstration of ideologically selective interpretation of data. But selective interest in education and research according to partisan or sectarian views is not unusual. Under the Johnson administration,

[8] Harold Noah documents other such misuses of data in his eminently readable article, "The Uses and Abuses of Comparative Education." *Comparative Education Review* 28 (November 1984): 350-362.

for instance, racial inequities were of prime importance and reflected in the research and other federal programs then encouraged by the administration. Under the Reagan administration, other educational concerns have taken precedence (Schwille and Burstein 1987, p. 605). In an excellent paper on the complexities of using IEA data in policy-making, long-time IEA chairman Torsten Husén (1987) notes that those in England and in Germany both for and against greater comprehensiveness in education used the same Swedish studies to support their positions.

Comparative educators, of course, point out that the initial choice of a research paradigm (the currently fashionable word for strategy or model) is heavily influenced by ideology. Many researchers in comparative education adopt such conceptual frameworks as dependency theory, which assumes center-periphery relationships worldwide, and, more generally, conflict theory, which suggests that tinkering with the education system will do little good in achieving economic and social equity and justice unless and until there is a revolutionary change in power relationships within societies.[9]

Arnove (1982) suggests that borrowing educational approaches around the world simply reinforces various kinds of economic and cultural dependencies. He argues that a world system of education exists within which nations both at the center and on the periphery are dominated by a few of those at the center. Indeed, it is difficult to identify educational structures or innovations that have been borrowed by rich countries from poor countries, but it is easy to catalogue the features of educational systems of rich countries that have been installed (some might say often inappropriately) in poor countries.

In fact, successful innovations we have borrowed from abroad often have little to do with solid research evidence. Many are simply logical approaches to solving technical and professional problems. Some have proven useful to the borrowing country, some have not. One positive example is the Open University in the United Kingdom, whose

[9] For an informative overview of trends in comparative studies in education, see Philip L. Altbach and Gail Kelly in *New Approaches to Comparative Education* (Chicago: University of Chicago Press, 1986); Torsten Husén, "Research Paradigms in Education," *Interchange* 19 (Spring 1988): 2-33; and Carlos E. Olivera, "Comparative Education: Toward a Basic Theory," *Prospects* 18, 2 (1988): 167-184).

approach to distance higher education has been replicated and adapted throughout the world. Another is the Teachers' Center concept pioneered also in the United Kingdom. Other examples include literacy methodologies pioneered in other countries and used successfully in domestic programs.

When we compare and, as a result, borrow, many questions surface. Frequently borrowings are based on intuition or the word of "authorities," however defined. The "new math" of the early '60s is an example. It was developed by mathematics scholars convinced that advanced mathematical theory and thinking should be taught early (before computational skills) to facilitate the learning of higher math in later years. With little research evidence, "new math" was accepted (with minor modifications) by countries throughout the world, including many poor countries, starting with the so-called "Entebbe" math, launched at a meeting in Entebbe, Uganda, in 1963. Now, years after the U.S. has largely abandoned the "new math," many other nations, especially in the developing world, are still struggling with it. Many teachers and students do not fully understand it, and it has not given students the math skills they need.

The Future

Clearly, the study of both the internal and external efficiency, effectiveness, and relevance of education will occupy the attention of policymakers and educational researchers in years to come. In the United States, NAEP attempts to provide increasingly complex data to help assess efficiency and effectiveness, and IEA attempts to do a similar job internationally. Researchers are calling for greater attention to input variables that can make a difference and that we can do something about. Other researchers and educators suggest more qualitative studies and more cautious interpretation of quantitative survey information. Some suggest greater cooperation between NAEP and IEA, but note the differences in approach. Especially difficult to reconcile is the difference in the way test items are selected. The NAEP instruments are constructed by panels of experts outside the schools system, while the IEA tests are curriculum-based and designed to measure only what is taught within a system.

The danger, of course, is that policymakers will expect prescriptions from such studies and may use them to justify innovations or reforms completely out of local context. Already calls are heard for more homework in U.S. schools since other countries that assign more homework appear to do better on IEA tests But would students do more homework in U.S. schools, even if it were assigned? Are the other objectives of the school and the many extracurricular activities in conflict with more homework? And even if all students were to do more homework, would it really make a difference?

A report just published of a seven-year study of homework in U.S. schools and its apparent effects suggests that rigid homework requirements may have little to do with increased learning. Homework must be related to the needs of each student, and slower students often get so far behind by junior high school that homework assignments are no longer relevant to them, so they simply stop doing them. More homework, this study suggests, could further discourage slow students to the point of encouraging additional school dropouts (Epstein 1988).

Another notion arose from international comparisons that show that some countries using national standardized achievement tests for selection purposes (especially for university admission) tend to do better in IEA studies than countries using a more decentralized selection process. If these tests encourage achievement, the reform that would make the most difference at the least cost would be to use centralized national testing as a basis for selection to upper levels of education. In a sense, the U.S. is moving in this direction, with a national curriculum to a large extent defined by textbook publishers and with an increasing emphasis on standardized testing to rate teachers and schools and to assess the ability of students to move up the educational ladder (Heyneman and Fagerlind 1988).

The IEA studies are at the center of an evolving worldwide data base of tantalizing "indicators" of quality, effectiveness, and efficiency. IEA and other groups will increasingly use ethnographic, case study, longitudinal study, and other methodologies to complement the survey research done to date. Policymakers, in turn, will select those "indicators" they see as relevant when they propose educational reforms. Where all of this will lead, nobody knows, but currently fashionable technocratic rationality creates an insatiable demand for hard data to use in making policy decisions. Let us hope we have the wisdom to interpret the data appropriately in our search for better schools and better futures for our children.

References

Arnove, Robert F. (1982). "Comparative Education and World Systems Analysis." In *Comparative Education*, edited by P.G. Altbach and others. New York: Macmillan.

Brickman, W.W. (1988). "History of Comparative Education." In *The Encyclopedia of Comparative Education and National Systems of Education*, edited by T.N. Postlethwaite. New York: Pergamon Press.

Epstein, Joyce. (1988). *Homework Practices, Achievements and Behaviors of Elementary School Students*. Baltimore, Md.: Center for Research on Elementary and Middle Schools.

Heyneman, S.P. (1986). *The Search for School Effects in Developing Countries: 1966-1986*, Seminar Paper Series No. 33. Washington, D.C.: The World Bank.

Heyneman, Stephen P., and Ingemar Fagerlind, eds. (1988). *University Examinations and Standardized Testing: Principles, Experience and Policy Options*. Washington, D.C.: The World Bank. (Technical Paper No. 78.)

Husén, Torsten. (February 1987). "Policy Impact of IEA Research." *Comparative Education Review* 31, 1: 31.

Keeves, J.P. (October 1980). *Issues and Problems in International Cooperative Research*. Hawthorn, Victoria, Australia: Australian Council for Educational Research, Radford House, Frederick Street.

Leestma, Robert, and others. (1987). *Japanese Education Today*. Washington, D.C.: U.S. Government Printing Office.

Lockheed, Marlaine E. (April 1988). "The Measurement of Educational Efficiency and Effectiveness." Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Marklund, Sixten. (1983). *The IEA Project: An Unfinished Audit*. Stockholm: Institute for International Studies in Education, University of Stockholm.

Schwille, John, and Leigh Burstein. (1987). "The Necessity of Trade-offs and Coalition Building in Crossnational Research: A Critique of Theisen, Achola, and Boakari." *Comparative Education Review* 31, 4: 602.

Suter, Larry. (1988). "Indicators of School and Student Performance in the United States and Japan." In "Selected Indicators of Education of the United States and Japan." Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Theisen, Gary L., Paul P.W. Achola, and Francis Musa Boakari. (1986). "The Underachievement of Cross-National Studies of Achievement." In *New Approaches to Comparative Education*, edited by Altbach and Kelly. Chicago: University of Chicago Press.

Windham, Douglas M. (January 1988). *Improving the Efficiency of Educational Systems: Indicators of Educational Effectiveness and Efficiency*. Tallahassee, Fla.: IEES Educational Efficiency Clearinghouse, Learning Systems Institute.

The Problems of Comparing Curriculums Across Educational Systems

IAN WESTBURY

Figure 2.1, a summary of some of the findings of the IEA Second International Mathematics Study (SIMS), illustrates the kind of comparison of achievement that played a key role in recent discussions of the quality of U.S. schools.[1] The form of the argument, now firmly standardized, goes like this:

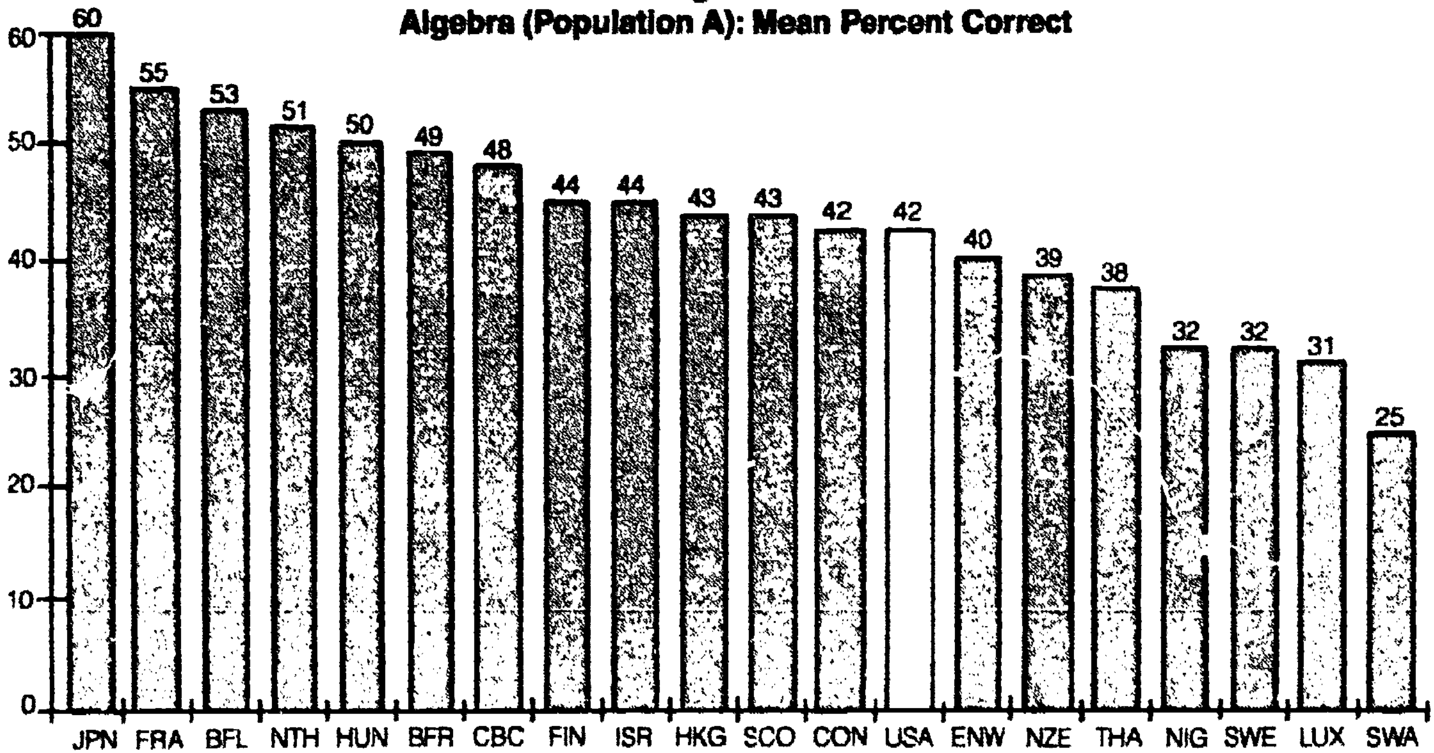
Compared with those in other countries, U.S. students are underachievers If our schools are to compare favorably with the schools of other nations, achievement must be raised. To do this, the argument continues, we must introduce more demanding curriculums, such as those found in more successful school systems This rhetoric is now so conventional that it became part of 1988 presidential election campaigns.

The argument has some merit. A persuasive body of theoretical analysis and empirical evidence suggests that curriculums are causally associated with achievement, and U.S. curriculums are, on their face, less demanding than others in their coverage. But before the implications that seem to follow from such findings can be pushed through to any policy prescriptions, a number of real and very thorny problems in

[1] These findings were included in the U.S. Department of Education's widely distributed report, *What Works?*

Ian Westbury is Professor of Secondary Education, Department of Secondary Education, University of Illinois at Urbana-Champaign.

Figure 2.1
Algebra (Population A): Mean Percent Correct



Key:

JPN	Japan	SCO	Scotland	FRA	Fr[illegible]	CON	Canada (Ontario)
BFL	Belgium (Flemish)	ENW	England & Wales	NTH	The N[illegible]lands	NZE	New Zealand
HUN	Hungary	THA	Thailand	BFR	Belgium (French)	NIG	Nigeria
CBC	Canada (British Columbia)	SWE	Sweden	FIN	Finland	LUX	Luxembourg
ISR	Israel	SWA	Swaziland	HKG	Hong Kong		

comparing curriculums across school systems must be faced.[2] Two of these issues are very important. They are, one, the need to discover what international comparisons of achievement of the kind seen in Figure 2.1 really tell us about what is happening in educational systems. And, two, does this have implications for policy-relevant thinking about the curriculum that could enhance U.S. achievement?

I first review some of the results of the comparative analyses of curriculums in mathematics that were undertaken in the course of SIMS (Travers and Westbury 1988).[3] Any claims about the scope or depth or quality of one school system's curriculum or its achievement vis-à-vis another must be made cautiously and reviewed critically. Paradoxically, however, when cross-system comparisons are made cautiously and carefully, they may indicate greater problems than might seem apparent from a quick pass over the findings.

The Second International Mathematics Study

Figure 2.1 depicts achievement of one of the two cross-national populations investigated in SIMS. Population A students were in the grade at which the modal age was 13. In the U.S., these are students in Grade 8. In addition, SIMS investigated the achievement of students taking an advanced course in mathematics in the terminal grade of each system's secondary system. Population B in the U.S. consisted of students taking courses for which Algebra II is a prerequisite.

In addition to testing students on a common set of cognitive instruments, SIMS collected a large body of data on the mathematics curriculums at the target grade levels in the participating systems. To organize

[2] Within IEA, the terms 'system' or 'educational system' are preferred to 'country' or 'nation.' 'Educational systems' which are not countries or nations, e.g., Canadian provinces of British Columbia or Ontario, participated alongside 'systems' like the U.S. or France. The use of a neutral term like 'system' eliminates the need to constantly qualify usage.

[3] The following systems participated in SIMS: Population A—Belgium (Flemish and French), Canada (British Columbia and Ontario), England and Wales, Finland, France, Hong Kong, Hungary, Israel, Japan, Luxembourg, the Netherlands, New Zealand, Nigeria, Scotland, Swaziland, Sweden, Thailand, and the USA; Population B—Belgium (French and Flemish), Canada (British Columbia and Ontario), England and Wales, Finland, Hong Kong, Hungary, Israel, Japan, New Zealand, Scotland, Sweden, Thailand, and the U.S.

this data and the resulting analysis, SIMS distinguished two curriculums: the *intended curriculum* and the *implemented curriculum*. The intended curriculum was defined as the formal prescribed curriculum contained in national courses of study. The implemented curriculum was the curriculum actually taught in the schools.

Comparing Curriculums Across Countries and School Systems

How can intended and implemented curriculums be described cross-nationally? Within SIMS, the pools of test items developed by the study determined the framework of the description. These menus of possible topics for a mathematics curriculum were developed by consensus of the representatives of the participating systems and contained topics covering a wide-ranging common core of mathematics as seen internationally. Two instruments provided the date. They are as follows:

1. Using each population's item pool as a starting point, national committees of curriculum experts from each system were asked to rate each item as 'appropriate' or 'inappropriate' to their system's *intended* curriculum.
2. Teachers in the sampled schools in each system were asked, via questionnaires, whether the content underlying each item on the test their classes were taking had been taught to their students (the *implemented* curriculum).

Analysis of these data revealed wide differences in some systems between the intended and the implemented curriculums. This has been a common finding in evaluations of many school systems. For example, Keitel reported that the Soviet curriculum held up to Americans as an example of what our principal international antagonist was teaching in mathematics—with all that might mean for both military preparedness and technological competitiveness—was an intended curriculum, not the curriculum actually taught in Soviet schools.

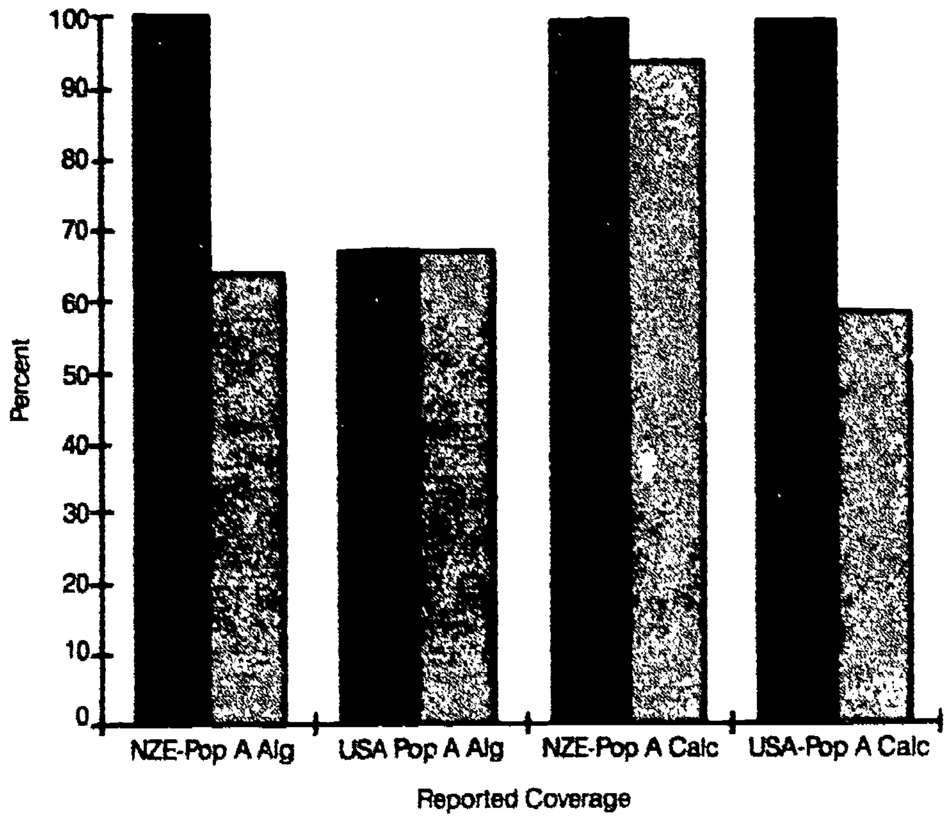
Differences between Intended and Implemented Curriculums

Figure 2.2 compares the U.S. and New Zealand committees' estimates of the proportion of the items in the SIMS content areas of elementary functions and calculus (Population B) and algebra (Population A) they considered 'appropriate' for their systems' students with the reports of New Zealand and U.S. teachers about the same items. This information makes an important point about the difficulties in securing a picture of the curriculum of a system. The U.S. committee provided a much more ambitious picture of what Population B students cover in calculus than did their teachers. In New Zealand, on the other hand, the two pictures are virtually the same. However, for junior high school algebra, the New Zealand national committee offered a very different picture of the curriculum than did classroom teachers. In contrast, the U.S. committee and U.S. classroom teachers presented virtually the same picture for algebra. SIMS evaluated the judgments of both national committees and teachers against what students actually learned as measured by the study's cognitive test and discovered that, *in almost all cases, the descriptions of the implemented curriculum provided by teachers were a better predictor of achievement than the appropriateness ratings provided by national committees.*

How could such committees so misjudge what was happening in their school? And why, as we have seen, were these committees so inconsistent in their judgments (or misjudgments) across content areas and populations? After all, each SIMS national committee consisted of "true" experts. In New Zealand, they were inspectors of mathematics education, whose job was to visit schools and assess curriculums and teaching, and experienced teachers. In the U.S., they were mathematics teachers, supervisors, and textbook authors.

In the U.S. national committee's description of the calculus course, we know the answer to these questions. The committee focused its attention *not* on Grade 12 mathematics classes seen overall (i.e., the curriculum of Population B classes), but on a subset of these classes, the 20 percent of Grade 12 classes where a formal course in calculus is taught. By implication, the committee unconsciously projected its sense of what *should* be taught in Grade 12 to the whole population. They did not make this mistake in the case of Grade 8 algebra. In New Zealand, we know that part of the national committee's problem stemmed from the range in levels of coverage of algebra that occurs in New Zealand's Grade 8 classrooms. Despite the presence of one formally prescribed national curriculum, significant *de facto* tracking occurs in New Zealand schools, which results in highly differentiated curriculums and highly variable achievement. The New Zealand committee has acknowledged that is members did not appreciate the extent of this tracking until they

Figure 2.2
Intended and Implemented Curriculums—USA and New Zealand



saw both teachers' reports of their implemented curriculums and the variation in achievement found in New Zealand's Population A classes.

Many SIMS national committees experienced similar problems in rating the 'appropriateness' of the SIMS item pools to their curriculums. As my examples illustrate, such judgments about the scope of an intended curriculum must always be evaluated in the light of qualification for understanding that might (or might not) be clear to insiders from a particular system, but will be unclear and potentially very confusing to outsiders. And, as I have shown, one of these qualifications is within-system variation in course coverage.

Describing Variations in the Distribution of the Implemented Curriculum

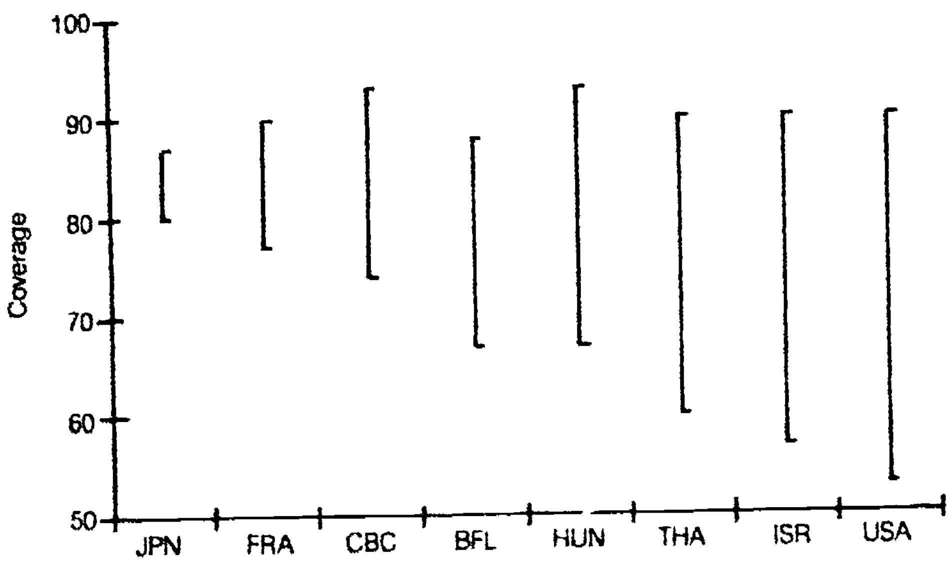
We have seen that the pictures of the implemented curriculum that came from teachers may or may not match descriptions of the intended curriculum based on expert testimony. But in addition to this validating role of teacher-based descriptions, we can also use teacher-provided data to secure a picture of within-system variation in curriculum.

Within-system variation and diversity are key parameters of any national curriculum and, as we have seen, they are parameters which can modulate in significant ways descriptions of a curriculum that focuses on modal patterns alone. Such patterns of distribution of "opportunity to learn" are, in their turn, related to achievement in very important ways.

Figure 2.3 on page 24, for example, describes the variation in coverage of algebra in a subset of systems. In these systems, teachers of high-coverage classes (defined as classes at the third quartile; i.e., Q3, of each system's distribution of teacher-reported "opportunity to learn" in algebra) indicated that their students had covered more than 85 percent of the content in the SIMS algebra item pool. The level of coverage in these classes is represented by the top of each line in the figure. As would be expected, little variation appears across these systems in the algebra taught in their highest-coverage classes. But a quite different pattern emerges when we look at the level in low-coverage classes (defined as classes at the first quartile; i.e., Q1, of each system's distribution of teacher-reported "opportunity to learn"). In Japan, low-coverage classes experience a curriculum that is virtually the same as the curriculum of the highest-coverage classes. Substantial equity in coverage exists across all classrooms in the system. In the U.S., on the other hand, curriculums are extremely diverse, and a substantial difference in the levels of coverage exists for both high- and low-coverage classes. U.S. teachers of high-coverage classes report teaching about 90 percent of the content represented by the SIMS algebra pool, whereas teachers of low-coverage classes report teaching only about 55 percent.

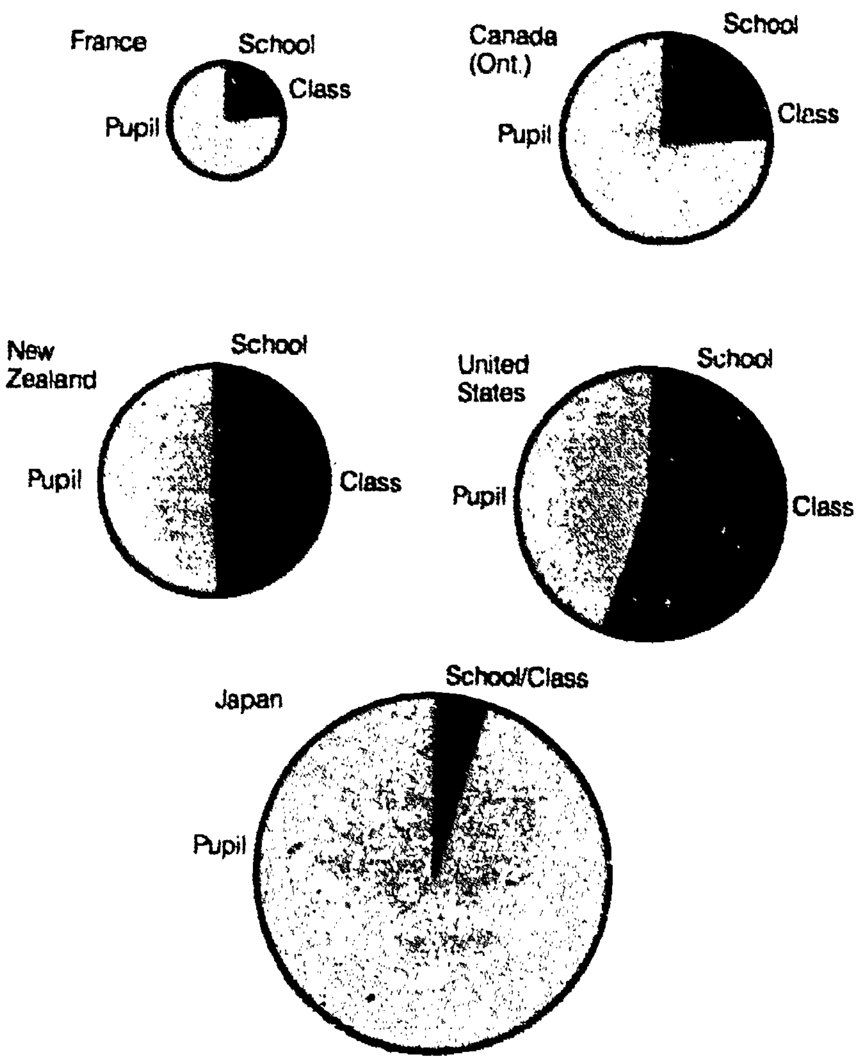
Figure 2.4 on page 25, taken from SIMS U.S. report (McKnight et al. 1987), suggests what this variation in coverage implies for the patterns of achievement of U.S. students when compared with the patterns of some other systems. It is an outcome of the between-class tracking omnipresent in Grade 8 classrooms. But when Figure 2.4 is viewed in the context of Figure 2.3, two important questions arise: Why is the pattern of curriculum offerings in the U.S. Grade 8 so much more diverse and inequitable than the pattern seen in other systems? And more important for our theme, once we recognize this inequity, on which classes or class types should we focus policy prescriptions to improve the curriculum? The curriculum offered to our highest-coverage (and highest achievement) classes is not a core problem when seen cross-culturally, but the curriculum offered to our lower (not lowest) coverage and achievement classes is. This answer does not emerge from a focus on mean or modal, intended or implemented coverage data alone.

Figure 2.3
Algebra: Coverage at Q3 and Q1 for High-Coverage (Q3) Systems



Key: JPN Japan
CBC Canada (British Columbia)
HUN Hungary
ISR Israel
FRA France
BFL Belgium (Flemish)
THA Thailand

Figure 2.4
Population A: Achievement Variance in Five SIMS Systems



These circles depict for each country the variance components associated with the core pretest. (For most countries, this test consists of 40 items. For Japan, it had 60 items.) The areas of the circles are roughly proportional to the total amount of score variation. The slices represent the amount of total variation that is attributable to differences between students, classrooms and schools. Since these are pretest data, the variance components represent how students were allocated to schools and classrooms and not to differences in teaching during the school year. Very similar patterns were found in the posttest data, as well.

Source: Curtis C. McKnight et al. (1987). *The Underachieving Curriculum. Assessing U.S. School Mathematics from an International Perspective.* Champaign, Ill.: Stipes, 1987, p. 108

The Yields of Mathematics Education

Who and What Should We Evaluate?

Within SIMS, the curricular and achievement questions that circle around the ultimate outcomes of mathematics education in each school system were posed in terms of two issues: How much mathematics do Population B students learn, and how many students are given the opportunity to learn that mathematics? We used the term *yield* to gather these questions within one frame. How much do systems vary in the yield of mathematics education?

Let us consider these questions by focusing on one course. How many students in the various systems are given how much opportunity to learn calculus and how much calculus do they learn? Figure 2.5, when viewed beside Figures 2.6 and 2.7 (pp. 28-29), gives us the starting point for considering what we have learned about this from SIMS.

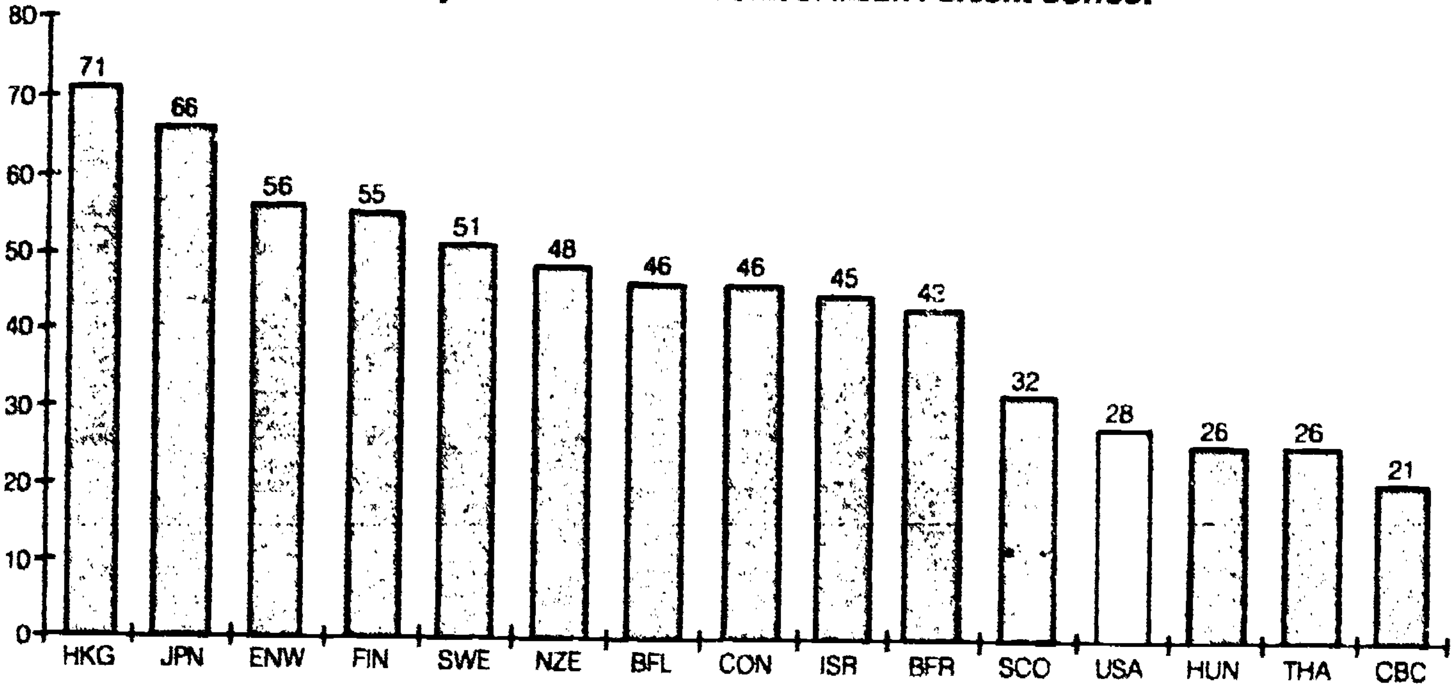
From the U.S. point of view, the findings implicit in the figures in these cross-system comparisons are, at first glance, disconcerting. U.S. achievement is low; the numbers taught in Grade 12 are not particularly high; and, when we view achievement and enrollment rates together, U.S. yield is among the lowest found in the study.

But three issues need to be considered as we try to assess these findings and consider what they might mean for U.S. curricular policy-making. First, the question arises of the meaningfulness and validity of using grade 12 as *the* point of reference for a cross-cultural consideration of yield. Second, who do we count in the pool of potential enrollees in a calculus course and, therefore, what denominator do we use to develop an index of an enrollment rate or retentivity? Finally, what is the significance of calculus itself as a criterion for assessing yield? These questions are difficult to answer.

Who and What Should We Compare?

In Europe and in systems such as New Zealand or Hong Kong, which have highly selective structures at the secondary school-college interface, the curriculum of the academic secondary school is tightly linked to the university and, typically, university programs have been so framed that students cannot pursue studies in most subjects without taking the highest level courses offered in their secondary schools. In mathematics, physical science, and engineering, for most of this century, that course has been calculus. This has not been, and is not, the U.S.

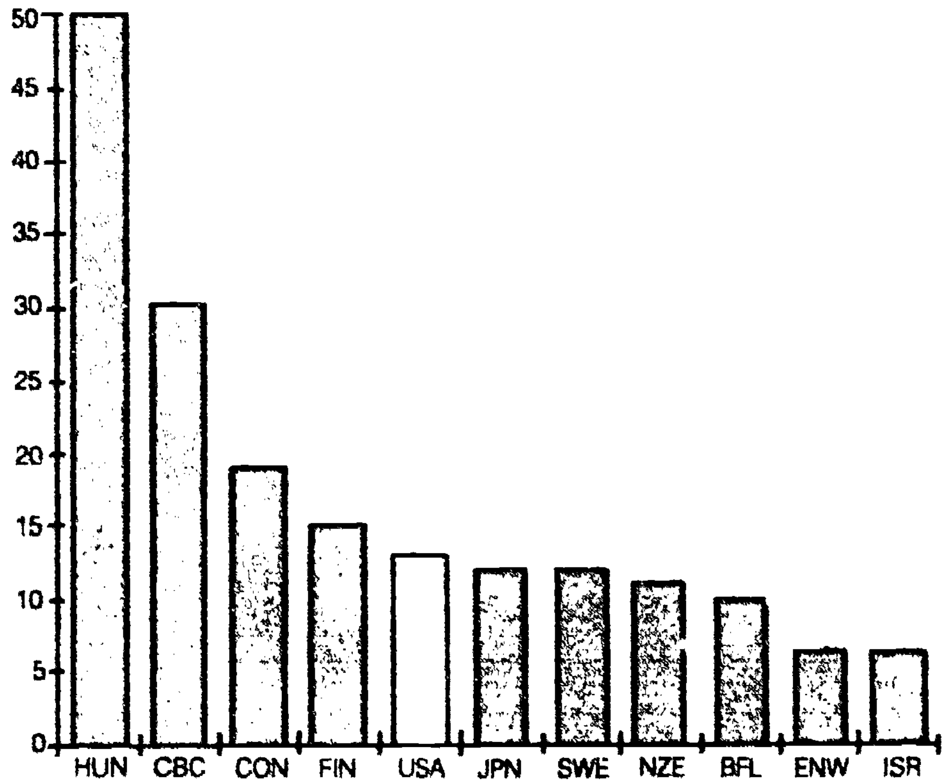
Figure 2.5
Elementary Functions and Calculus: Mean Percent Correct



Key:

HKG	Hong Kong	SCO	Scotland	CON	Canada (Ontario)
ENW	England & Wales	THA	Thailand	BFR	Belgium (French)
SWE	Sweden	JPN	Japan	HUN	Hungary
BFL	Belgium (Flemish)	FIN	Finland	CBC	Canada (British Columbia)
ISR	Israel	NZE	New Zealand		

Figure 2.6
Percent of Age Cohort Retained in Advanced Mathematics

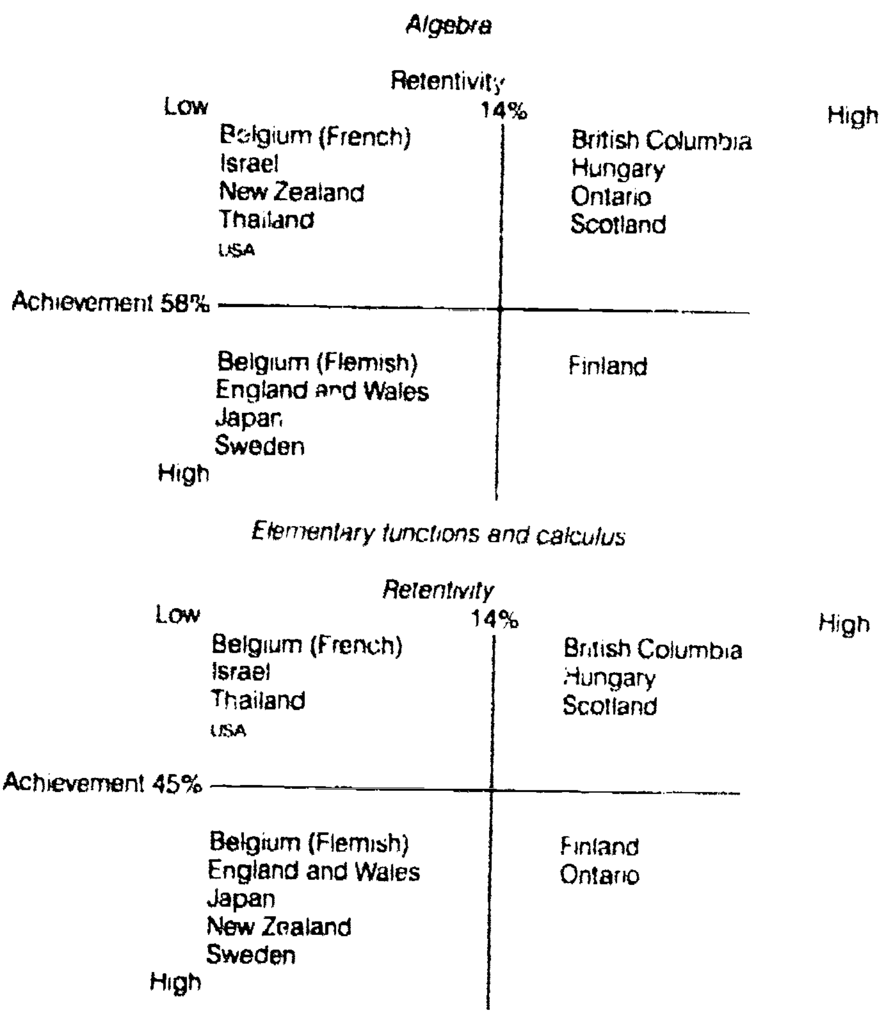


Key: HUN Hungary
CBC Canada (British Columbia)
CON Canada (Ontario)
FIN Finland
JPN Japan
SWE Sweden
NZE New Zealand
BFL Belgium (Flemish)
ENW England & Wales
ISR Israel

pattern. In the U.S., a highly permeable boundary exists between the school and the college or university; consequently, colleges offer many courses that Europeans would regard as quite inappropriate. Furthermore, in the U.S. mathematics curriculum, introductory calculus has not been a school course. It has been a university course and, until the post-war years, taken in the sophomore year by most students. Only since the late 1940s and early 1950s has even the idea that calculus should be taught in schools been seriously entertained. The reality of school courses in calculus is even more recent, and today such courses are not so widely taught that college programs can presume that students have taken more than precalculus. Moreover, the loose overall framing of college programs means that students may enroll in mathematics courses with only, say, Algebra II in their school programs. This implies that students need not take, nor schools offer, more than a senior-level precalculus course.

Given such fundamental historical and structural differences between European and U.S. curriculums, what meaning can be ascribed

Figure 2.7
The Yields of Mathematics Education in SIMS Systems



to cross-cultural comparisons of curriculums and achievement? U.S. schools are not expected to offer courses in calculus, even to students heading towards mathematics-based college programs, and many schools do not. Wouldn't it have been more meaningful, Americans might ask of a study like SIMS, to compare opportunity to learn and achievement after the point at which the relevant cohorts had had real opportunities to learn calculus? Furthermore, given the requirements for mathematics courses found in most U.S. colleges, could we not expect U.S. yields of mathematics course-taking and achievement to look quite different for students at age 22 than at age 18? In other words, structural and cultural factors inevitably come into play in very central ways as one considers what the questions and the target populations might be in any cross-cultural studies involving the U.S. and in interpreting the finds of such studies. Let me illustrate this observation in another way. How many are taught and learn mathematics in the terminal secondary grade?

One criterion for evaluating the success of a subject within a school system is student [illegible]nent. Thus, some years ago we often heard, for example, that in[illegible] [illegible]0 percent or more of the students enrolled in advanced mathematics courses in that system's grade 12. That figure was often compared to the apparently much lower enrollments by U.S. students in upper level school mathematics courses—with all the dire implications that seemed to follow for the future of our high technology economy. And that brings me to a discussion of the narrower question: How many students are taught mathematics in the different school systems that participated in SIMS?

The most striking finding to emerge from SIMS was the wide differences in the numbers of students enrolled in the terminal grade in school. We faced considerable difficulty in coming up with appropriate estimates of this seemingly simple statistic (Westbury 1988). In the U.S. at the time of the study, approximately 85 percent of the age cohort were enrolled in the 12th grade in an undifferentiated high school system. In Japan, 92 percent were enrolled in the equivalent of Grade 12 but, at this level, the Japanese school system is highly differentiated into school and program types. In England, on the other hand, approximately 18 percent of the age cohort were enrolled full time in the sixth form in secondary schools, preparing for advanced level examinations—but another 12 percent took advanced level courses in colleges of further education. Yet another 20 percent or so took full- and part-time occupational or professional training outside the academic system.

Such statistics illustrate the complexity both of developing a normative sense of what *should* happen in the last year of high school, an institution that takes different forms in different systems. We must ask, then, who is in the high school? What part of the age cohort do these students represent? What kinds of courses are taught in high schools and why? Only then can we attempt to derive meaningful *comparative* estimates of how many students are enrolled in any course at the terminal level of secondary schooling (Westbury 1988). In, for example, the statistic for Japan—that 30 percent of students enroll in advanced mathematics courses—the question is which students make up the denominator in this percentage. On further investigation, it turned out that the denominator was made up of *only* students in academic programs and schools. When all students in school were used as the denominator, the retentivity index became 12-13 percent—effectively the same as that of the U.S.

Because of these kinds of interpretive and technical issues, SIMS sought to discover how many students were enrolled in advanced mathematics as each system defines it by using the age cohort (all young people of the modal age of Population B students in the country) as the constant base for all comparisons. Figure 2.6 shows SIMS' estimates of enrollments in Population B mathematics derived in this way, and Figure 2.7 relates these findings to achievement on the SIMS cognitive tests. As we see, substantial between-system differences appear in the numbers of students who persist with mathematics to the advanced courses of the terminal year. (The U.S. stands in the middle of the range with an enrollment rate of 13 percent.) And, as seen in Figure 2.7, lower enrollment rates are generally associated with higher levels of achievement and vice versa. However, some of the lower achievement systems in the top right quadrant have very high enrollment rates, and some systems achieve both high enrollment rates *and* high achievement.

The between-system differences seen in Figure 2.7 also highlight the range of policy goals that might be available to national educational systems in any subject and, by implication at least, invite us to ask questions that are fundamental for both assessment of these differences and policy prescriptions that might follow: What kinds of goals for enrollment and retentivity in advanced mathematics are most desirable? Is it always desirable to emphasize high achievement as a goal at the expense of participation? Is it self-evident, for example, that school systems should choose policies that defer to achievement at the expense of high levels of enrollment? Is it self-evident that achievement in any single subject or area such as calculus should be *the* criterion to invoke as we consider how to rank or rate any educational system vis-a-vis others?

The importance of the last question is highlighted for North Americans as we consider the pattern of yield seen in the SIMS' findings from British Columbia (see Figure 2.7): 30 percent of the province's 18-year-olds take the system's terminal level advanced mathematics course. Although (as seen in Figure 2.5) these students' achievement is low in calculus, it is middling in algebra. How should we evaluate such a pattern of yield? In the words of McLean, Wolfe, and Wahlstrom (1987), writing from the perspective of Ontario:

> One hears complaints [in Ontario] that entering university students are not adequately prepared in mathematics, that they need to relearn what they have studied in high school. A particular issue in mathematics is the utility and efficacy of the instruction in calculus. In Ontario, virtually all mathematics specialists take calculus in high school. This is also true of most other countries, with the notable exceptions of the USA and British Columbia. In British Columbia, calculus is simply not available to students, and the senior mathematics program is a rather thorough course in algebra and trigonometry. The universities expect to teach calculus. . . . Ontario students study calculus in grade 13 and then study it again in first-year university. Their performance was poor on the SIMS trigonometry item subset.
>
> *Is it really appropriate to have so much emphasis on calculus in a system [like Ontario] that has such a high participation rate [18% of the age cohort]? . . . time could be better spent in the high schools consolidating pre-calculus mathematics or on other topics altogether.* (Emphasis added)

In Ontario, such reasoning provided a rationale for a new Grade 13 course in finite mathematics, covering matrices, combinatorics, probability, and statistics. Many mathematics educators in the U.S. and elsewhere would agree that such a course, with the implications it might have for mathematics for all, represents the direction in which thinking about modern curriculum in advanced school mathematics is going—this is in contrast to the selective assumptions of the late 19th–early 20th century curriculum that we have inherited, in which calculus is its culmination (see Damerow and Westbury 1985, 1987).

As we consider the issues raised by McLean, Wolfe, and Wahlstrom (1987) in the context of this paper, we again have to ask what they might mean for the ways in which we interpret the implications of the SIMS findings on U.S. yield in advanced mathematics. We can still be concerned about the position of the U.S. in Figure 2.7, but, as we consider what should follow from this concern, we must wonder whether we should develop curriculums which thrust, on the one hand, in the direction suggested by the course revisions in Ontario or, on the other hand, in the direction suggested by low achievement–high enrollment systems such as British Columbia. In other words, do the findings of SIMS suggest that we should retool our conceptions of what should be taught and how many students we might try to reach in the light of our educational structures and assumptions and contemporary thinking? And, of course, we must constantly question what such thinking means for the ways in which we interpret the results of studies like SIMS.

* * * * *

In this account of the problems of interpreting SIMS results, I have highlighted the difficulties we face in deciding what the findings mean. All thinking about schooling faces the problem of escaping the limitations we share as a result of our experience in one school system. By necessity, involvement at any level in a cross system study forces us to stand outside our own framework, for a time at least. The inevitable result is that we see our worlds differently. Although I have emphasized the problems that must be faced in interpreting the results of a study such as SIMS, I would insist that these problems are not a weakness of the study but, rather, its strength. By the considering questions I have raised here, we secure new understandings of the limitations of our often unexamined assumptions about what schools are, what they do, what they achieve, and, most important, what they are striving to achieve.

References

Damerow, P., and I. Westbury. (April-June 1985). "Mathematics for All—Problems and Implications." *Journal of Curriculum Studies* 17: 175-184.

Damerow, P., and I. Westbury. (1987). "Conclusions Drawn from the Experiences of the New Mathematics Movement." In *Mathematics for All: Problems of Cultural Selectivity and Unequal Distribution of Mathematics Education and Future Perspectives on Mathematics Teaching for the Majority.* Science and Technology Education Document Series No. 20. Paris: UNESCO, Division of Science, Technical, and Environmental Education.

Keitel, Christine. (April-June 1982). "Mathematics Education and Educational Research in the USA and USSR: Two Comparisons Compared." *Journal of Curriculum Studies* 14: 109-126.

McKnight, Curtis C., F. Joe Crosswhite, John A. Dossey, Edward Kifer, Jane O. Swafford, Kenneth J. Travers, and Thomas J. Cooney. (1987). *The Underachieving Curriculum: Assessing U.S. School Mathematics Achievement from an International Perspective.* Champaign, Ill.: Stipes Publishing Co.

McLean, Les, Richard Wolfe, and Merlin Wahlstrom. (1987). *Learning about Teaching from Comparative Studies: Ontario Mathematics in International Perspective.* Toronto, Ont.: Ministry of Education.

Travers, K.J., and I. Westbury. (1988). *Second International Mathematics Study. Vol. 1: International Analyses of Mathematics Curricula.* Oxford: Pergamon Press.

Westbury, Ian. (July-August 1988). "How Should We Be Judging the American High school?" *Journal of Curriculum Studies* 20: 291-315.

U.S. Department of Education. (1986). *What Works? Research about Teaching and Learning.* Washington, D.C.: U.S. Government Printing Office.

44

How Two Educational Systems Learned from Comparative Studies: The Swedish Experience

INGER MARKLUND

To clarify the impact of IEA's multinational and other international and national research on Swedish educational policy, I must first describe several characteristic features of Swedish society that influence the political impact of research. Some of these may, indeed, be international, while others are quite distinctively Swedish. Of the following five factors, three are definitely Swedish; the other two probably apply to many countries.

Reform Strategies in the Educational Sector

Important changes in Swedish social policies have, until the past very few years, invariably been the result of studies prepared and proposed by large government commissions. This is eminently true of our educational reforms since the 1940s. These commissions often include researchers; indeed, are sometimes chaired by them. In addition, researchers are engaged as expert advisers, research secretariats are set up, and research is generally undertaken at the request of government commissions. Research and researchers are powerful determinants for change in Sweden.

Inger Marklund is Research Director, Department of Coordination and Planning, National Board of Education, Stockholm, Sweden.

The System of Central Administrative Boards

In Sweden (and Finland), the administrative, executive part of national government consists of separate, independent central administrative boards which, formally, are directly responsible to the government and not subordinate to any individual ministry. The national Board of Education (NBE) is one such authority. The boundaries between designing and implementing a policy are not immutably fixed. The authority that implements political decisions also interprets them, which gives it the additional role of policy maker. As far as education is concerned, this was to a great extent true during a tryout period in the 1950s, before the nine-year compulsory, comprehensive school was decided upon from 1962 onwards. In recent years, however, great emphasis has been put on a clearer distinction between the political and administrative decision making, the former resting with the Ministry of Education and government, the latter with NBE.

Swedish Research Policy

Research in Sweden in the liberal arts and social sciences is conducted almost exclusively within our universities. They receive grants from four sources: the state, research councils, national authorities, and industry. Of the four sources, the grants from national authorities are most important for education.

The great majority of national administrative boards, such as NBE, receive special research allocations and are required by their standing instructions to initiate and finance research. This form of funding is referred to as "The Sectorial Principle." NBE has received a special research allocation since the comprehensive compulsory school was introduced in 1962. Its standing instructions require it to plan research, to define research priorities, and to specify, finance, and utilize research as part of its responsibility for school development. This duty distinguishes this sectorially-funded research from the activities financed, for example, through research councils. The national authorities, however, are normally not required to engage directly in research. Thus we do not have any in-house research activities nor do we have separate research institutes for our educational research.

These three characteristics are those distinctively Swedish. The following probably apply to many countries.

International Cooperation

In the educational sector, four overlapping agencies are important in Sweden, though educational questions account for only part of their activities. These agencies are the Organization for Economic Cooperation and Development (OECD), the Council of Europe, UNESCO, and the Nordic Council. Of these, OECD unquestionably means most and has the greatest impact. To a great extent, OECD constantly refuels the Swedish sensitivity concerning our position in various international ranking orders.

Finally, a word about the impact of research on politics, policy, and practice and the conditions favoring or impeding this kind of impact. One can distinguish both the nature and the degree of different effects of research, as well as correlates of the effects. In fact, a governmental committee that reviewed NBE's research and development work between 1978 and 1980 employed this distinction. Incidentally, that committee was chaired by a professor of education, another of its members was a professor, and two of its special expert advisers were professors. In all, about 20 percent of the professorial corps in education at that time was that particular committee.

Effects and Effect Correlates

Effects can vary in kind from general to specific, from attracting a certain amount of interest to attracting interest in a very specific aspect, say a new teaching aid. Effects can also vary in degree from no effect at all to either a very strong positive or negative effect.

Effects correlates, that is, what helps create effects or lack of them, are related to both kind and degree. Usually, only positive effects are looked for, but equally important is looking for what "helps" create negative effects.

Because Swedish characteristics make it easier for research and research findings to get across to politicians and other decision makers, researchers and research are strongly represented in the process of educational reform and in our system of policy-executive and policy-making bodies; research and research-utilization are important fields of responsibility. This adds up to a receptiveness of research and its findings. Include in this equation our dependence on the outside world and

Figure 3.1
Kind of Effects

General effects
|
Attract interest, create debate
Influence attitudes
Give rise to new organizational models
Give rise to administrative rules and routines
Influence curriculum
Introduce new instructional methods
Create new teaching aids
|
Specific effects

DEGREE OF EFFECTS

strong negative ---------- 0 ---------- strong positive
no effects

Figure 3.2
Effects Correlates

to kind of effects

- problem area (broad, narrow, general, specific)
- target group (general public; well-defined, small, interest group)

to degree of effects

- the distinctiveness of the message
- information processes
- the availability of efficient channels of communication
- the readiness to receive the results of research
- the increase of knowledge, both practical guidance and theoretical insight

our need for an economic and cultural place within the international community, and you'll see the principal mechanisms explaining why multinational research, above all within the IEA, and general international research have had an impact on our educational policy and will probably continue to do so. The impact on educational practice has, at times, been a good deal less or, perhaps more accurately, it has been less direct.

IEA Studies: First Generation

My first example is the first generation of IEA studies—mathematics and the six-subject survey. The first mathematics survey (FIMS) was conducted in Sweden in 1964 and reported in 1967. It ought to have aroused greater attention than it actually did, considering what happened when the second mathematics survey (SIMS) was reported between 1983 and 1985.

The results of the six-subject survey were published in early autumn of 1973. Ever since 1970, a fierce political debate had raged in Sweden about compulsory schooling and the decline in standards of achievement which, it was alleged, resulted from substituting this school system for the system of parallel schools. Four major issues occupied the focus of attention: marks, discipline, foreign language teaching, and the rising proportion of special teaching

This debate fell silent in 1973. Sweden's position on the international score sheet belied the great majority of adverse pronouncements on Swedish schooling—in an international light, be it noted.

From 1973 onwards, IEA has been referred to frequently in the Swedish debate on educational policy. IEA proved a source of outstandingly durable arguments in favor of the comprehensive compulsory school. Perhaps the most important finding of all concerned the between-school variance in Sweden compared with other countries. To those of us who have made equivalent educational standards a cornerstone of our educational policy, IEA furnished convincing evidence that, in this particular respect at least, our educational reforms had up to then been successful.

Using the categorization of effects and effect correlates, we get the following picture in Figure 3.3:

49

Figure 3.3
First-Generation IEA Studies
(mathematics and six-subject survey)

EFFECTS

General: +	Created debate in educational politics and policy about educational standards and evaluation
Specific: 0	No effects

EFFECT CORRELATES

To the general effects:	Target groups (politicians), IEA results published during a heated political debate about standards in the comprehensive school
To the specific effects:	The distinctiveness of the message; IEA variables and data to general as compared to national curriculum

IEA Studies: Second Generation

A major commission to prepare for a reform of teacher education programs was already appointed in 1974, but not until 1985 was a government bill on the subject introduced in the Riksdag (parliament). Once again IEA became very much part of the picture, but this time its position was completely different from 10 years earlier. The Swedish results in the second mathematics survey (SIMS) had been published in 1983, but little attention had been paid at the time. No appreciable changes, for better or worse, were apparent. By now, the finding that Swedish compulsory school pupils had not done very well in mathematics in the 1964 survey had been forgotten. Only when the international averages "leaked out" in 1985 did the survey become "hot stuff." In it, the opponents of the proposed reform of teacher education found such convincing evidence of the failure of the Swedish compulsory comprehensive school that the then Minister of Education appointed a special study group to investigate *why* Sweden was doing so badly. That our upper secondary school students were doing quite well remained unnoticed.

Using the same categories of effects and effect-correlates we obtain a somewhat different picture from the one resulting from the six-subject survey.

Here we can see that, this time, the IEA has also had specific effects, and I think these will also become apparent in education practice. The proposals put forward by the mathematics group were exceptionally well received, not only by politicians but also by teachers as well. Those proposals are now being put into practice, and all schools and teachers are affected. Also in SIMS, the between-school variance was very low, which could be interpreted as showing that the standard of mathematics education was equally low all over the country. But even in this context, the low between-school variance was both politically and practically helpful. The measures taken in the resulting national development program, administered by NBE, concern *all* schools, *all* students. Without the IEA survey, mathematics teaching would probably have remained fundamentally unsatisfactory for several years longer. Its unsatisfactory state was apparent before 1983; the IEA findings and the international

Figure 3.4
The Second IEA Mathematics Survey

EFFECTS

General +	Created heated debate in Parliament, government, mass media, schools, and among parents
Specific +	Governmental committee on mathematics education resulted in a study day in *all* schools, compulsory for *all* teachers in grades 1-6 and *all* mathematics teachers in grades 7-9, earmarked resources during four years for INSET and development work and curriculum changes

EFFECT CORRELATES

To the general effects:	A specific *problem area* - mathematics, *target groups* - opponents of teacher education reform
To the specific effects:	*The distinctiveness of the message* - the international low scores, *efficient channels* - very alert media, *readiness to receive results*, teachers, teacher educators

51

ranking of Swedish students, in particular, were an unpleasant, but evidently necessary, reminder.

Second Science Study

The main Swedish findings for the second science survey (SISS) were reported in August 1986. From a *national* viewpoint, it stimulated a number of important, positive results, but until March 1988, all political, media, or other impact was conspicuous by its absence. What were people waiting for? Answer: The international results. After two years spent living in the wreckage left by the mathematics survey, I sincerely hoped that the positive signals we had received from the Swedish analyses would not be swamped by nasty surprises from the international comparisons. I could not help asking myself, though, what, if anything, would have happened if the science survey had been a completely national study, if it had not been known that international results were also expected? Would the silence have been the same?

The full national report of SISS was published on March 1, 1988, together with the main international results. All in all, the general impressions from the first national publication in August 1986 remained positive, though now, for some parts at least, confirmed and strengthened by the international results. The compulsory comprehensive school was judged "approved" by the international ranking order. Not so, though, in science in the final year of upper-secondary school (pre-university), where Swedish students scored below average on biology, chemistry, and physics.

But this time, we could discuss the international and national results simultaneously, and this contributed much to the, on the whole, balanced discussion that took place. The shortcomings of the science students could be discussed in terms of what and who they were compared with and this started a discussion of, among other things, the pros and cons of a highly specialized science education, such as those of England and Hong Kong, versus the almost encyclopedic upper-secondary education in Sweden.

It is too early to predict the effects of the SISS study in Swedish science education. National and, above all, international results are still to come. The most important lesson to be learned from SISS is this: It is imperative that national and international results are published and discussed simultaneously. If not, international results will remain an Olympic game with very limited possibilities of being explained by and added to the understanding of national results.

Written Composition

We are now working on the third IEA study, Written Composition, and neither the national nor international analyses have yet been completed.

This study is basically different from its predecessors; the extensive scope of the qualitative data is giving us the pupils' picture of part of their reality both in and outside of school. Through their own production we will gain a picture of Swedish mother tongue instruction that no conventional tests or questionnaires could give us. I believe that this multinational study will have considerable impact on educational practice. Whether or not it will have any political impact remains to be seen. Perhaps we can look forward to bottom-up effects. Perhaps this time the changes in teaching which may result from both the national and international analyses can be directly initiated by the pupils.

Overall Conclusions

I would like to conclude with a few reflections. In Sweden, we can observe a development of the impact of IEA's multinational research on educational policy and practice. The IEA studies of the 1960s and 1970s had a great deal of political impact, but did not make any practical impression. They did not lead to curricular changes or changes in teaching methods, teacher education, inservice training, teaching materials, or ultimately, of course, in the pupils' command of mathematics.

The science survey can have both a political and a practical impact, and I hope and believe that written composition will also have an impact on teaching practice.

How have these developments come about?

The IEA studies have partly changed character. Aspirations have risen, and, within the IEA, analytical models are being developed that emphasize systematic understanding and the truly comparative aspect beyond the international "top of the form."

Sweden has intensified its national follow-up. We are no longer content with international averages and ranking orders; we are trying, more and more, to penetrate the background of these data. We are not going to have mathematics teaching which adheres to a kind of international core curriculum. IEA has not been effective in this way, but IEA forced us to take a long, hard look at what is happening to the teaching of mathematics in Sweden.

In total figures, IEA is attracting more attention today and from more people. The interest shown by the media, admittedly, is not always salutary. Here we have to plead against the self-assumed duty of journalists to highlight the bad news. Perhaps this dismal image can, to some extent, be offset by a desire on the part of politicians to see a positive outcome from the decisions they make.

Finally, I would like to quote what the Director General of NBE, Erland Ringborg, had to say on this subject in his address at the 26th IEA General Assembly, Open Day Seminar, in Stockholm:

> Critics of the IEA sometimes bewail the tendency of politicians to latch onto superficial comparisons between countries as a means of scoring political points. Personally, I feel that we should refrain from this kind of moralizing and acknowledge the realities of politics. If the surveys have the scientific substance we assume, then let us instead turn the whole thing to the best possible account. Let us supply our politicians with national averages and ranking orders and also with those analyses of the underlying systematic factors that may serve to explain the results.

How Two Educational Systems Learned from Comparative Studies: The Hungarian Experience

ZOLTÁN BÁTHORY

Continuous measurement of the effectiveness of public education has been regarded as a primary concern of Hungarian educational researchers for almost 20 years now. By "effectiveness," we mean learning achievements rather than the quantity of marks or certificates. We contend that the accountability of an educational system can justifiably be described in terms of learning achievements.

We measure this effectiveness in the frameworks of *national* and *international surveys*. To do that, two important methodological problems must be solved: which *grades* and which *fields of education* (school subjects) do we survey?

In both Hungarian and international surveys, we usually appoint three strategically important grades: Grade 4 as the final year of elementary school education (Population 1), Grade 8 as the final year of the eight-grade general school (Population 2), and Grade 12 as the school-leaving year of secondary school education (Population 3).

The selection of fields of education to be surveyed also needs careful consideration. In Hungary, the view is fairly common that the fields of education chosen should be both fundamental in the learning process and those in which a high degree of validity can be ensured. In our

Zoltán Báthory is Head, Research Department, Országos Pedagógiai Intezet, Budapest, Hungary.

opinion, reading comprehension, mathematical-logical skills, and the cognitive side of science education can satisfy these two criteria.

We are, of course, aware that this system provides a rather reduced view of the evaluation of school education. We know that the effectiveness of formal education cannot be reduced to some basic skills and knowledge (minimal competence). Nevertheless, in order to achieve "softer" educational effects in the development of students' personalities, a basic standard of fundamental skills and knowledge is needed as a reliable system of guarantees. Our surveys, therefore, intend to inform teachers, school administrators, and researchers if this basic standard exists for the softer, more complex effects. Therefore, we consciously accept the reductionism of our philosophy of evaluation.

I have already mentioned that we measure the effectiveness of our formal educational system in the frameworks of national and international surveys. Here, I concentrate solely on *international surveys*—describing important methodological and organizational conditions and some research findings. And finally, I discuss the significance of the international surveys from a national point of view. It is my view that, in the European culture, the international comparison of learning achievements is not appreciated as much as it should be, given the importance and weight of the issue. Educational comparisons often deal with educational philosophies, input factors, and other issues of the educational process, but neglect the output side. Therefore, let me grasp this opportunity to speak about the *importance of the comparison of output* and its assumed advantage.

Methods and Set-Up

We have been conducting international surveys under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) for years. IEA studies involve students of the same age, attending the same grades, but studying in different educational systems. Their learning achievements are measured and compared to arrive at conclusions about development in the fields of education to be studied and the improvement of the whole school system. By comparing achievements and the learning environment, we can mutually benefit from each other's experience. This, of course, presupposes a high degree of standardization of methods and the realization of the mutually worked-out project.

Research findings are first published in an international volume as studies of the participating countries. Interpretations range from simple descriptive analyses to multivariant mathematical-statistical analyses that uncover cause-and-effect connections as well. The relatively simplest and most easily understandable interpretation is the rank order of national averages. This sometimes appears as if IEA were conducting some kind of "intellectual Olympic Games." The international analyses, however, show that the aim of the studies is to understand and improve the educational system of the participating countries. Our account will rest primarily upon the rank order analysis.

The first IEA survey in 1970 tested reading comprehension skills and science knowledge. Mathematics was the next field of education to be studied in 1980 and 1981, and in 1983 science achievements were measured again. (This list is not complete as there were other IEA studies.)

The Main Research Findings

Reading Comprehension

The 1970 findings of the reading comprehension study showed that Hungarian students' average achievements in all grades involved were generally below those of other IEA countries. It follows, therefore, that education in the Hungarian language and communication skills was one of the weakest fields in effectiveness. Subsequent national reading comprehension studies (in 1979, 1980, 1986) essentially supported the IEA finding.

The rank orders of national averages, also indicating the range of the two inner standard deviations, in the international survey showed that at the 4th grade and 12th grade levels, Hungarian students were next to the bottom; at the 8th grade level, they were eighth out of twelve.

The poor results of the reading comprehension survey were for us like a bolt from the blue. Both professional circles and the public reacted strongly. Experts mainly considered changes in teaching methods, and as a result, silent reading became accentuated. Also, we believe that the IEA findings contributed to the spread of diverse, alternative teaching methods of initial reading from the late '70s. Until then, only one teaching method of initial reading had been used, in accordance with the central curriculum. These research findings, besides other factors, also contributed to considerable innovation of education in Hungarian language-communication skills.

Mathematics

The IEA has so far studied the effectiveness of mathematics education twice. Hungary did not participate in the first study in the '60s. In the second survey, however, the achievements of 8th grade students (Population 2) were measured in 1980, and in 1981 those of 12th grade students (Population 3) were measured. Unfortunately, the international volume has not yet been published, and national evaluation studies are also missing. Nevertheless, we have rank order of national averages from American sources at our disposal. They show that Hungarian students in the 8th grade are well above the international average and that at the 12th grade level, the top 5 percent score fifth out of twelve.

Rank order shows that Hungarian students' achievements are usually better than those of the averages of other IEA countries. Though we do not know the real causes and circumstances, the results clearly indicate the effectiveness of Hungarian maths education in the period when we discarded the traditional arithmetic-geometry teaching for the sake of new maths (1975-1985).

Science

The first international science study took place in 1970 and the second in 1983. A preliminary international report on the latter was published in March 1988 (*Science Achievement in Seventeen Countries*, Pergamon Press). The main points of the book were made public at press conferences in London and a few days later in Budapest. Both IEA science studies indicated, at least according to the achievements scored in the test papers, that this field of education in Hungary can be regarded as effective. Our results stand the comparison with those of Japan! This conclusion is unquestionably backed by the following rank orders:

In 1970, the 4th graders were in the middle, the 8th graders second, and the 12th graders seventh. In 1983, their respective ranks were fifth, first, and third (but with a larger number of Asian countries in the field).

The relatively high achievements of the Hungarian students result from several factors. Among the most conspicuous are that the political and professional concerns in science education coincided, that teaching materials (curriculums, textbooks) are on a relatively high level, and the relative quantity of time devoted to science teaching is high. Science education can, at the moment, be regarded as the prime branch of Hungarian education, thanks to the work of schools, teachers, and students. I would not like to say, however, that we have nothing to improve. For instance, we need more student experiments and practical knowledge.

Overall Conclusions

In our case, the various IEA studies led to contradictory results. On the one hand, while it turned out that our science education and partly mathematics education are more effective than in many IEA countries that are more highly developed than Hungary, language-communication education, on the other hand, lagged considerably behind. So we got into a schizophrenic state of mind. Several publications came out to explain the reasons, and especially since March 1988, when the results of the 1983 science study were published, there have been fierce debates. The press, television, and radio all began to deal with the problem. The interpretations called researchers' attention to new approaches and new aspects, and researchers could justifiably argue for the importance of differentiated evaluation of the educational system. This was almost unprecedented in the Hungarian mass media.

The Benefit of International Studies

Based on the Hungarian experience, we can note three concrete advantages of empirical-comparative educational surveys. Perhaps the most important outcome is that we have succeeded in getting accepted the principle that Hungarian students' level of knowledge can, and must, be compared to international standards, that is, to the level of developed countries and not to provincial, politically set criteria. Many new facts and connections were revealed during the design and realization phases of the IEA studies that contribute to the always-present fear of provincialism, a natural consequence of the national character of our education. Very often, it becomes obvious that a number of similarities exist between educational systems though they function differently in various cultures. Thus the IEA-type of international educational studies is an important means of improving the cooperation between the member countries.

Another benefit is also obvious and connected with the previous. In light of the results, we can more profoundly analyze the input-output factors of educational systems. We can not only point out whether a particular content or organizational factor functions in a given culture, but we can also make comparisons regarding its effectiveness. IEA studies have contributed to the understanding of such issues as the effectiveness of comprehensive versus selective school systems or the connection between excellence and mass education. The more students attend school, the more talented children can be identified. Understandably, though, the increasing numbers of attendance affects the quality of education.

A third merit is concerned with research methods. During the years of cooperation with IEA, we have adapted to our own circumstances a research methodology that makes possible valid and reliable measurement of learning achievements. In Hungary, it has led to a new educational evaluation system that can provide regular and reliable information to schools, politicians, teachers, and the public about the actual state of students' knowledge.

What IEA Studies Say about Curriculum and School Organization

EDWARD KIFER

The international arena in general and the results of IEA studies in particular are fertile grounds for those who wish to cite successful practices that deviate from existing ones. As stated in one of the first official publications of IEA:

> If custom and law define what is educationally allowable within a nation, the educational systems beyond one's national boundaries suggest what is educationally possible (Foshay et al. 1962).

French schools are closed on Wednesday afternoons. New Zealand children begin school on the day of their sixth birthday not on the first day of school that year. Japanese mathematics teachers have comparably large numbers of students to teach, but they teach just three of the six days of the school week. All children in Sweden are introduced to a second foreign language. Elementary school pupils in the United States recite a Pledge of Allegiance each morning.

The list could be expanded almost indefinitely, but these are enough to make the point that variation in policies, procedures, and practices found in other educational systems provide a counterpoint to any one system's views and a basis for change, innovation, and reform. These differences are embedded in cultural contexts that act as restraints to their wholesale adoption. They serve, however, as a beacon for what can be done.

Edward Kifer is Professor, Department of Policy Studies and Evaluation, College of Education, University of Kentucky, Lexington.

Curriculum broadly construed and diverse organization of schooling are major themes in IEA reports. The curriculum not only forms the backdrop to IEA's cognitive tests, but also is scrutinized thoroughly for its contents, resulting in innovative empirical measures of its implementation. School organization is a focus, as well, since histories and social forces produce very different types across the 40-odd systems studied by IEA in its short but productive history.

I deal first with issues surrounding the curriculum and then with those surrounding the organization of schools. How schooling is organized is intimately connected to curriculum since differences in school structure are often associated with who in a system gets what kinds of experiences. The extent to which a system's curriculum is differentiated and how that relates to school organization forms the third major section of this chapter.

Issues Surrounding the Curriculum

The IEA approach to curricular issues is scientific, outcome-driven, and focused on finding commonalities among curriculums of the participating educational systems. It is scientific because the rationale for the studies involve systematically collecting objective, empirical evidence about each system. The approach is broadly based on views formulated by Ralph Tyler in the 1930s and emphasizes behavioral specifications of what is to be measured empirically.

It is outcome-driven because a major portion of the curriculum analysis accompanying each survey is devoted to developing a content by behavior grid to provide a basis for formulating an objective test (usually multiple choice) that can be used to compare cognitive outcomes of the various systems. Whether it is a mathematics survey where the content dimension might include, among others, arithmetic and algebra, and the objective dimension would include, again among others, computation and solving problems, or a literature survey that includes specific texts as well as literary theory for content and familiarity with and expression of a consistent response to as the objectives, the grid forms a kind of Procrustean bed for the outcomes of a survey.

For obvious reasons, IEA searches for comparability and commonalities across educational systems. Objective, scientific study of educational systems is based on the assumption that commonalities exist among the systems and a sense of fair play suggests that comparisons should be made about those things that are similar. In the search for common instances, diverse and unique phenomena must be left behind.

As with all such work, however, to the delight, one assumes, of educational reformers, idiosyncratic practices and diverse opinions spilled over the edges of IEA science. So we not only have an opportunity to talk about things that seem more or less related to differences between systems (the scientific results), but also about evidence of unintended outcomes of the studies.

Time and the Opportunity to Learn

Two related but different notions are central to IEA investigations. The first is the amount of time allocated to different subject matter; the second, a unique and innovative IEA measure, *Opportunity to Learn*, that attempts to define what students are exposed to in a particular curriculum area.

Amount of Time

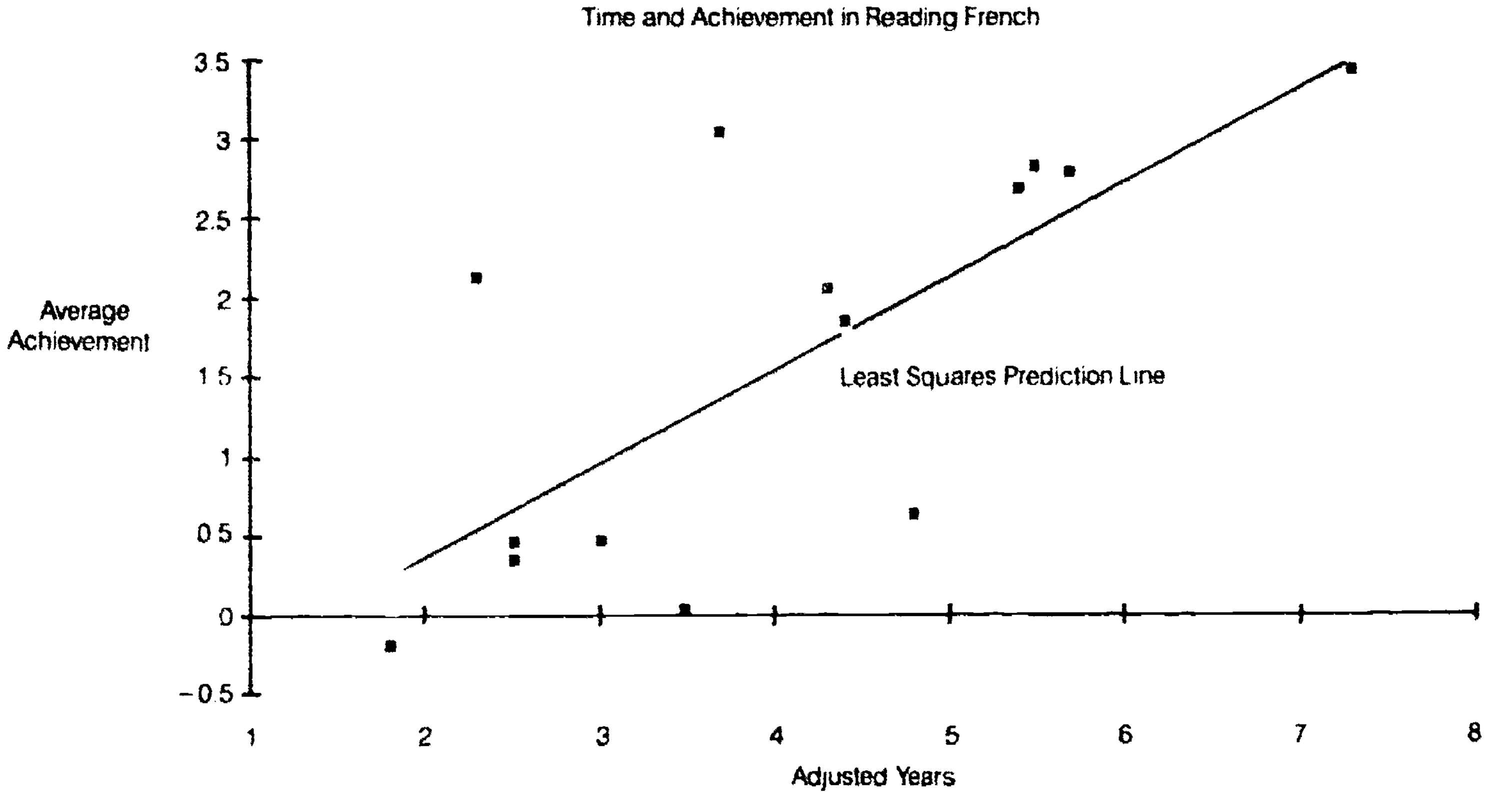
A variable, the amount of time allocated to instruction, features in a number of IEA studies, but is most central to results of the study of French as a Foreign Language. Carroll (1975) presents a series of analyses that, over educational systems, relate the number of years studying French to average performance in reading, writing, and listening.

Figure 5.1, based on the data reported by Carroll, depicts a strong relationship between number of years of French study and French reading performance. That figure is similar to those published in the IEA report for each of the criterion variables and leads to an obvious conclusion: the longer students study a subject, the more proficient they become in it.

Carroll estimated that an average student would take between six and seven years to attain a "satisfactory" (and high) level of performance in French. That amount of time could be reduced if the student were fortunate enough to (a) have a fluent teacher (an effect of almost one year); (b) be taught mainly in French, not in a mother tongue (an effect of a year and a half); and (c) have serious aspirations to learn French (an effect of about a half year of instruction).

The implications for reform are both obvious and sensible. If a particular learning outcome is desired, then a frame or window in the curriculum must provide a sufficient amount of time for the learning to occur. That learning is enhanced, and the size of the window decreased,

Figure 5.1
The Relationship Between Number of Years Studying French and Average Achievement in Reading French Across Countries



by the presence of effective teachers using appropriate methods with students who are highly motivated.

The elegance of these findings should not be lost because they appeal so wel to common sense. A curricular reform, for example, based on a fixed standard of performance, using empirical estimates of the amount of time needed to reach that level of performance, that could specify what qualities are needed by a teacher and what teaching methods are most effective, and responded specifically to the needs and aspirations of students would, indeed, be a reform. These findings, therefore, provide a beginning, a direction, and a challenge to reform.

Opportunity to Learn

IEA studies contain an Opportunity to Learn (OTL) measure that is an ingenious and innovative attempt to learn the extent to which a particular curriculum is implemented. Based on achievement tests constructed from the grids mentioned earlier, OTL is measured by asking each teacher in a survey to respond to each test item in terms of whether or not his or her students have been exposed to the material needed to be able to answer the question correctly.

Figure 5.2 on page 56, from the science study (Comber and Keeves 1973), shows the relationship between the OTL measure and average achievement across the various educational systems. While there are exceptional cases (the two low outliers being systems in developing countries), there is also a strong positive relationship between achievement levels and the opportunity to learn the material.

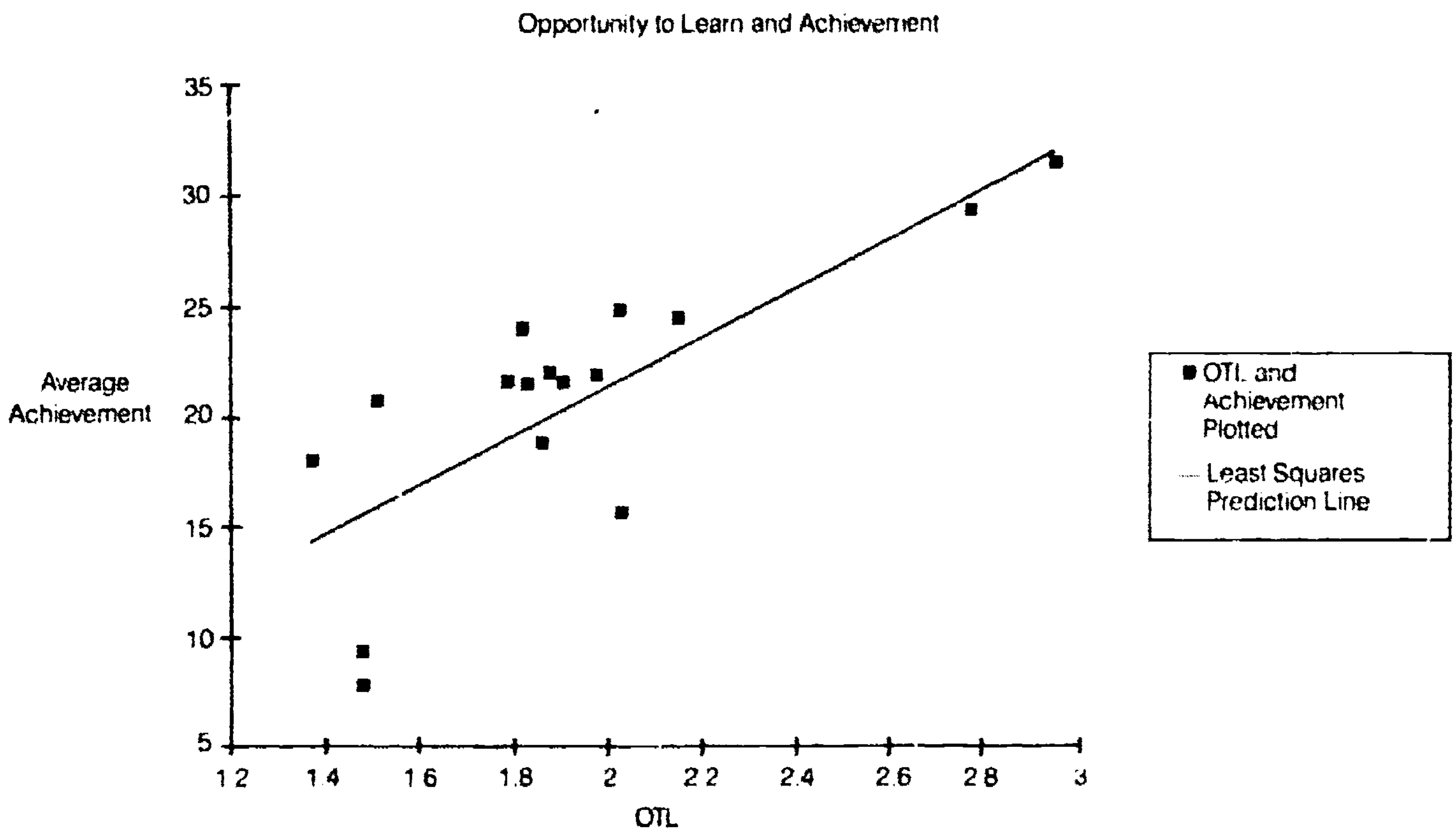
This general finding also appeals to common sense. Students tend to learn what they are taught. Yet, all is not so simple. This finding occurs across systems that vary greatly in terms of national practices, such as having or not having national curriculums or common national examinations.

Reformers in a system without a national curriculum would find little evidence in IEA studies for the efficacy of a curriculum more centrally determined or more broadly applied. In the second mathematics study (Burstein in press), for instance, both the Japanese and New Zealand systems are based on national curriculums. While the academic performance of the former is the highest of the 20 systems studied, the latter's achievement is more modest.

OTL, then, is a measure that speaks to the issue of what students are exposed to in a classroom. Relationships between OTL and achieve-

Figure 5.

Opportunity to Learn and Average Achievement Across Systems: 14-Year-Olds in Science



ment are related to the amount of content covered in classrooms, not necessarily to the extent the curriculum is standardized. Another way to state the finding is to distinguish between curricular aims at whatever level they are specified (local, province, or system) and what a teacher actually does in a classroom. OTL reflects what is implemented; not necessarily what is desired or intended.

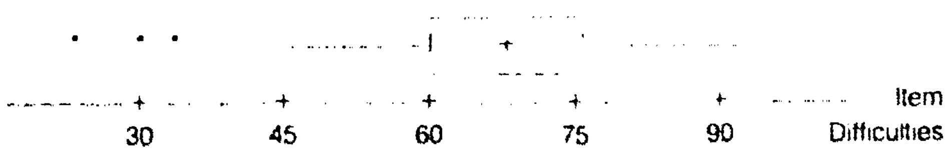
Results of OTL, as well as other related analyses across surveys and systems, suggest also that the amount of material covered, not the sequence, is a predictor of levels of achievement. Several sets of results point to the secondary importance of sequencing. In the "feasibility study" (Foshay et al. 1962), correlations between item difficulties across systems were produced for four subject areas: mathematics, reading, geography, and science. Figure 5.3 presents both a stem and leaf display and boxplot of the correlations between item difficulties for the geography test. The correlations between these item difficulties should be uniformly high if different systems approach the content in similar ways (i.e., that sequencing of material is as important as amount of material covered). While there are differences between subject matter (mathematics correlations tend to be higher than the others), the results are notable for variation between systems rather than similarity.

Results of other surveys, too, suggest similar things. Both the first mathematics study (Husén 1967) and the science study report single item results from systems in which a particular item is either more difficult or easier than expected if performance across the item set were consistent between countries. While the literature survey (Purves 1973) found that opportunity to learn is important in terms of explaining achievement differences between systems, students in those systems typically respond differently to the various reading passages. That is, what they are taught is related to achievement differences, but how they are taught is not.

Evidence for the negative effect of redundancy, a counterpoint to OTL (too much time spent on too narrow a set of outcomes and not enough exposure to new topics), comes from two recent sources. McKnight (1987) pegs the relatively low performance of U.S. students on mathematics achievement tests in the second mathematics study to a curriculum that moves very slowly, contains huge amounts of redundancy, and, compared with other systems in the study, introduces fewer new topics. Kifer and Wolfe (1985) found a correlation between pretest scores and OTL. This demonstrates that students exposed to material in

Figure 5.3
Stem and Leaf and Boxplot of Correlations Between Item Difficulties Across Systems: 13-Year-Olds in Geography

Stem	Leaf
2	6
3	1
3	5
4	0
4	9
5	444
5	5568
6	0122344
6	55667777789
7	00012444
7	6777
8	022444
8	5689
9	03
9	5



a previous year have a higher likelihood of being exposed to the material again than do students who have never been exposed to it.

Again, the OTL results suggest that the amount of material covered counts more than how it is covered. That, of course, makes a good deal of sense. While students do not learn all they are taught, they do not learn what they have not been exposed to.

Summary

Two crucial alterable variables of the IEA surveys are the amount of time allocated to a subject and the opportunity to learn important features of it. Although both of these measures are related to how educational systems use time, they have different implications.

Where Carroll's findings include implications for the allocation of time, teacher competence, and teaching methodology, the OTL findings suggest that the relative importance of the variables place exposure to content very high on a list and teaching variables and "sequencing" of the curriculum rather low. This suggests that those interested in curriculum reform should work diligently to include coverage of "desirable" content in educational systems, but worry less about how the content is handled once there. Exposure to the material is the necessary condition for learning. Sequence, pacing, and teaching variables follow in importance. The mere existence of the subject matter placed before students is a crucial ingredient of any curriculum reform.

These findings emerge from a "scientific" view of how to study diverse systems internationally. Such findings mute differences between the contexts in which schools operate and emphasize common activities that can be put into operation and measured more or less precisely.

Educational activities in schools are not necessarily conducted in optimal ways. For example, the average French child acquires competence in his or her language in two years, but it takes seven in other educational systems to produce arguably less facility. For the French child who masters the language in two years, the issues are not those of learning French nor how he or she is taught. Dominating those two years is a social world that the child is trying to understand.

Not only are educational systems different from a social world in terms of efficiency, but they also vary in the extent to which scientific models apply to what they do. In each of the findings cited, some systems deviate substantially from what is found on the average. Typically, systems in developing countries do not perform as predicted by a scientific model. Other systems, for example Rumania in the French study, do much better than a model would predict. This reminds us that any reform should be as cognizant of the needs of students, the idiosyncratic aspects of a social context, and the peculiarities of a culture as it is of the scientific findings.

These findings, then, represent a starting point for curricular change and reform. They do not define the limits of what can be changed or how such change should be effected. They point to a way for a reformer to begin to think about such matters.

School Structure

Educational systems surveyed in IEA studies vary dramatically in their structure—how they are organized and who they serve. School structure and its implications formed a major focus of early studies.

Interest waned during the period of the six-subject surveys, but rekindled in the latest surveys. While the next section deals with the inter-relatedness of school structure and curriculum, the focus is primarily on issues raised by different types of schools in the various systems. The following three sections deal with findings related to type of school, selectivity and retentivity, and the notion of yield of a system.

Type of School

Virtually all the IEA surveys measured the type of school in which students were enrolled. For academic purposes, such types could vary from comprehensive schools (all students attend the same type of school) to selective schools (schools where only an elite can enroll). In systems with selective schools, other types of schools are usually available to students not among the academic elite. Examples include schools with a vocational or technical curriculum.

Systems vary not only by types of schools, but also in terms of when students are channeled into the different schools. For example, by age 13 students in Flemish Belgium are enrolled in different types of schools. In the United States (at least at the descriptive level) all students are enrolled in comprehensive schools for as long as they attend school.

The variable school type figures prominently in IEA publications, especially in the six-subject survey, where it was included among a block of variables in a series of regression analyses. It was a powerful predictor of achievement, mainly because systems, on the average, place more or less capable students in different types of schools. While school type is a major and powerful predictor of achievement, its implications can be traced through two different and related issues popular in IEA volumes.

Selectivity

One major issue addressed by the first mathematics study was to discover the circumstances under which the most talented students develop best. The argument for selective schools is that talented students, learning together, will not be held back by less talented members of their cohort; therefore, to develop the best talent a system should be selective not comprehensive.

The analyses of the data, however, produced a resounding condemnation of selective schools. Although average scores in selective educational systems were higher than averages in comprehensive schools, a variety of analyses show that the "best" students did equally well regardless of the type of school in which they were enrolled.

While there is no obvious advantage of elite schools in terms of the highest performing students, a comprehensive system benefits generally if the criterion is the yield of the system (Husén 1967). That is, if one could posit an indicator representing the sum of mathematical knowledge (as opposed to the average amount), then comprehensive systems are clearly preferred. More students knowing some mathematics produces a higher yield than a few students knowing a lot of mathematics.

The findings from the first science study are equally unambiguous on questions related to selectivity of educational systems. A strong model to predict average achievement levels includes a large adjustment for selectivity, but small adjustments for differences between systems for the performance of top students. These results support those mentioned earlier: comprehensive education appears not to harm the highest achieving students and increases the overall amount of knowledge produced by an educational system.

Retentivity

Retentivity, another issue addressed in IEA studies, is related to selectivity and yield, but focuses on a different set of questions. While most developed countries have compulsory schooling until at least age 16, a varied percentage of the student cohort attends some type of school until time to enter a college or university. At the terminal year of secondary school, anywhere from about 10 percent to 90 percent of the cohort is enrolled. When who is still in school varies so dramatically between systems, it comes as no surprise that those with a small percentage of students still in school have higher achievement averages than those that have high retentivity—a larger proportion of the cohort still enrolled. Rather the question is whether enrolling more students in school lowers the achievements of a few. Again, the IEA results across surveys suggest "No." While increased retention of students does influence negatively average achievement values, it appears to have no substantial influence on the highest scoring portion of the achievement distribution. Hence, a system can increase the number of students who complete secondary school without harming the education of an elite. The net result of increased retention is lower average scores, more knowledge produced, and no appreciable influence on the highest scoring group.

Summary

IEA results suggest that an expansion of schooling should be both comprehensive and nonselective. The system would produce more knowledge without penalizing its most capable students.

The findings are based on assumptions, however, that more schooling for longer periods of time for more students represents defensible needs of the young and is a desirable goal for an educational system. Challengers to these assumptions argue that more schooling for more people is both unnecessary and fundamentally coercive: many other settings are available in which people can learn; the more technological the society, the less schooling needed; and, adolescents need real work and real experiences rather than more schooling. Implementing change based on IEA results on retention and selectivity, then, should consider first the assumptions upon which the arguments against selectivity and for increased retentivity are made.

Curriculum and School Structure

Participation versus Achievement

Issues surrounding selectivity and retentivity push the educational reformer away from questions about achievement and talent as criteria for accomplishment to a position that makes participation in schooling and the terms of that participation major focal points. If one wants to educate only an elite then the system should be selective and retain fewer students. Only after the decision has been made that the masses are capable of being educated and making a commitment to educate each individual does selectivity look undesirable and retentivity become a goal. Hence, participation in schooling becomes a criterion for a desirable educational policy.

Participation, however, is an ambiguous term when it comes to schooling. A system can easily demand that students be present in schools without offering them the best the system has to offer. Systems which are neither elite nor specialized and have high retention rates can still reserve the best schooling for a small number of students.

Tracking and Sorting Students

Tracking and sorting of students becomes an issue similar to selectivity when retention is high and schooling is long because not all students are offered opportunities to participate fully in school life. In his feasibility study, Pidgeon (1962) used the standard deviation of test scores to draw conclusions about the effects of "streaming" in England. While trying to understand the low average performance but great variance in achievement, he concluded:

> Perhaps exerting a greater influence, however, is the belief a teacher may have that innate ability is of paramount importance in determining the level of attainment to be expected from a child. Streaming by ability, which is viewed as an administrative device resulting from the acceptance of this belief, will merely tend to enhance its effects. When all these factors act in the same direction the effect will clearly be greatest and this is what happens in England. Here, it is claimed, the aims and, more especially, the beliefs of most teachers and educational administrators lead them to expect wide differences in performance, and this is what is therefore achieved. Where, on the other hand, the grade placement system operates and especially where, within such a system, teachers do not attempt to measure innate ability and therefore do not expect their pupils' attainments to be matched to it, then the dispersion of achievement will be much less.

Whatever educational assumptions underlie the use of tracking and streaming, these practices have clear consequences in terms of curricular issues. The flip side of tracking is a differentiated curriculum.

Differentiated Curriculum in Grade 8 Mathematics

Analyses from the Second International Mathematics Study (SIMS) produced a substantial number of results addressing the effects of a differentiated curriculum in Grade 8 (Burstein in press). A pretest was administered in eight educational systems at the beginning of the school year, making it possible to describe how students are allocated to classrooms or schools within those systems. In another paper (Kifer in press), I describe how the variance decomposition of Figure 5.4 reflects decisions made to track students either into homogeneous classrooms or different types of schools.

In Figure 5.4, areas of the circles are roughly proportional to the total variance in achievement for each system. The wedges within the circles represent percentages of total variation found between students, classrooms, and schools. The circles containing only two wedges depict systems that did not sample two classrooms per school. In those cases, the variation is labeled *student* and *school variation* although, theoretically, the wedge for school contains both the classroom and school variance. That is, the between-classroom variation, if any exists, is part of the between-school variation, not the between-student.

The differences between systems are dramatic. Not only are the total variances (individual differences within a system) of strikingly different magnitudes, but also that variation (how the system responds to individual differences) is divided in distinct ways. In Japan, for instance, almost all of the large total variation is between students. Since such a small amount of between-school variation exists, variation between classes in the same school must likewise be small.

Apparently, the Japanese either ignore individual differences when assigning students to classrooms or their educational policies produce equality among classrooms and schools. Neither homogeneous grouping in mathematics nor sorting by school takes place in Japan in Grade 8.

The U.S. provides the most extreme contrast to the Japanese pattern of variation. The magnitude of the between-classroom component, which represents tracking in U.S. schools, is its single largest component and exceeds comparable values in all of the other systems.

While the distinctive patterns of other systems have been discussed elsewhere, the major conclusion is worth repeating. These eight systems demonstrate a wide variety of practices in sorting students into different classrooms or schools. They range from what is done in Japan, France, and Ontario (where no sorting has yet occurred by Grade 8) to what is done in the United States where more variation in achievement is allocated to sorting practices than to individual differences.

Consequences in the U.S. Since the U.S. shows the largest effect due to tracking, it is worth reporting what is known about the consequences of thepolicies that produce it. A typical pattern is to form, in any school with a sufficiently large enrollment, four types of mathematics classrooms—algebra, pre-algebra, general, and remedial classes. The typical rationale for this isthat homogeneous grouping provides better instruction. What is ignored isthat the practice also produces fundamentally different patterns ofopportunity, participation, and exposure. In Grade 8, some students areexposed to mathematics content and experiences that others may never get. This is opposed to what happens in other systems with no comparable tracking.

Figure 5.4

Variance decomposition of the population A (grade 8) pretest core test for each of the eight longitudinal systems. The area of the 'pie' is roughly proportional to the size of the overall variance in each of the systems The 'wedges' of the pie reflect the proportions of the variance that are attributable to differences between students within classrooms, classrooms within schools, and between schools. Those systems that have only two wedges did not sample two classrooms per school; hence, one cannot distinguish between the classroom and school variances. Given the nature of the samples, however, the proportion of the pie attributable to school differences contains the classroom differences.

Japan's pie is larger than the others in part because of a longer pretest. Theirs had 60 items versus the 40 items for the other systems. Other analyses suggest that with a test of 40 items, the Japanese variance would be about equal to or perhaps slightly larger than that of the USA.

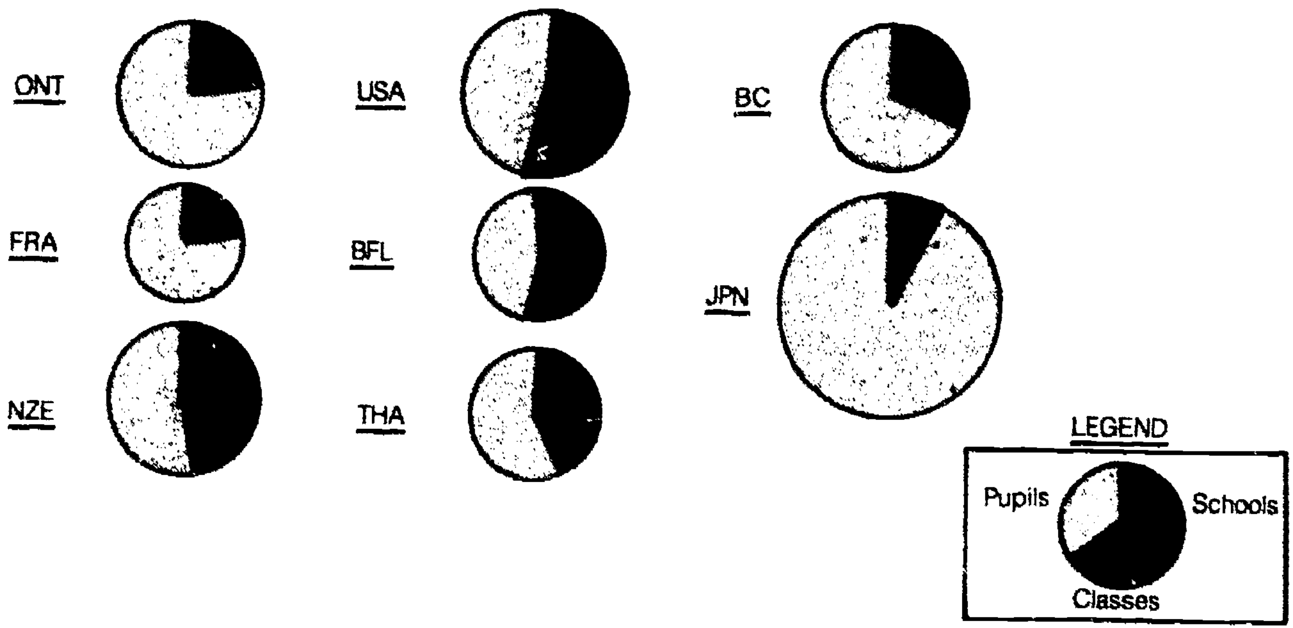


Figure 5.5 contains a set of boxplots showing OTL ratings by classroom type. It is evident from Figure 5.5 that tracking students and differentiating the curriculum are two sides of the same coin. Students in algebra classes are exposed to very different kinds of material from those in other classes. Students in remedial classes are exposed to very little mathematics in comparison with other groups and are likely to have been tracked so that they never will be exposed to the material provided to the algebra students.

These practices deny opportunities for the many to experience what the few get, plus other equally undesirable side effects (Kifer 1984). These include the following:

1. Sorting is replete with misclassifications. Only about 50 percent of the top 10 percent of students are placed in algebra; the others are spread among the remaining class types.

2. Sorting favors white women students from wealthy homes. If placement in high level classes were strictly meritocratic, more boys, members of ethnic groups other than white, and more children from poor families would be in those classes.

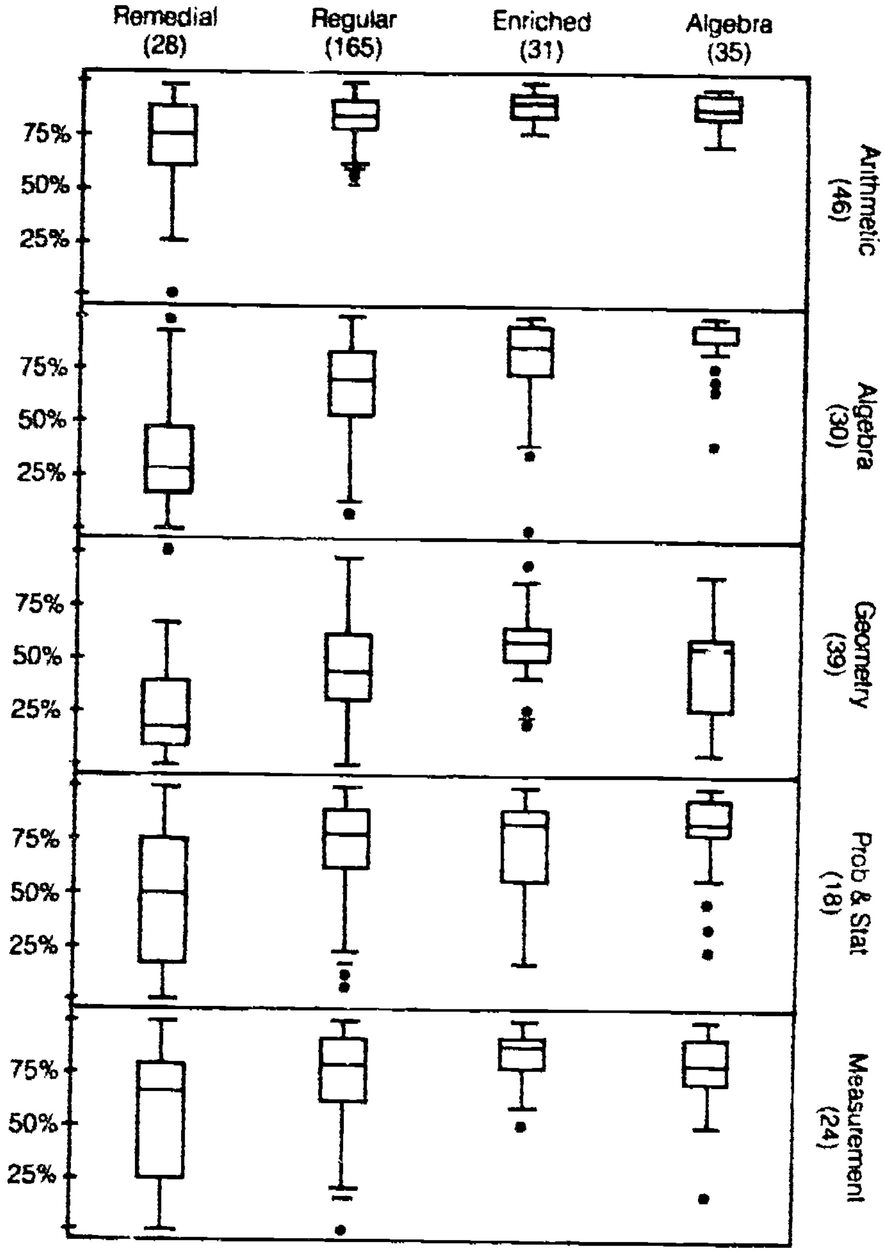
3. Selection is so rigid and so early that it effectively eliminates the major proportion of the cohort from participating in the most advanced mathematics offered by the school system. (This is discussed in more detail below.)

All school systems, not just the U.S., sort students. These results show the effects in a system that does the earliest and greatest amount of tracking. Others tend to sort later in a student's career. No system studied by IEA does no sorting.

Population B

Figure 5.6 shows that issues surrounding participation in Population B (the terminal year of secondary school in SIMS) are very different from those of Population A (Grade 8). Virtually all students take some type of mathematics in the 13-year-old population. By the end of secondary school, depending on the system, a large proportion of the cohort is either no longer enrolled in school or not taking mathematics or both.

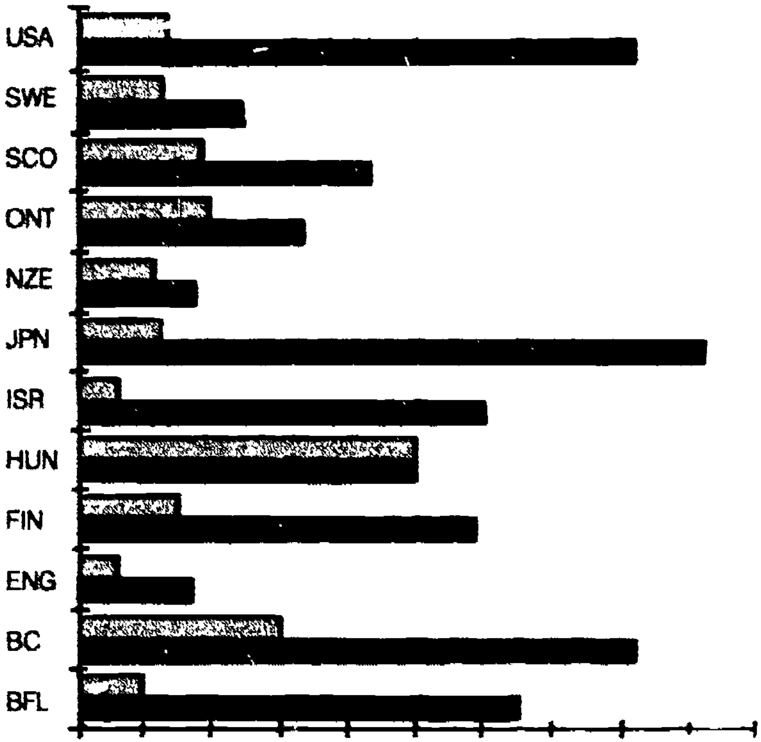
Figure 5.5
OTL by Content Area for U.S. Class Types



Source: Richard Wolfe

Figure 5.6
Participation in Schooling and Advanced Mathematics Population B

Country	In School	In Advanced Mathematics
United States	82	13
Sweden	24	12
Scotland	43	18
Ontario	33	19
New Zealand	17	11
Japan	92	12
Israel	60	06
Hungary	50	50
Finland	59	15
England	17	06
British Columbia	82	30
Belgium Flemish	65	9.5



The figures above represent the percentages of the cohort that are still in school and that which is taking advanced mathematics courses. The average age, number of years of schooling, and definition of advanced mathematics vary among educational systems. Students in Ontario are about one year older and are enrolled in grade 13, for instance. Although in most systems calculus is the standard content, that differs as well. In the United States, for example, only about 2 percent of the cohort is enrolled in a calculus course; less than one-half of them are enrolled in advanced placement calculus.

The estimated percent of the cohort still in school ranges from a high value of over 90 in Japan to lows of 17 in New Zealand and England. The percent taking advanced mathematics courses ranges from a high of 50 percent in Hungary to lows of 6 percent in Israel and New Zealand.

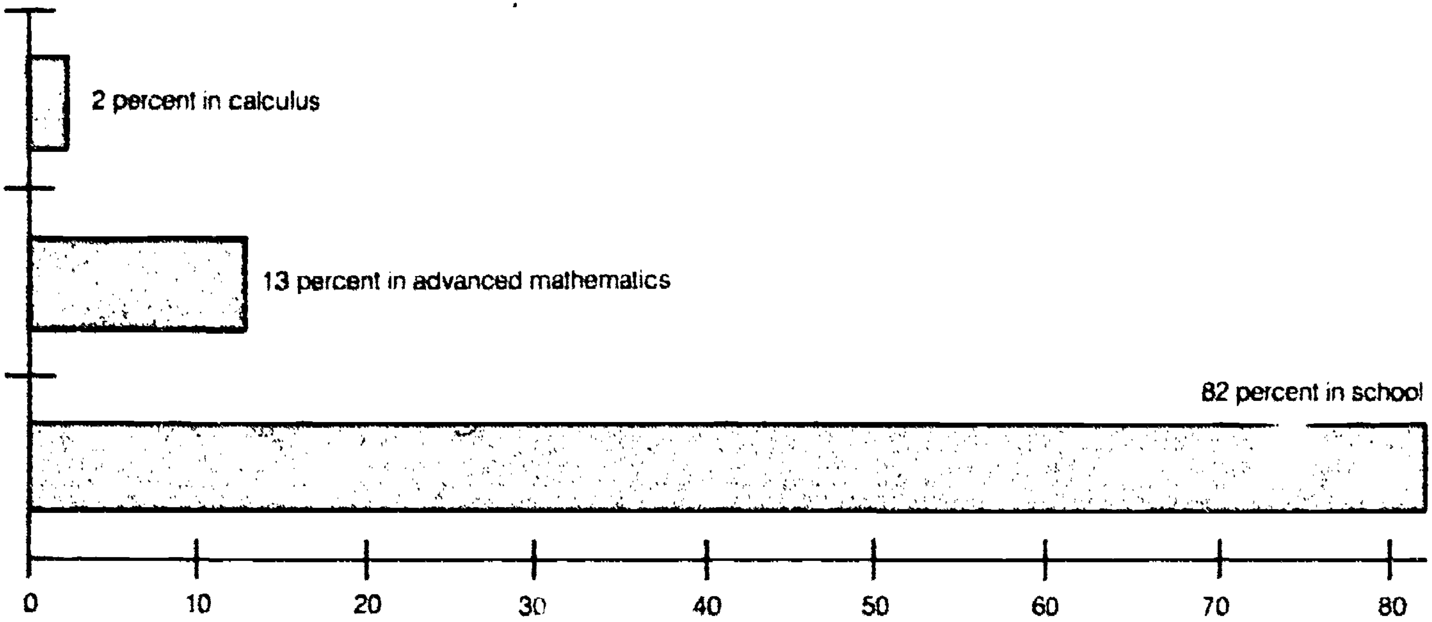
The Hungarian Example. While most systems are very selective at this level, Hungary is a striking exception. Although "only" 50 percent of the cohort are still in school, all of them take advanced mathematics. This system apparently is telling students that advanced mathematical knowledge is so important that every student must have it.

Miller and Linn (1985) examined achievement patterns in light of different retention rates and report two things that are relevant. First, although the average level of achievement for Hungary's students is close to the bottom among the systems, it is very close to the U.S. mean. Second, the top 1 percent and top 5 percent of Hungarian students perform near the top of the distribution of scores for the systems in the study. From a participation perspective, the Hungarians have it both ways. Not only do they provide advanced mathematical experiences to a large percentage of the cohort, and thereby increase dramatically the sum of mathematical knowledge in the culture, but they also do it without sacrificing the talents of their most capable students. As a model for both providing opportunity and creating a pool of talent, Hungary's bears scrutiny.

The U.S. Case. Since the U.S. case is practically the opposite of the Hungarian one, it again deserves scrutiny. The United States has a high retention rate and an average percentage of students taking advanced mathematics. The latter count, however, is misleading. The curriculum is differentiated at this level as well. Figure 5.7 shows how the U.S. stands when content areas are broken broadly into calculus and other courses. The results suggest that a smaller proportion of the cohort takes advanced mathematics (advanced mathematics is the calculus in virtually all systems) than in any other system. Since calculus is standard fare for the other systems, the U.S. percentage is really much lower than it appears. Calculus courses can be broken down even further to those considered advanced placement courses and those that are not. If the advanced placement calculus were considered life in the fast lane, then an extraordinarily small number participate in the best of schooling in the U.S.

Figure 5.7
U.S. 12th Grade Participation in Schooling and Mathematics

Percentages of the cohort in the United States enrolled in school, advanced mathematics classes, and calculus, respectively. The calculus courses can be divided further into calculus and advanced placement calculus. Courses other than calculus are composed of mixed content matter. Population B is defined as those students in the final year of secondary schooling who are enrolled in mathematics as an integral part of university preparation.



Early Tracking

How early and to what degree systems sort and track students is an issue of major consequences for what a system can produce. If findings about mathematics represent what happens in schools, early tracking of students has a profound effect on chances for many to be exposed to learning experiences offered to a tracked elite. By Grade 8 in the U.S., for instance, less than 15 percent of the students are in a track that will allow them, for instance, to take calculus in Grade 12, so there is no way that system can produce as much knowledge as do systems without early tracking. The practice of tracking so early effectively eliminates the possibility for most students to experience what is considered the best a school system has to offer.

Early tracking, and its consequences, appears from an IEA perspective to be an issue very similar to that of an earlier one of selectivity versus comprehensiveness. Elitism is less evident in a comprehensive organizational scheme with high retentivity. But it is no less questionable in a social arena where issues of achievement, opportunity, and participation are important.

Such issues, given the earlier caveats about IEA science and the assumptions upon which it is based, are most salient when one has the benefit of an international perspective. They can be addressed with the evidence provided by IEA surveys. People interested in change, reform, or innovation should continue their work with the benefit of evidence from the IEA experience.

Bibliography

Bloom, B.S. (1973). *Implications of the IEA Studies for Curriculum and Instruction.* Boston: Conference on Educational Achievement, Harvard University.

Burstein, Leigh, ed. (In press). *Student Growth and Classroom Process in Early Secondary School.* London: Pergamon Press.

Carroll, J.B. (1975). *The Teaching of French as a Foreign Language in Eight Countries.* Stockholm: Almqvist and Wiksell; New York: John Wiley.

Comber, L.C., and J.P. Keeves. (1973). *Science Education in Nineteen Countries.* Stockholm: Almqvist and Wiksell; New York: John Wiley.

Foshay, A.W., R.L. Thorndike, F. Hotyat, D.A. Pidgeon, and D.A. Walker (1962). *Educational Achievement of Thirteen-Year-Olds.* Hamburg: UNESCO Institute for Education.

Husén, T., ed. (1967). *International Study of Achievement in Mathematics (Vols. 1 and 2).* Stockholm: Almqvist and Wiksell; New York: John Wiley.

Kifer, Edward. (1984). "Issues and Implications of Differentiated Curriculum in the Eighth Grade." Paper delivered at the Atherton Conference, National Conference on Teaching and Learning of Mathematics in the United States. Monticello, Ill.

Kifer, Edward. (In press). "Opportunities, Talents, and Participation." In *Second International Mathematics Study: Student Growth and Classroom Process in the Lower Secondary Schools*, edited by Leigh Burstein.

Kifer, Edward, and R. Wolfe. (1985). "Classroom Processes and Patterns of Growth." Paper delivered at the National Conference on Classroom Processes in Secondary School Mathematics in the United States. University of Illinois at Urbana-Champaign.

Lewis, G.E., and C.E. Massad. (1975). *The Teaching of English as a Foreign Language in Ten Countries*. Stockholm: Almqvist and Wiksell; New York: John Wiley.

McKnight, C., F.J. Crosswhite, J.A. Dossey, E. Kifer, J.O. Swafford, K.J. Travers, and T.J. Cooney. (1987). *The Underachieving Curriculum*. Champaign, Ill.: Stipes Publishing.

Miller, M.D., and R.L. Linn. (1985). "Cross-national Achievement with Differential Retention Rates." Mimeograph copy. Urbana, Ill.

Pidgeon, D.A. (1962). "A Comparative Study of the Dispersions of Test Scores." In *Educational Achievements of Thirteen-Year-Olds in Twelve Countries*, edited by A.W. Foshay, R.L. Thorndike, F. Hotyat, D.A. Pidgeon, and D. Walker. Hamburg: UNESCO Institute for Education.

Purves, A. (1973). *Literature Education in Ten Countries*. Stockholm: Almqvist and Wiksell; New York: John Wiley.

Torney, J.V., A.N. Oppenheim, and R.F. Farnen. (1975). *Civic Education in Ten Countries*. Stockholm: Almqvist and Wiksell; New York: John Wiley.

Thorndike, R.L. (1973). *Reading Comprehension Education in Fifteen Countries: An Empirical Study*. Stockholm: Almqvist and Wiksell; New York: John Wiley.

What IEA Studies Say about Teachers and Teaching

LORIN W. ANDERSON
T. NEVILLE POSTLETHWAITE

From the earliest IEA studies to the present ones, researchers have been interested not only in describing student achievement and the differences in achievement both within and across countries, but also in understanding the reasons behind these differences. Their search has led them to consider the influence of teachers and teaching practices on student achievement.

In the following discussion, we will draw heavily from three generations of IEA studies: the first mathematics study (Husén 1967), the six-subject surveys (Peaker 1975, Walker 1976), and the Classroom Environment Study (Anderson 1987; Anderson, Ryan, and Shapiro 1989). Although the initial studies were surveys, the Classroom Environment Study relied extensively on classroom observations to answer two major questions. First, how similar are classroom conditions and teaching practices in countries throughout the world? Second, which, if any, of these conditions and practices are associated with higher student achievement and more positive student attitudes? Furthermore, while United States students were involved in the first series of studies, no United States classrooms were observed during the Classroom Environment Study. As a consequence, American readers may be able to "distance" themselves from the results without feeling a need to explain or justify them.

Lorin W. Anderson is Research Professor of Education, University of South Carolina, Columbia. T. Neville Postlethwaite is Professor of Comparative Education, University of Hamburg, West Germany.

What has emerged from these studies is a set of generalizations about teachers and teaching that hold promise for understanding and improvement. These generalizations apply to students, teachers, and classrooms "on the average," rather than to any particular student, teacher, or classroom. Exceptions quite definitely exist and may, in fact, abound. To place our generalizations within a proper context, we will first describe what IEA studies say about the relative contributions of home and school to student achievement.

Home, School, and Student Achievement

In combination, the results of the IEA studies support the well-known fact that students' backgrounds exert a powerful influence on their achievement. The studies, however, go beyond mere replication of this and provide several explanations for it.

First, far greater differences exist in the homes in which students are reared than in the schools they attend. As Postlethwaite (1987) asserts:

> After all, teaching is not like parenthood. There are no entry qualifications to become a parent, while in many countries there are strict entry qualifications to join the teaching profession. Consequently, although teacher differences often look large to students or their parents, these differences are, in fact, small compared with differences between homes. Thus, with little variability between teachers, it will be more difficult to detect teacher effects compared with [home background] effects (p. 1).

Second, the impact of the home on achievement is often amplified because students from different backgrounds are sent to different types of schools (e.g., private versus public comprehensive). In several of the countries included in the Classroom Environment Study, the type of school that students attended was strongly related to their achievement. In those countries in which this finding was not supported, only one type of school was typically included.

Third, the impact of home on achievement is also amplified by differences in the aspirations students bring to school. Students from middle- to upper-class homes tend to perceive education to be of great value and seek more of it. As a consequence, they tend to enter schools and classrooms better prepared academically and expend greater effort in learning. These initial student differences are strongly associated with differences in overall achievement (Anderson et al. 1989).

Finally, the impact of home on achievement is different for different subjects. The relationship between background and achievement appears to be strongest for subjects such as reading comprehension, literature, science, and the cognitive aspects of civic education. This relationship is much weaker for foreign languages and the affective aspects of civic education (Walker 1976).

While the IEA studies shed some light on the reasons for the strength of the relationship between home and achievement, they yield no evidence that would deny that relationship. Home background, either directly, or indirectly through its effect on aspirations and initial achievement, influences student achievement in most subjects. Furthermore, because of the strength of this relationship, the impact of schooling variables (that is, those pertaining to teachers and teaching) is likely to be quite small in studies, such as IEA's, in which students from a wide variety of backgrounds are included. As a consequence, when interpreting these results, one must remember that associations between teachers, teaching, and student achievement that are apparently small in magnitude may be quite significant. With this caveat in mind, let us now turn to our major generalizations.

Generalization 1—Greater teacher experience and subject matter knowledge are associated with greater skills in classroom management and higher levels of student achievement.

One of the few findings consistent across all countries in the Classroom Environment Study was that experienced teachers spent less time on classroom management. Less time in classroom management, in turn, resulted in greater achievement by their students (Anderson 1987).

The relationship between teacher knowledge and student achievement was supported in several of the studies in the six-subject survey (Walker 1976). In the science study, for example, science *teachers* responsible for teaching 10- and 14-year-old students in one country had lower achievement than the 10- and 14-year-old science *students* in another country. As might be expected, the average student achievement in the first country was much lower than that in the second. Similarly, in the stud of French as a foreign language, student performance in reading, listening, writing, and speaking was strongly related to teachers' ratings of their own skills in each of these areas.

The finding concerning teaching experience is consistent with the developmental model of teachers proposed by Fuller and Brown (1975). In this model, less experienced teachers are concerned primarily with personal survival. As teachers become more experienced, they become more concerned with the acquisition and refinement of teaching techniques and, ultimately, with the development and learning of individual students. Therefore, novice teachers are likely to spend more of their time on classroom management. Without proper management, their very survival is threatened.

The finding concerning teacher knowledge is related to what Shulman (1986) refers to as the "missing paradigm problem." According to Shulman, the recent trend in research on teaching has been toward the identification of generic teaching behaviors and skills that transcend the boundaries of subject matter knowledge. Knowledge of a particular subject and the ways in which this knowledge can be transformed by the teacher to make it "educationally digestible" to students are major components of the missing paradigm. The relationship between teacher knowledge and student achievement is supported by several of the IEA studies. The relationship between teachers' ability to transform their knowledge into forms that students are most likely to understand and student achievement remains to be explored.

This generalization offers at least two implications for teachers and teaching. First, teachers with different levels of experience are likely to have different needs for inservice training in topics such as classroom management, instructional strategies, and the like. This points to the need to individualize inservice training. Second, increased attention should be paid to the subject matter knowledge that teachers bring to the classroom. This attention should focus on both the subject matter to be transmitted and the ways in which it can be transmitted in order to ensure high levels of achievement by large numbers of students.

Generalization 2—Students who have greater opportunities to learn the knowledge and skills included on the achievement tests are likely to have higher levels of achievement.

This rather commonsense statement was supported in the majority of countries in all three sets of IEA studies. In fact, opportunity to learn was one of the first schooling variables included in the IEA studies (Husén 1967). "Opportunity to learn" was defined as teachers' ratings of the extent to which their students had received instruction in the knowledge or skill associated with each item on the achievement test. More specifically, the opportunity-to-learn index for each class of students was the percentage of items that tested knowledge or skills that the teacher believed he or she had taught. The higher the percentage, the greater the opportunity to learn.

In the Classroom Environment Study, we made a second estimate of opportunity to learn. Mandeville (1984) suggested that students in a classroom could be said to have had an opportunity to learn the knowledge and skills tested by an item if five percent more of the students answered the item correctly on the post-test than did on the pretest. This index of opportunity to learn (referred to as the five-percent rule) was the percentage of items for which the increase in the percentage of correct answers from pretest to post-test exceeded five percent. As might be expected, the estimate of opportunity to learn provided by teachers (termed the "judgmental estimate") typically was higher than the estimate based on the five-percent rule (termed the "empirical estimate"). The true opportunity to learn is likely to be somewhere between these two estimates.

The concept of opportunity to learn has at least two implications for teachers and teaching. First, opportunity to learn is a potentially powerful extraneous variable. That is, differences in the effectiveness of teachers and teaching practices are likely to be hidden by differences in opportunity to learn. Students having minimal opportunity to learn what is included on a test are likely to perform poorly on that test regardless of either the quality of the teachers or the quality of their teaching.

Second, and as consequence of the first implication, educators would be wise to attend to issues surrounding opportunity to learn as they struggle to understand and ultimately improve the quality of teachers and teaching. In particular, the concept of "alignment" is relevant. Stated simply, alignment is the extent to which the goals to be achieved, tests to be administered, and instruction to be provided are consistent with one another. In order to maximize alignment, one typically begins by identifying and holding constant either the goals, tests, or instruction. Thus, one can have "goal-driven" alignment, "test-driven" alignment, or "instruction-driven" alignment. As educators, we all have our personal preferences concerning the proper starting point of these alignment efforts. The Tyler rationale (1950), for example, emphasizes the need to begin with clear goals. Many experienced teachers, however, tend to begin either with textbooks or instructional activities (Clark and Peterson 1985). Regardless of the starting point, however, alignment needs to be addressed. Increases in alignment will quite likely result in increases in opportunity to learn which, in turn, are likely to affect student achievement.

Generalization 3—The greater the amount of time spent on teaching and learning, the greater the achievement of students.

At first blush, this generalization seems as "commonsensical" as the previous one. However, it is far more complex. For example, the IEA studies produced no evidence that increasing the *number of hours* per year spent on the study of a subject produces an increase in achievement. On the other hand, the *number of years* students spend studying a subject such as French has been associated with proficiency in that subject (Carroll 1975).

The amount of time spent on learning may be extended by homework. In fact, homework was one of the few variables in the six-subject survey that was consistently and positively related to student achievement, even after differences in home background, type of school, and other variables were taken into consideration (Peaker 1975). Pacing, defined as the amount of content covered per unit of time (e.g., per day, per month, per year), is another component of instructional time. Much research to date has focused on factors that influence the pace of instruction, such as the organization and length of the textbook or the aptitudes of groups of students, and some of it suggests that a more rapid pace is associated with higher levels of achievement (Barr 1987).

Teachers in the Classroom Environment Study were asked before each observed lesson whether new content would be presented or previously taught content reviewed. In most countries, the majority of the lessons focused on review of previously taught content, and, in most of these countries, average gains between pre- and post-tests were modest. Time spent on review, therefore, slows the pace of instruction. If new content were presented every other lesson or every third lesson, modest gains in achievement could be expected.

Perhaps the component of time most frequently studied during the past two decades has been student engaged time or time-on-task. In the Classroom Environment Study, the correlation between time-on-task and student achievement was essentially zero in the vast majority of the countries. However, when differences in class ability and opportunity to learn were taken into consideration, a weak but consistently positive relationship between time-on-task and student achievement appeared. Apparently the relationship of time-on-task with student achievement is easily masked when other, more potent, variables are included but not controlled in the study.

The major implication to be drawn from these findings is that educators should resist the simplistic, linear, "more is better" view of instructional and learning time that has been so pervasive in recent years. Adding minutes to the school day, adding hours to the school year, and finding ways of keeping students "on-task" at all costs are not likely to be a wise application of these findings. While adding years of study in a particular field is likely to pay dividends in student achievement, one must add a sufficient number of years. Based on the results of the study of French as a foreign language, Carroll (1975) concluded that six or seven years of study are needed to acquire satisfactory proficiency in a foreign language. Adding one more year to an already skeletal foreign language curriculum is not likely to pay off.

Similar arguments can be made for pacing and homework. The results of the Classroom Environment Study suggest that the pacing of most teachers in most countries is quite slow, due in part to the large number of review lessons. While some increase in pacing seems appropriate, one must avoid the tendency to move to the opposite extreme. Too fast a pace is likely to interfere with, not enhance, student achievement (Barr 1987). The relationship between pacing and achievement is not likely to be linear: one more lesson per week devoted to new content resulting in one more item correct on the end-of-year achievement test. Nonetheless, achieving an optimal rate of pacing seems a worthy goal.

Similarly, to facilitate student achievement, homework tends to be superior to no homework. This finding does not justify a call for more homework (except in those circumstances in which homework is not being assigned)! Rather, the purpose and difficulty of the homework (including length), and whether it is graded by the teacher (Walberg 1984) must be addressed before this research finding can be translated into a sound and defensible practice.

Generalization 4—Differences in teachers' behaviors are unrelated to differences in student achievement.

An earlier version of this generalization was more succinct but inaccurate. It stated simply that teacher behavior was unrelated to student achievement. As our colleague, Michael Dunkin of the University of Sydney, was quick to point out, this generalization is not in fact supported by the data gathered during the Classroom Environment Study. The study focused on *differences* in teacher behavior, rather than teacher behavior *per se*. The distinction between these two versions of the generalization is far more than semantic.

The statement that teacher behavior is unrelated to student achievement implies that what teachers do in their classrooms does not influence their students' achievement, that teachers do not make a difference in the academic lives of their students. This may or may not be the case; however, the IEA studies shed no light on this issue. Rather, the Classroom Environment Study attempted to determine whether differences in the behaviors noted by observers were associated with differences in student achievement. Immediately, then, behaviors on which teachers varied little, if any, were removed from consideration. If there were no differences among teachers on certain behaviors, they could not be related to differences in student achievement.

As a consequence, two groups of behaviors were excluded. The first consisted of behaviors exhibited fairly frequently by large numbers of teachers in virtually all countries. These included giving explanations to large groups of students, asking questions that required students to recall what they had been taught, monitoring or supervising students as they worked at their desks or tables, and attending to housekeeping matters (e.g., taking attendance, distributing papers).

The second group contained behaviors rarely exhibited by any of the teachers in any of the countries. These consisted of using examples to concretize an idea or abstraction, asking questions that require students to think or express and defend their opinions, checking to see whether students understand what they are being told or taught, and stating that a student's response to a question is incorrect or inadequate.

The first group of behaviors includes those that define the whole-class, recitation-seatwork approach to classroom teaching. Apparently, teachers throughout the world rely heavily on this approach. The second group includes behaviors that hold promise for increasing student achievement, but whose use in the classroom occurs so seldom that their relationship to student achievement remains unknown. Herein lies the fallacy of using correlational studies to prescribe teaching practices.

In summary, then, teachers tend to rely on a limited number of behaviors as they seek to transmit knowledge and skills to their students. These behaviors, which for experienced teachers appear to be part of a well-defined repertoire of behaviors that they use as needed or desired, are apparently consistent with the current emphasis on whole-class instruction with the teacher assuming a directive role.

Once these behaviors were excluded, we analyzed the remaining categories further. During the Classroom Environment Study, each teacher was observed for several lessons, typically eight. In one of the earliest stages of analysis, the consistency of teachers' behavior across these lessons was examined. The results suggested that, despite the limited number of behaviors observed, teachers used them quite inconsistently from one lesson to the next. In fact, this inconsistency was so great that differences in behavior could be predicted more reliably from differences in the lessons than from differences in the teachers. This inconsistency implies either that teachers behave somewhat randomly from one lesson to the next or, more likely, that their behavior is consistent with the demands placed on them by each class of students, their subject matter or matters, the specific objectives they have for their students, and their school and classroom organization.

The results of the Classroom Environment Study, then, suggest that teachers (1) do not differ at all on many of the behaviors thought to influence student achievement, and (2) do not differ reliably on other behaviors. As a consequence, the finding that differences in teacher behavior are unrelated to differences in student achievement should not be surprising.

The Classroom Environment Study produced some evidence that differences in teacher behavior are somewhat related to differences in student behavior, particularly in the extent to which students spend their classroom time on-task. Data on student behavior, like the data on teacher behavior, were collected during each lesson. Thus, changes in teacher behavior were quite useful in examining the relationship between teacher behavior and student behavior. In most countries, for example, teachers who spent more time interacting with their students during seatwork and teachers who spent more time giving explanations had students who spent more of their time on-task. These findings make sense because students are likely to take their behavior cues from their teacher's behavior (e.g., stay on-task when teachers are physically near you or pay attention when they are explaining something important to you).

Two implications can be drawn from these findings. First, and one that may not be greeted with cheers in light of the current American educational *zeitgeist*, is that educators should stop looking for direct links between variation in teacher behavior and variation in student achievement. Far too many factors mitigate against such a direct relationship. At the same time, however, evidence supporting the relationship between teacher behavior and student behavior is more substantial and is consistent with the earlier finding that teachers tend to judge their effectiveness more on the basis of student behavior than student achievement (Jackson 1967).

The second implication, consistent with the first, is that an exclusively behavioral view of teaching is potentially misleading and quite likely counterproductive. Jackson (1986) states his position on this issue quite succinctly.

> There is no such thing as a behavioral definition of teaching and there never can be. We can never simply watch a person in action and be sure that something called teaching is going on. . . . Our attempt to say when a person is or is not teaching is always an act of interpretation. We are forever "readers" of human action, seeking to determine which "reading" is correct from among those possible (pp. 77-78).

Three different "readers" of teachers' actions were in classrooms included in the Classroom Environment Study: the observer, the teacher, and the student. Which of these "readers" provides the most accurate reading?

Generalization 5—Students' perceptions of what transpires in the classroom are more valid than those of trained observers and, perhaps, teachers.

Slightly more than a decade ago, Doyle (1978) differentiated among three paradigms for research on teacher effectiveness: the process-product paradigm (which posited a direct link between teacher behavior and student achievement), the mediating process paradigm (which suggested that the ways in which students understand and act on teaching influence their achievement), and the classroom ecology paradigm (which asserted that students attempt to negotiate a set of environmental demands as they seek to survive and prosper academically in the classroom). While the first paradigm gives the central role to teachers, the other two give the central role to students.

Several of the findings of the Classroom Environment Study support the central role of students in their own learning. For example, students who see their classrooms as businesslike and academically focused tend to spend more of their time on-task and learn more. This relationship was generally stronger than that between observers' estimates of the percentage of students who were on-task and their achievement.

Similarly, those students who report that their teachers provide the necessary structure for their learning (e.g., informing them of the goals and objectives, focusing their attention on relevant points or features, or helping them understand the steps they should follow to arrive at correct answers or solutions to problems) tend to have more positive attitudes toward the subject matter. In contrast, none of the behavioral indicators associated with teacher structuring on the observation instrument were consistently related to student achievement.

Some educators may suggest that few implications can be drawn fro m this generalization since the students' perceptions rather than the teachers' behaviors are more important in determining student achievement. After all, what can teachers do to influence student perceptions? We believe this to be a narrow interpretation of the generalization.

What this finding suggests to us is the need for teachers to constantly monitor their students' efforts, to provide feedback to students concerning the adequacy (or inadequacy) of their learning, and to work with them to correct their errors and misunderstandings. The role of feedback and correctives (Bloom 1968, Rosenshine and Stevens 1986) in student learning has long been recognized.

Teachers need to be sensitive to students' reactions to presentations and explanations. The extent to which a teacher's explanation is clear depends primarily on the students, not the viewpoint of an external observer. Similarly, teachers must supervise students working at their desks or tables, point out errors as they occur, provide suggestions for improvement, and, generally, encourage students to expend the time and effort needed to complete the task successfully. Just as observers and students are expected to "read" the behaviors of teachers, noting their intention and consequences, teachers are expected to "read" the behaviors of students, noting when and how to intervene, and when not to.

Recognizing the central role of students in their own learning does not in any way diminish the importance of teachers. It simply redefines the teacher's role. Instead of assuming that their presentations are clear and comprehensible, teachers need to verify this assumption with their students. Instead of assuming that assignments are appropriate and lead to the desired learning, teachers need to monitor their students' work, providing feedback and correctives as necessary. This generalization suggests that teachers should be less egocentric and more sensitive to the impact they have (or fail to have) on their students.

Conclusion

Recently, Berliner (1987) proposed a "simple theory of classroom instruction" that includes the following as its primary variables:

1. opportunity to learn
2. students' engagement in learning
3. students' experience of success
4. pacing
5. structuring
6. monitoring

IEA studies support including these variables in such a theory. We suggest, however, that a much-needed emphasis on the central role of the students in their own learning is missing from Berliner's theory. Each of the variables is actually a student, not a teacher, variable. Students need opportunities to learn the knowledge and skills included on important tests. They need to be actively engaged in learning, not passive. They need to experience a high rate of success, not failure. The pace at which the instruction is delivered needs to be adapted to their rate of learning—not too fast or too slow. Students need to know what is expected of them—what they are to do and what they are to learn. Finally, the quality of completed work must be examined. If standards are not met, students must be told of this fact and encouraged, with the teacher's guidance, to make modifications so that the standards will be achieved.

Some educators might object to the view of teaching and learning suggested by Berliner and supported by the results of the IEA studies, labeling it simplistic. We conc.. with Berliner, who said

> To such criticisms, I offer two replies. First, simple ideas may have considerable utility. Second, from extensive observations in classrooms, I have learned that commonsense knowledge does not always result in either common (that is, widespread) or sensible practices (pp 95-96).

The strength of the findings of the IEA studies resides not in their insights, but in their grounding in the everyday experiences of school administrators and teachers. Krathwohl (1985) suggests that research in the social and behavioral sciences, including education, has its greatest impact when it allows us to

> compare the knowledge claim against our own experience for consistency. Much behavioral science knowledge is tested against our own experience to see whether it "rings true" (p 8).

We contend that the knowledge provided by the IEA studies does, in fact, "ring true." The extent to which the "commonsense" knowledge provided by these studies becomes "commonplace" remains the responsibility of practitioners.

References

Anderson, L.W. (February 1987). "The Classroom Environment Study: Teaching for Learning." *Comparative Education Review* 31: 69-87.

Anderson, L.W., D.W. Ryan, and B.J. Shapiro. (1989). *The IEA Classroom Environment Study*. Oxford: Pergamon Press.

Barr, R. (1987). "Content Coverage." In *The International Encyclopedia of Teaching and Teacher Education*, edited by M.J. Dunkin. Oxford: Pergamon Press.

Berliner, D.C. (1987). "Simple Views of Effective Teaching and a Simple Theory of Classroom Instruction." In *Talks to Teachers*, edited by D.C. Berliner and B.V. Rosenshine. New York: Random House.

Bloom, B.S. (Spring 1968). "Learning for Mastery." *Evaluation Comment* 2: 1-12.

Carroll, J.B. (1975). *French as a Foreign Language in Eight Countries*. New York: John Wiley and Sons.

Clark, C.M., and P.L. Peterson. (1986). "Teachers' Thought Processes." In *Handbook of Research on Teaching*, 3rd ed., edited by M.C. Wittrock. New York: Macmillan.

Doyle, W. (1978). "Paradigms for Research on Teacher Effectiveness." In *Review of Research in Education*, vol. 5, edited by L.S. Shulman. Itasca, Ill.: F.E. Peacock Publishers.

Fuller, F.F., and O.H. Bown. (1975). "Becoming a Teacher." In *Teacher Education: The Seventy-fourth Yearbook of the National Society for the Study of Education*, part II, edited by K. Ryan. Chicago: University of Chicago Press.

Husén, T. ed. (1967). *International Study of Achievement in Mathematics*, Vols. 1 and 2. New York: John Wiley and Sons.

Jackson, P.W. (1967). *Life in Classrooms*. New York: Holt, Rinehart, and Winston.

Jackson, P.W. (1986). *The Practice of Teaching*. New York: Teachers College Press.

Krathwohl, D.R. (1985). *Social and Behavioral Science Research: A New Framework for Conceptualizing, Implementing, and Evaluating Research Studies*. San Francisco: Jossey-Bass.

Mandeville, G.K. (1984). "Reanalyzing Teaching Research Data: Problems and Promises." *Evaluation in Education*, 8: 153-166.

Peaker, G.F. (1975). *An Empirical Study of Education in Twenty-one Countries: A Technical Report*. New York: John Wiley and Sons.

Postlethwaite, T.N. (September 1987). "Teachers, Teaching, and Teacher Education: Some Hints from IEA Research." Paper delivered at the IEA General Assembly Meeting, New York.

Rosenshine, B.V., and R. Stevens. (1986). "Teaching Functions." In *Handbook of Research on Teaching*, 3rd ed., edited by M.C. Wittrock. New York: Macmillan.

Shulman, L.S. (February 1986). "Those Who Understand: Knowledge Growth in Teaching." *Educational Researcher*, 15: 4-14.

Tyler, R.W. (1950). *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press.

Walberg, H.J. (February 1984). "Improving the Productivity of America's Schools." *Educational Leadership*, 41: 19-27.

Walker, D.A. (1976). *The IEA Six-Subject Survey: An Empirical Study of Education in Twenty-one Countries*. New York: John Wiley and Sons.

The IEA Studies and Reforms in Reading, Writing, and Literacy

ALAN C. PURVES

Throughout its history, IEA has conducted three studies of education in the mother tongue: one in reading, one in literature, and one in written composition. Studies of foreign language learning (French and English) were also conducted, but I shall not cover those here. The first two surveys were completed in the early 1970s and the third in the mid-1980s. Their results offer ways to use comparative international data for charting the direction of curriculum reform.

The IEA Reading Study

The Reading Comprehension Study (Thorndike 1973, Purves et al. 1980) used a straightforward test of comprehension of passages, followed by multiple-choice questions, plus reading speed and word knowledge tests. The instrument was administered to three populations: 10-year-olds, 14-year-olds, and students in the pre-university program.

The overall results show that United States students performed about as well as the students in most industrialized countries, but that New Zealand students stood out on the comprehension measure above all others. The various initial analyses of the reading comprehension data pointed clearly to the influence on achievement of the home back-

Alan C. Purves is Director of the Center for Writing and Literacy and Professor of Education and Humanities, The University of Albany, State University of New York.

ground variables, particularly home literacy variables. This influence was apparent both within and between countries.

Although these analyses indicated little of consequence for school reform, they prompted a number of subsequent studies, particularly those of Purves 1973, Purves et al., 1980, Baten 1981, and Guthrie 1981. The results of some of these studies supported IEA's initial analysis and suggested little could be done to raise a system's reading scores save a massive restructuring of society to make print literacy more predominant in the culture than it is, and that such an effort, even if possible, would fail because of the influence of electronic media. At the time of testing, New Zealand was, relatively speaking, television free and heavily print oriented. That is no longer the case only six years after the initial test.

A more promising line of inquiry for educational reform was to re-examine the passages and test items (Purves et al. 1980, Baten 1984). The results of this inquiry show that, for the 14-year-old population in most countries, certain passages are systematically more difficult. Their difficulty, however, can not be explained by the usual formulas, such as readability, but it can when the structural organization of a passage is considered.

For this age group, passages with other than a sequential or narrative structure are systematically more difficult. Passages with cause-and-effect or comparison-contrast structures are seldom available in school texts through junior high. These texts see much more frequent use in upper secondary schools and form the basis of academic writing in most of the industrialized and developing nations. Students with little previous experience of them appear to have great trouble reading them on a test.

The analysis of item types shows a similar result. Of the four types—meaning of parts, meaning of the whole passage, inference beyond the text, and analysis of rhetorical or structural devices—the fourth proved to be much more difficult for 14-year-olds in many of the countries in the study. Analysis of the textbooks and the curriculum show that, although questions of rhetorical device and structural analysis appeared in many standardized tests, they were not part of the instructional program for students until upper-secondary school. Again the students performed poorly because they did not have the opportunity to learn what was covered on the test, even though testmakers in many of the systems said that these were appropriate items and passages for the age group.

The findings of this study have since been supported by the work of many researchers in the United States and elsewhere and are included in the volume *Becoming a Nation of Readers*. Clearly the implications for curricular and pedagogical reform are present here, but whether textbooks used in the United States (especially the basal readers) have been revised to include nonnarrative material and questions that deal with rhetoric and structure has not become evident even in the more recent editions.

The IEA Literature Study

The IEA Study of Literature (Purves 1973) conducted in the early 1970s involved 10 educational systems: Flemish-speaking and French-speaking Belgium, Chile, England, Finland, Iran, Italy, New Zealand, Sweden, and the United States. The result was similar to that of the reading study. Again students from the United States were well within the average, and the best students in the last year of secondary school outperformed the students in all other countries.

Of the variety of literary works presented to the students, however, one proved much more difficult than the others. It was the one elliptical and metaphoric story, written in a style that might be termed poetic rather than literal. Examination of the textbooks shows that this sort of writing was virtually absent from the fare presented to students, particularly in the United States, through secondary school.

The analysis also shows that students, particularly at the 14-year-old level, did much more poorly on items that asked for the application of literary terms to the material than they did on items dealing with interpretation, either literal or inferential. Text analysis and data from teachers show that literary terminology is not taught these students even though it is considered an appropriate part of the curriculum. As in the reading study, the literature study shows that opportunity to learn is an important factor in student achievement and is clearly an area where educational reform can occur.

In a recent reanalysis of tests in reading and literature in the United States (Brody, DeMilo, and Purves 1988), unfortunately little change appears in test items and passages. Nonnarrative and figurative passages are still lacking, and items dealing with the application of analytic terms and concepts from rhetoric or literary study are in the minority. A cursory look at the textbooks again shows that the majority of texts contain literal narratives and the terms used for analyzing style and structure are not emphasized in the teaching-learning apparatus surrounding the texts. Opportunity to learn these more complex aspects of writing is still nonexistent for most students in the United States.

The literature study also examined the preferred responses of students to literary texts (e.g., what students think is important to write about). On the basis of a content analysis of each sentence in compositions written by over 1000 students and teachers in response to a variety of short stories and poems (Purves and Rippere 1968), 20 core issues emerged concerning literary texts that could be framed as questions. A larger sample of students in each country were then asked to select the five of these questions that they thought most important to raise in writing or talking about a short story. Each student made three responses—one each following the reading of two stories and one in a questionnaire about fiction in general. From these we derived a picture of what a student sees as important in the abstract and after reading each of two stories. The questions were rated by their teachers and by a group of experts on literature and its pedagogy in each country.

The results show that, as students progress through secondary school, they become more consistent in their responses across texts and thus come to approach all the fiction they read from a single perspective. Students in any one country tend to have increasingly similar responses as they progress through school. This is not to say that the different stories had no effect on the students' responses, but that in each of the systems of education studied, the effect of the text was less strong on the older students than it was on the younger ones.

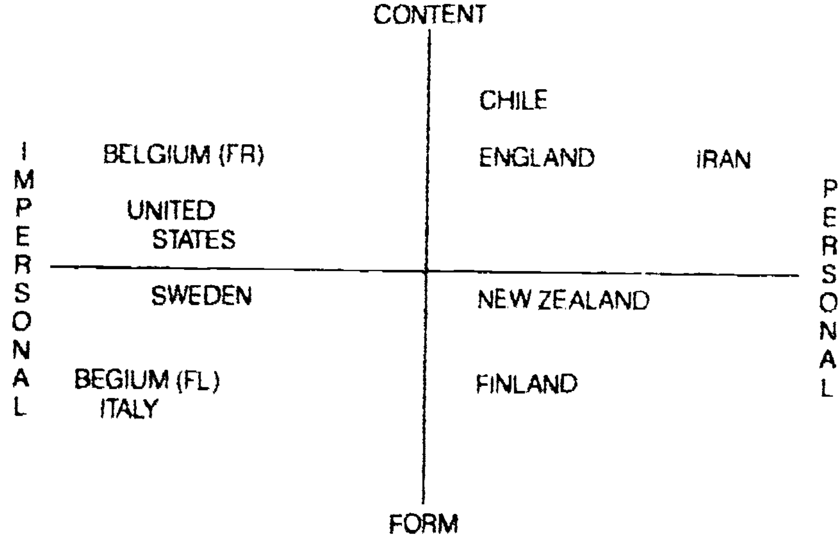
In terms of national differences, the results show that there are both an international interpretive community and national interpretive communities. Students in all countries generally rejected the questions "Is this a proper subject for a story?" and "Is anyone in this story like the people you know?" They generally found attractive the questions "Has anything in this story a hidden meaning?" and "What happens in the story?"

Aside from these general rejections and predilections, students in different countries exhibit sharply different profiles of response. Two sets of questions form the coordinates on which we plotted the major differences between and among countries (Figure 7.1). They are, first, a "personal-impersonal" continuum and, second, a "form-content" continuum. Students in Belgium and Italy emphasize the impersonal and the formal; in Chile, England, and Iran, the personal and the content-oriented response; and United States students are concerned with content, but not from a personal point of view.

Responses to reading then, although partially dependent on the particular text read, appear more crucially dependent on the culture of the reader, the reader's membership in a community of readers. Such dependency increases as people progress further through their educational system. Students' responses become increasingly like the responses preferred by teachers and, to a lesser extent, experts (critics, curriculum makers, and teacher trainers whom we polled). For the younger students in the IEA Literature Study (age 14), the average rank-order correlation between the question preferred by student and teacher was .35; for the older students (in the pre-university year) it was .54. The teacher, therefore, serves both as a major force in influencing how students respond to a text and as the individual who inducts students into the interpretive community of a particular society.

The question this result raises for educational reform is whether the limited and rigid pattern of response that students learn is the most appropriate or desirable to take into adult life. Reform documents call for critical reading and varied approaches. United States students learn

Figure 7.1
Position of Countries Relative to Continuums of Personal-Impersonal and Form-Content Responses



that each literary text has a hidden symbolic meaning containing a moral or lesson. Such an approach to reading has consequences beyond the literature curriculum, making it difficult for people too explore reading material that does not fit the limitations of the pattern. At an extreme, it leads to censorship and book burning. Is that truly the goal of the literature curriculum of American schools?

The IEA Written Composition Study

The IEA Study of Written Composition currently involves students who were tested in 1985 in the following countries: Chile, England, Hamburg (The Federal Republic of Germany), Finland, Hungary, Indonesia, Italy, Netherlands, Nigeria, New Zealand, Sweden, Thailand, the United States, and Wales. Three populations participated: students at the end of compulsory education, roughly the ages of 15 to 17 and the main focus of this study; students at the end of primary education (grade 12); and students at the end of the pre-university year. Because the results are still being analyzed, comments made at this point may be modified by future findings.

The students wrote on parallel tasks and the scoring used a common scheme and scale across languages and cultures. As a result, certain comparisons can be made but it is impossible to make a direct comparison of the writing performance of students in different countries, in part because of language differences and in part because of the problems of scoring compositions internationally.

In the main population (ages 15-17), each student wrote on three tasks, one a functional task, one a descriptive-expository letter, and one either an argument, a narrative, or a reflective essay. The relative performance across tasks was not uniform. Although the narrative was the easiest and the argument or reflective essay the most difficult in eleven of the systems, this was not the case universally. In three systems, the reflective or argumentative task was relatively easy. These were the systems in which such types of composition had replaced the narrative as the dominant form.

Another initial finding shows that, in systems with streaming or tracking, students in lower tracks, as might be expected, come from homes where literacy practices are few. In class, they write less and receive less instruction in writing. These students tend to perceive success in writing in terms of handwriting, neatness, spelling, and grammar rather than structure, style, or content. Such a result, combined with the first, clearly suggests the importance of both classroom opportunity to learn and the curriculum to ensure that they have a chance to learn and practice what they are measured on. One can attack teaching to the test, but shouldn't one also attack testing the untaught?

Another aspect of the study is exploring the extent to which students are taught to be members of rhetorical communities. It is examining the criteria used by teachers of writing to see if there are systematic differences that might help define rhetorical communities, and it is examining the perceptions of students about what constitutes "good" school writing.

The hypothesis underlying this exploration is that students learn to write according to certain conventions, many of which have little to do with the structure of the language and more to do with the literary and cultural heritage of the society. That is to say, many aspects of texts are not bound by the morphology and grammar of a particular language, but by custom and convention. Such is the line of reasoning behind *contrastive rhetoric*.

As an initial step, we gathered a number of compositions written by approximately 100 students at or near the end of secondary school in several of the countries in the study and three countries no longer in the study (Australia, Japan, and Israel). The compositions were on "My Native Town" and "What is a Friend?", both thought to be relatively neutral topics that would not force a particular pattern of organization or style on the students. A team of three researchers then examined the compositions and found they tended to differ systematically by culture of origin along a number of continuums, some of which matched those of earlier researchers, particularly Carroll (1960), Kaplan (1966), Glenn and Glenn (1981), and Hofstede (1980). They are the following:

PERSONAL-IMPERSONAL depends primarily on the frequency of references to the writer's thoughts and feelings about the subject.

ORNAMENTED-PLAIN may also be defined as "figurative-literal" and depends on the frequency of use of metaphor, imagery, and other figures of speech.

ABSTRACT-CONCRETE refers to the amount of specific information and detailed references in the text, as well as to the general level of abstraction.

SINGLE-MULTIPLE refers to whether the text focuses on one aspect of the subject or tries to cover a large number.

PROPOSITIONAL-APPOSITIONAL: This continuum, which is similar to Glenn and Glenn's abstractive-associative and some of Kaplan's diagrams, refers to the types of connectives that hold the text together and whether the composition is developed by a standard order. A comparison-contrast composition, for example, would be propositional. An oppositional composition uses few connectives besides *and* or *but* and often omits cohesive ties other than idioms and repetitions. The continuum from appositional to propositional is one that moves from stream of consciousness through narration and description to the syllogistic proof.

These characteristics appear to distinguish the writing of students in different cultures and to form some of the dimensions that delimit models of text in certain cultures (see Figure 7.2). These differences are not inherent in the language, but result from cultural learning because they occur between students writing in the same language (such as English) but living in different cultures. Because full analysis has not yet been completed on the main body of compositions, these comments must be seen as suggestive rather than definitive.

In the pilot phase of the study, teachers from the participating countries were asked to both rate and comment on a number of the compositions. Their comments were analyzed for content, followed by a factor analysis. Four "general merit factors" emerged apart from mechanics, spelling, and handwriting: (1) content, (2) organization and structure, (3) style and tone, and (4) personal response to the writer and the content (Purves 1984). These factors of judgment appeared in all countries, but the relative emphasis and interpretation varied from system to system and from type to type of composition.

To check the teachers' criteria, students were asked to write a letter of advice to people younger than they who were about to attend their schools. The letter was to suggest ways to succeed in school writing, and content analysis of the resulting compositions proved most revealing, although not necessarily complimentary to the schools. Students in all the countries noted that spelling, handwriting, grammar, and neatness were of paramount importance.

The letters were also analyzed for specific types of advice, and it was divided into the broad categories of content, organization, style and

Figure 7.2

Country	FACTOR Personal	Ornamental	Abstract	Single	Propositional
Australia	High	High	Low	High	Low
England	Medium	Low	Low	Low	Low
Federal Republic of Germany	High	Low	Low	Low	Low
Finland	Low	Low	Low	Low	Low
Israel	High	Medium	Low	High	High
Italy	High	High	High	High	High
Ivory Coast	Medium	Low	Low	Low	Low
Japan	High	Low	High	High	Medium
Netherlands	High	Low	Low	Low	Low
New Zealand	Low	Low	Medium	Low	Low
Nigeria	Low	Low	Low	Low	Low
Scotland	Low	Low	Low	Low	Low
Thailand	High	High	Medium	Medium	Low
United States	Low	Low	Low	Medium	High

tone (following the scoring scheme), and presentation, including grammar, orthography, layout and neatness, process and classroom tactics (Takala 1983).

Although the analysis for all countries is not complete, the results from a few of them strongly suggest that students within a country are aware of the existence of certain operating norms. Variation occurs within a country, of course, but specific patterns exist nationally and internationally. Figure 7.3 identifies these patterns.

This figure reveals the general perception of students that their teachers are preoccupied with matters of presentation. There are strong national differences in perception, however, such as the relatively low emphasis on organization in Chile and on style and tone in The Netherlands. In New Zealand and Sweden, teachers appear to emphasize process more than in other countries, but in Sweden more of this emphasis concerns choice of topic than is true in New Zealand. Differences occur within other categories as well. In The Netherlands, the emphasis under presentation is weighted more towards grammar than in the other countries, whereas in Chile, the balance tips heavily toward appearance and spelling. Under organization, a strong concern is shown in The Netherlands for using an organization set by the teachers, an issue of less importance in the other countries. In Sweden, teachers seem to favor simple sentences, a feature that does not figure much in the other countries. From this evidence, it appears, according to their students, that teachers favor particular ways of preparing and presenting written discourse.

Each of the three aspects of the Written Composition Study (the examination of compositions, teacher questionnaires concerning crite-

Figure 7.3
Summary Percentage Distributions of Student Advice in Five Countries

Category	Chile	England	Netherlands	New Zealand	Sweden
Content	13 3	14 7	9.2	20 5	13 9
Organization	4 1	15 0	17 0	14.5	12 9
Style and Tone	13.0	13.0	7 8	13 5	13.7
Presentation	49 1	40 0	47 5	31 3	33 9
Process	11.4	13.0	11.0	18 5	19 6
Tactics	1.1	2 0	1 4	1 6	0 0
Unclassifiable	1.3	0 0	6.1	0 0	0 3

ria, and the letter of advice) suggests that students in an educational system do indeed learn to become members of a rhetorical community, that they learn not only how to write, but also what aspects of their writing are valued in their society. At times, there is direct instruction; at times, it is implicit, but either way, the students learn that being able to write is being able to produce texts that match certain models and that these models serve as student and teacher criteria. If we put these findings together with the findings of the literature study, we can see that learning literacy in a society is not merely a matter of acquiring a set of skills. The implications for reform are clear. The educational community should determine whether the models of text and of reading and writing are what is desired and then the extent to which the testing and teaching programs (including teacher feedback on student writing) reinforce or undermine these goals.

References

Baten, L. (1981). "Text Comprehension. The Parameters of Difficulty in Narrative and Expository Prose Texts: A Redefinition of Readability." Unpublished doctoral diss., University of Illinois.

Brody, P., C. DeMilo, and A.C. Purves. (1988). *The Current State of Assessment in Literature.* Technical Report Number 1. Albany, N.Y.: Center for the Learning and Teaching of Literature.

Carroll, J. (1960). "Vectors of Prose Style." In *Style in Language*, edited by T.A. Sebeok. Cambridge, Mass. and New York: Technology Press and Wiley.

Glenn, E.S., and C.G. Glenn. (1981). *Man and Mankind: Conflict and Communication Between Cultures.* Norwood, N.J.: ABLEX.

Guthrie, J. (February 1981). "Reading in New Zealand: Achievement and Volume." *Reading Research Quarterly.* XVII: 6-27.

Hofstede, G. (1980). *Culture's Consequences.* Berkeley, Calif.: Sage.

Kaplan, R.B. (March 1966). "Cultural Thought Patterns in Intercultural Education." *Language Learning* 16: 1-20.

Purves, A. (1973). *Literature Education in Ten Countries. An Empirical Study: International Studies in Evaluation.* Stockholm: Almqvist and Wiksell.

Purves, A. (1984). "In Search of an Internationally Valid Scheme for Scoring Compositions." *Journal of College Composition and Communication* 35: 426-438.

Purves, A., B. Bauer, D. Harnisch, and D. Quirk. (1980). *Reading and Literature. United States Achievement in International Perspective.* Research Report No. 20. Urbana, Ill.: National Council of Teachers of English.

Purves, A., and U. Rippere. (1968). *The Elements of Writing about a Literary Work.* Research Monograph No. 10. Champaign, Ill.: National Council of Teachers of English.

Purves, A., and S. Takala, eds. (1982). *An International Perspective on the Evaluation of Written Composition: Evaluation in Education: An International Review Series*, Vol. 5, No. 3. Oxford: Pergamon Press.

Takala, S. (1983). "Achievement in Written Composition." (Unpublished manuscript). Urbana, Ill.: IEA Study of Written Composition.

Thorndike, R. (1973). *Reading Education in Sixteen Countries: An Empirical Study: International Studies in Evaluation*. Stockholm: Almqvist and Wiksell.

Science, Mathematics, and National Welfare: Retrospective and Prospective Achievements

HERBERT J. WALBERG

In *The Wealth of Nations*, Adam Smith (1776) argued that national welfare depends on ability. Given the importance of this often-assumed dependence, the direct evidence for it seems slight. In *A Nation at Risk*, for example, the National Commission for Excellence in Education (1983) warned that educational mediocrity jeopardizes U.S. prospects for growth in economic productivity and welfare, but cited no correlations of education and economic data. In view of current public and legislative concern, this study explores the relations of education and economic indicators.

Economic indicators are hardly the only criteria for evaluating the quality of life or national welfare. Income, however, substantially determines the amount and quality of domestic and foreign goods and services that people can buy. Moreover, even small differences in sustained economic growth cause large differences in national standings.

From 1870 to 1950, for example, as school and university enrollments expanded in percentages of eligible age groups in the U.S., the gross national product per worker grew 0.7 percent faster than that of Europe as a whole. This small advantage in growth sufficed to place the U.S. first in per capita income and economic growth by mid-century (Schultz 1981).

Herbert J. Walberg is an educational psychologist and Research Professor of Education, University of Illinois, Chicago.

As reported below, growth in income differs widely among nations. Such differences, if sustained, will cause vast changes in national wealth rankings. At 1977–1982 U.S. and European growth rates of around 1.5 percent, for example, income would take 47 years to double. At the Japanese rate of 4.4, it would take 16 years. These figures may underestimate ultimate differences since Japan's income now exceeds that of the U.S. and was growing at the rate of 11.3 percent in the first quarter of 1988 (ba 1988, p. 29).

Education Indicators

In considering how to measure the possible education-growth connection, what should be chosen as education indicators is not immediately obvious. National agencies (for example, the U.S. Department of Education 1988) report quantitative indexes of input, process, and outcome such as per-student expenditures, enrollments of age-eligible groups, and mean years of education in adult populations. Yet the National Commission, other reform groups, and legislators want quality indicators such as results on achievement tests.

Such insistence may not be without merit. Per-student expenditures, for example, appear unlinked to achievement gains in an analytic review of more than 150 studies (Hanushek 1986). Moreover, despite poor science and mathematics performance, the U.S. recently spent a greater fraction of gross national product and of total government expenditures on education than many other countries. At 18 percent, the U.S. spent a larger portion of its total tax dollars on education than any other country recently surveyed by he United Nations (U.S. Department of Education 1988, pp. 16-17, 35). Indeed, considerably more than the second and third ranked countries, Norway and Sweden.

Nor can enrollments and average adult years of education be taken as valid indicators of results. In 240 days per school year (National Commission 1983), for example, Japanese students earn in 12 years the time equivalent of 16 years of U.S. students' 180-day year. This is aside from other possible Japanese advantages such as a rigorous national curriculum and educational encouragement at home (U.S. Department of Education 1987, p. 79).

In view of the $850 billion educational investment embodied in the U.S. labor forc from 1900 to 1970 (Schultz 1981), it seems ironic that nations have such spotty and noncomparable indicators of human capital in contrast to the many well-standardized indexes of labo. and physical and financial capital—many available in modern countries monthly over decades (see Becker 1976 and Mincer 1981, on the concept of human capital). Several national and international organizations, however, are beginning to collect comparable information over time and place. These include the National Assessment of Educational Progress, similar programs in other countries, and the International Association for the Evaluation of Educational Achievement (IEA).

The Paris-based Organization for Economic and Cooperative Development, moreover, inaugurated a large-scale indicator program in mid-1988 that will enable scholars and policymakers to compare the progress of countries with respect to educational inputs, processes, and outcomes (U.S. Department of State 1988). Thus, we will be able within a few years to relate educational investments and practices to accomplishments in the hope of improving policies and investigate the possible contributions of accomplishments to national income and welfare to suggest allocations of educational investments.

In the meantime, this study explores the second possibility by relating indicators of educational outcomes to growth in national income and other welfare or quality-of-life indicators. These data were compiled from indicators collected by staff researchers of the (London) *Economist* (1983) and in three IEA studies of science and mathematics achievement (Comber and Keeves 1973; McKnight, Crosswhite, Dossey, Kifer, Swafford, Travers, and Cooney 1987; and IEA 1988).

Level and Growth of Income

After World War II, the U.S. was first in national income and economic prowess, but other countries caught up and surpassed us. As late as 1965, for example, General Motors earned twice as much as the combined totals of the top 30 German and top 30 Japanese industrial companies, but in 1980 GM reported its first deficit since the Great Depression and in 1982 announced that it would import and sell Japanese cars. From 1965 to 1980, the U.S. share of companies with world sales over $1 billion declined from 70 to 40 percent (Freund 1982).

By 1981, as shown in Figure 8.1, the U.S. slipped to fourth place in per capita income and ranked in the bottom half in economic growth among 16 nations with comparable data. By 1988, the dollar had lost substantial value in relation to other currencies; therefore, the current U.S. rank is now considerably lower.

Figures 8.1 and 8.2 show that high-income countries grew slowly. The low-income countries—Brazil, Hungary, and Mexico—averaged 4.3 percent growth, and the high-income countries—Switzerland, Sweden, and Germany—averaged only about 1.5 percent. Simple regression shows that each additional $1,000 in income is associated with an average drop of 0.25 percent in growth (see Figure 8.2 for bivariate correlations and significance levels).

The important exceptions are the Atlantic nations—Britain and Spain—with relatively inferior income and growth (below the fitted regression line) and the Pacific nations—Japan and Singapore—which grew much faster than their incomes would predict. Japan remains unusual since its recent growth has been extremely high, 11 percent, despite being first or near first in per-capita income since the near doubling of its national currency in relation to the dollar.

As Kennedy (1987) argued in his widely noted history, *The Rise and Fall of the Great Powers: 1500-2000*, the mighty Western nations (Spain, Portugal, Holland, England, France, Germany, and the U.S.) have risen and fallen in wealth and power in relation to one another. Despite its economic ascendancy in the first half of the century, the U.S. in 1981 was 26 percent below Switzerland in per-capita income and 64 percent below Japan in economic growth (Figure 8.1). What factors account for economic growth and national welfare?

Creative Elites?

Among philosophers who attribute primacy to elite aristocracies or oligarchies, Plato believed that philosopher-kings brought out the best in a Greek city-state. Among historians of the rise and fall of civilizations, Arnold Toynbee (1959) held a "great person" theory of national development. In his voluminous *A Study of History*, Toynbee hypothesized that national ascendancy is the result of the emergence of a "creative minority" that leads the masses to greatness.

One possible world index of such leadership in modern times is the Nobel Prize, given in recognition of outstanding leadership in the natural sciences, economics, medicine, literature, and the pursuit of peace. However, the association between Nobel prizes and economic growth contradicts the great-person theory. As shown in Figure 8.1, the nations that have accumulated the highest number of Nobel prizes per 10 million population—Sweden, Switzerland, and Britain—grew

Figure 8.1
Economic, Educational, and Cultural Indexes of 16 Nations

Nation	Income	Growth	Unemploy-ment	Nobels	College	Tests	Work Hours	Divorce	TV
Switzerland	17,430	1.5	0.4	21.6	17		44.1	31	31.4
Sweden	14,870	1.3	3.1	32.4	37	22	35	53	38.1
Germany	13,450	1.6	6.1	9.5	26	24	40.7	9	33.7
U.S.A.	12,820	1.6	9.5	6.9	55	22	34.8	50	62.4
France	12,190	2.0	8.0	0.7	25		39.8	22	35.4
Canada	11,4[illegible]0	1.0	10.9	1.3	36		38.2	32	47.1
Australia	11,080	2.4	7.1	2.7	26	25	35.0	44	37.8
Japan	10,080	4.4	2.4	0.3	30	31	40.2	18	53.9
Britain	9,110	0.5	12.5	13.9	20	21	43.0	36	40.4
Italy	6,960	2.2	8.9	1.9	27	19	38.8	4	38.6
Spain	5,640	0.9	15.9	0.3	22		43.3	0	25.2
Singapore	5,240	9.6	6.7	0.1	8		48.8	1	16.6
Israel	5,160	3.6	5.1	2.5	26		35.4	14	15.0
Mexico	2,250	6.6	4.2	0.1	15		45.0	5	10.4
Brazil	2,220	3.9	6.9	0.0	3		46.0	7	12.2
Hungary	2,100	2.5	1.0	2.7	13	29	40.0	32	25.8

Income: Gross domestic product (GDP) per capita, 1981
Growth: GDP average annual growth rate percent, 1977-82
Unemployment: Unemployed percent of labor force, 1982
Nobels: Number of prizes per 10 million population, 1901-1982
College: Percent of 20- to 24-year-olds enrolled in higher education, 1979
Tests: Science test scores of 14-year-olds, 1970-71
Work Hours: Average hours worked per week (nonagricultural), 1982
Divorce: Percent of marriages ending in divorce, 1979
TV: Number of television sets per 100 population, 1980

Note. Only 9 of the 16 countries participated in the 1970-71 science survey of the International Association for the Evaluation of Educational Achievement.
Sources: Economist, December 24, 1983, Walberg 1983

Figure 8.2
Correlations of Economic, Educational, and Cultural Indicators

	Income	Growth	Unemploy-ment	Nobels	College	Tests	Work Hours	Divorce	TV
Income	1.00	−0.51	−0.09	0.64	0.58	−0.30	−0.41	0.58	0.65
Growth	−0.51	1.00	−0.30	−0.41	−0.48	0.77	0.52	−0.50	−0.51
Unemployment	−0.09	−0.30	1.00	−0.29	0.19	−0.74	0.01	−0.16	0.16
Nobels	0.64	−0.41	−0.29	1.00	0.25	−0.39	−0.21	0.59	0.21
College	0.58	−0.48	0.19	0.25	1.00	−0.29	−0.80	0.58	0.81
Tests	−0.30	0.77	−0.74	−0.39	−0.29	1.00	0.16	−0.06	−0.03
Work Hours	−0.41	0.52	0.01	−0.21	−0.80	0.16	1.00	−0.63	−0.58
Divorce	0.58	−0.50	−0.16	0.59	0.58	−0.06	−0.63	1.00	0.61
TV	0.65	−0.51	−0.16	0.21	0.81	−0.03	−0.58	0.61	1.00

Note: Correlations of .50 or more are statistically significant at the .05 level, except those involving Tests, which require a correlation of .71 for .05 significance.

slowly, and those with few prizes—Singapore, Mexico, Brazil, and Japan—grew most rapidly.

The negative relation, earlier noted by Walberg (1983), has been exhaustively replicated with Nobel science and medicine prizes, mathematics prizes, and publication and citation rates for various periods by the Congressional Research Service (CRS 1986) for the U.S. House Committee on Science and Technology. Those who advocate spending for science to promote economic growth may overvalue discovery, undervalue technology and application, and fail to recognize that scientific information is rapidly disseminated around the world. If human capital is scarce, moreover, funds expended on discovery may not exploit it (Walberg 1983, CRS 1986).

Hardworking Elites?

Another index of the effect of the elite on a nation is the percentage of young adults enrolled in colleges and universities. This indicator also predicts slow economic growth (figures 8.1 and 8.2).

Each percentage point of additional enrollment was associated with a 0.1 percent drop in growth. Across the range of college enrollments, an increase from 3 percent in Brazil to 55 percent in the U.S. was associated with an estimated reduction in economic growth from about 7 to 1 percent. The U.S. had by far the highest fraction of 20- to 24-year-olds enrolled. Canada and Sweden also had high college enrollments and correspondingly low growth rates.

Larger enrollments in higher education, shown in the figures, are also associated with a shorter work week. Each additional percentage point in enrollment equals 0.3 hours (18 minutes) less work per week. Across the range of enrollments, from 3 percent in Brazil to 55 percent in the U.S., this amounts to an estimated 15-hour reduction in the work week from about 47 to 32 hours.

In the countries with high concentrations of college-educated adults, how is leisure time spent? College enrollments predict high rates of television ownership, up to 62 per 100 persons in the U.S. at the time the data were collected.

Finally with respect to cultural correlates of enrollments, another indicator was significant: high college enrollments are associated with high rates of divorce per 100 marriages, up to 50 and 53 in the U.S. and Sweden.

Achievement Test Scores

As shown in Figures 8.1 and 8.2, the scores of 14-year-olds on the standardized IEA of Science Achievement collected in 1970-71 predicted economic growth a decade later. Each one-point rise in the average test score is associated with 0.2 percent higher growth.

Among the eight nations that participated in the IEA of Science Achievement, Japan had by far the highest score, 31, and also the highest growth rate, 4.4. Italy, Britain, Sweden, and the U.S. had scores below 23 and correspondingly low economic growth below 2.5 percent.

High test scores, moreover, predict low rates of unemployment. A one-point rise in the national average test score was associated with 0.7 percent lower unemployment. Japan is again notable in having a low unemployment rate (2.4 percent) while Western nations with low scores—Italy, Britain, Sweden, and the U.S.—had high unemployment rates (3.1 to 12.5 percent).

Thus, national affluence and growth are demonstrably intertwined with indexes of education and culture. The rankings seem to indicate important characteristics of nations—their present and future. It is difficult, nonetheless, to determine their causal relations, value, and implications. Historians and social scientists have reached little consensus on why nations wax and wane. Reasonable people may also differ in how they interpret and evaluate the various indicators and their causes.

Yet it is important to try to understand these phenomena. Educators, in particular, should be concerned with the possible influence of enrollments and test scores on national welfare. How changes in education might affect our society and economy is worth thinking about.

The Information Economy

Modern life is increasingly concerned with the generation and processing of information. This prospect seems likely to demand an increasingly autonomous, cooperative, intelligent, and motivated work force (Walberg, 1983). Agriculture, mining, and heavy manufacturing occupy fewer of us today than in the past, and high value-added manufacturing such as electronics and pharmaceuticals and services such as insurance and data processing are growing throughout modern nations.

Although scientific and engineering breakthroughs occasionally raise productivity of these industries dramatically, progress comes mainly from adapting already-discovered ideas. These consist of small, incremental improvements in activities, materials, and components; redesign or substitution of activities, components, and products; reductions in costs; and advances in performance, quality, style, and consumer appeal.

To discover, plan, implement, and measure the results of such changes in traditional or high technology requires manufacturing, service, and sales forces who are knowledgeable and motivated. It also requires competent workers—not necessarily superior scientists, managers, and other leaders—skilled with materials, technology, and quantitative data. They also have to be able to absorb and propagate knowledge as well as communicate and cooperate with one another for individual, corporate, and national benefits.

These growing demands on the work force may explain why the knowledge and skills of the population, rather than the accomplishments of the elite, are responsible for economic growth and employment. The diligence and skills achieved by high standards and hard work in language, mathematics, and science in ele entary and secondary schools, combined with continuing on-the-job training, may foster economic productivity more than higher education and scientific discovery.

Trained Incapacity?

In addition, higher education may even be counterproductive for economic growth, as suggested by current data in Figures 8.1 and 8.2 and several causal possibilities. In his *Theory of the Leisure Class*, for example, University of Chicago economist Thorstein Veblen (1899) wrote of the "trained incapacity" of the highly educated and the elite's unproductive and "conspicuous consumption" of their time and money on the useless and esoteric to set themselves apart from the working people.

The continuing distinction of Cambridge and Oxford in the basic sciences and humanities has done little for Britain's plight: social-class tensions, unemployment, and declining commerce and industry. University emphasis on the classics and disdain of practical matters traditionally has set social classes apart and divided college-educated leaders from blue-collar workers.

In England, once the mightiest empire and workshop of the world, a curriculum theorist wrote that teachers "are right to resist the view of curriculum as being determined by the needs of society and especially the needs of industry" (Lawton 1980, p. 134). Is this part of the "English disease" that other Europeans fear?

Higher education enrollments are also associated with a shorter work week, high rates of divorce, and television consumption. Although affluence or modernity may cause all these indicators, possibly the leisure of the "new class" or the ivory tower of theory encourages cultural permissiveness and decline of the work ethic.

School Achievement and National Welfare

As suggested by recent IEA and other studies (Walberg 1983, McKnight 1987, and IEA 1988), the short school year, the lack of curriculum rigor, slow repetitive teaching, and little homework and parental involvement, seem the major causes of poor U.S. achievement. Even if this were certain, however, reasonable people may disagree on the value implications. Some may see in these trends the continuing economic and cultural decline of the West. Others may perceive a reorientation of society and education for a new era of liberation, leisure, and consumption in contrast to the deferral of gratification to attain long-term ends (Freud 1920) and prominence of the work ethic that has characterized the capitalistic Western economies since the Protestant Reformation (Weber 1958).

However, high school students apparently are unaware that the leisured, permissive millennium has arrived. When asked to name the most important American problem in a recent survey for the National Association of Secondary School Principals (Heard 1984), more than half mentioned unemployment, and a fifth said inflation. These ranked first and second in frequency among 14 problems mentioned. When asked what they most want out of life, more than a quarter answered career success, a quarter said happiness, and 21 percent mentioned marriage and family—to name the top three.

Japan's indexes—nearly three times the growth, one-fourth the unemployment, and one-third the divorce rate of the U.S.—are more consistent with the expressed values of the American youth than are the U.S. indexes. Japan provides an example of the possible causal link of school achievement and economic vitality. With the highest test scores, nearly the highest growth, and nearly the lowest unemployment of the nations with complete information, Japan may have set the educational and economic standards for the rest of the world.

Unlike Europe, Japan now graduates a greater fraction of its population from high school than does the U.S. (96 vs. 76 percent, U.S. Department of Education 1987, p. 79). Notwithstanding these often-expressed opinions, little evidence shows that Japanese students are uncreative, overstressed, and suicidal because of high standards and homework (Walberg 1983, U.S. Department of Education 1987).

Prospects

The most recent data on science and mathematics achievement offer gloomy prospects if Adam Smith and others are right about the importance of national ability for welfare. In *The Underachieving Curriculum*, McKnight and others (1987) reported that, considering both test scores and enrollments, U.S. math achievement is "very low." Of 8th grade students in 20 countries, those in the U.S. ranked 12th in algebra, 16th in geometry, and 18th (above Swaziland and Nigeria) in measurement. Of 12th graders in 15 countries, those in the U.S. scored 14th in advanced algebra and 12th in elementary functions and calculus.

Nor can we take comfort in our brightest students. Among the top 1 percent of 12th graders in the 15 countries, U.S. students scored worst in algebra and exceeded in functions and calculus only British Columbia, which doesn't offer these subjects in the high school curriculum. These findings are consistent with the National Assessment of Educational Progress (1988) conclusion that 1.5 million American 17-year-olds reach the end of high school each year unable to reason mathematically (p. 17).

The U.S. does little better in science (IEA 1988). Although near average at grades 4 and 5, U.S. students exceed only those in the Philippines and Singapore among 17 countries at grades 8 and 9. Among advanced 12th grade students in 14 countries, moreover, those in the U.S. ranked last in biology, 11th in chemistry, and 9th in physics.

In conclusion, both the importance and the uncertainty of national indexes and their causal relationships deserve emphasis. The U.S. has lagged in economic growth, which was well predicted by previous science test scores. We lag even further in the most recent comparisons. The indexes in which we excel—college enrollments, leisure, television consumption, and divorce—may not be conducive to future prosperity and national welfare.

It is possible, though, that during the first 18 years of life, families and schools *can* better foster knowledge, diligence, cooperativeness, and other psychological and social traits that promote a higher quality of future life. Although the correlations of indexes hardly prove this interpretation, they are consistent with it and with what many citizens believe. We may ignore its implications at our children's and our own peril.

References

Becker, G.S. (1976). *The Economic Approach to Human Behavior.* Chicago: The University of Chicago Press.

Chiba, S. (June 18, 1988). "Japan is Entering Talks in a Position of Strength." *New York Times*, sec. 2.

Comber, L.C., and J.P. Keeves. (1973). *Science Education in Nineteen Countries.* New York: John Wiley.

Congressional Research Service. (1986). *The Nobel-prize in Science as a Measure of National Strength in Science.* U.S. House of Representatives, 99th Congress, Second Session, Serial S.

Economist. (December 24, 1983). "Nation by Numbers," pp. 53-59.

Freud, S. (1920, 1961 reprint). *Beyond the Pleasure Principle.* New York: Norton.

Freund, W.A. (1982). *U.S. Economic Performance in Global Perspective.* New York: New York Stock Exchange, Office of Economic Research.

Hanushek, E.A. (1986). "The Economics of Schooling: Production and Efficiency in Public schools." *Journal of Economic Literature* 14:351-388.

Heard, A. (April 4, 1984). "Students' Interest in Education, Careers Found To Be Increasing." *Education Week* 8:8.

International Association for the Evaluation of Educational Achievement. (1988). *Science Achievement in Seventeen Countries: A Preliminary Report.* Oxford: Pergamon Press.

Kennedy, P. (1987). *The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000.* New York: Random House.

Lawton, D. (1980). *The Politics of School Reform.* London: Routledge and Kegan Paul.

McKnight, C.C., F.J. Crosswhite, J.A. Dossey, E. Kifer, J.O. Swafford, K.J. Travers, and T.J. Cooney. (1987). *The Underachieving Curriculum: Assessing U.S. Mathematics from an International Perspective.* Champaign, Ill.: Stipes Publishing.

Mincer, J. (1981). *Human Capital and Economic Growth.* Cambridge, Mass.: National Bureau of Economic Research Working Paper No. 803.

National Assessment of Educational Progress. (1988). *Mathematics. Are We Measuring Up?* Princeton, N.J.: Educational Testing Service.

National Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform.* Washington, D.C.: U.S. Government Printing Office.

Schultz, T.W. (1981). *Investing in People.* Berkeley, Calif.: The University of California Press.

Smith, A. (1776, 1976 reprint). *An Inquiry into the Nature and Causes of the Wealth of Nations.* Chicago: The University of Chicago Press.

Toynbee, A. (1959). *A Study of History.* New York: Oxford University Press.

U.S. Department of Education. (1987). *Japanese Education Today.* Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Education. (1988). *The Condition of Education: A Statistical Report,* 1987 edition. Washington, D.C.: Government Printing Office.

U.S. Department of State. (May 27, 1988). *OECD Summary Report.* Paris: Unclassified Memorandum.

Veblen, T. (1899, 1934 reprint). *The Theory of the Leisure Class: An Economic Study of Institutions.* New York: Modern Library.

Walberg, H.J. (1983). "Scientific Literacy and Economic Productivity in International Perspective." *Daedalus* 112:1-28.

Weber, M. (1958). *The Protestant Ethic and the Spirit of Capitalism.* New York: Scribner.

Afterword
A World of Assessment, A Universe of Data

CHESTER E. FINN, JR.

Americans have not always appreciated international information about education. We have at times been self-absorbed, even insular. But we value it now. Indeed, we crave it. Every time an IEA report comes out, for example, people are riveted by the data. We now recognize the need for more, better, and more regular international information than is currently available.

Cross-national research widens our horizon and can sometimes suggest nonincremental means of improvement. The discovery that a number of other countries devote more time to schooling than we do, for example, has stimulated domestic debate about lengthening the school day and the school year.

Educational performance in other countries provides benchmarks for us to use as we wish. Japan, for instance, does something many Americans had not thought possible. The Japanese graduate an extraordinarily large proportion of their students from secondary school while simultaneously producing a very high average level of academic achievement. This finding challenges our conventional wisdom. Contrary to the assumption of many American educators, equality and excellence can both be achieved within the same system. Japan illustrates the possibility.

Another instance in which an international study removed our blinders is the IEA's Second International Mathematics Study. It con-

Chester E. Finn, Jr., is Director of the Educational Excellence Network in Washington, D.C., and Professor of Education and Public Policy in the Vanderbilt University office in Washington, D.C.

firmed that achievement follows content coverage, that students tend to learn what is taught. The study also revealed something that many Americans had not supposed possible—that students can be taught complex mathematics at a relatively early age.

International comparative data frequently indicate areas that need our attention. The 1972 IEA reading and literature studies, for instance, showed that a major weakness among U.S. secondary students is comprehension. Findings from the latest IEA science study show U.S. students discouragingly far behind their peers from other lands in physics chemistry, and biology.

International comparisons of student performance generally stir debate in the United States, and they sometimes trigger action at the state and local levels. In the aftermath of the Second International Mathematics Study, school districts in Maryland formed a consortium to examine the study results and to initiate reforms. An organization in Minnesota, the Mathematics Quorum, informed each school district in the state about the study's conclusions. Three other states—Virginia, Florida, and Iowa—conducted their own surveys of mathematics to compare themselves with other countries. The results helped Floridians develop a new geometry curriculum.

International information has helped fuel efforts to strengthen education in the U.S. and in other democracies as well. Progress in social welfare, economic growth, and cultural vitality is something that concerns all of us, and progress in each of these domains depends in no small part on improvements in education. At a time when technology has shrunk the globe—when TV, telephones, satellites, facsimiles, and jet travel have turned once-distant lands into neighbors—we need, more than ever, to learn from each other, to benefit from each other's experiments and experiences, triumphs and errors. We need to undertake collaborative efforts wherever possible.

We also need more data—and more kinds of it—on education around the world. We need, as one example, a reliable way to chart education's percentage of the gross national product in each country. Not just government expenditures on education: *all* expenditures on education. Why can't this be done? It is among the most fundamental points of reference for comparing education systems and their relative productivity. Another international blind spot is the curriculum. Can we not do a decent comparison of course content, subject by subject, across nations?

Some of the reasons we lack international data are understandable, in their way even legitimate. Countries organize their systems differently and measure things differently. Data on school dropouts, for example, are difficult enough to collect within the U.S. It's hard even to get the 50 states to agree on a definition and on the elements that should be measured. International agreement is that much harder to reach, yet it is vital that we work out a way to do this. Attrition and completion rates are among the essential indicators of any education system.

If some of the reasons we lack international comparisons are reasonable, at least for the time being, others are not. Too many countries—the U.S. has been vulnerable to this at times—have been secretive, uncooperative, or simply lazy. We need joint efforts to galvanize support for more valid, more detailed, more varied, and more useful international information.

The international agency best positioned to accomplish this is the Organization for Economic Cooperation and Development (OECD). During my time in the Education Department, we pressed OECD hard to develop a better set of international indicators, and, in the past [illegible] years, we have seen promising movement in this direction.

In the fall of 1986, we hosted an International Meeting on Educational Indicators in Washington under OECD auspices. Representatives from most of the member countries attended, many indicating an interest similar to ours in securing timely, useful information about education quality, performance, and results.

As a result of this meeting and others that followed under the aegis of OECD's Education Committee and its Center on Educational Research and Innovation (CERI), OECD has initiated a project to develop educational indicators. The first step was to divide the universe of possible indicators into five categories and invite countries to participate in the development of those they are most interested in.

The United States' greatest interest is in cognitive achievement; therefore, we volunteered to coordinate the network on cognitive outcomes. Our National Center for Education Statistics is currently working with a dozen countries in the design of a set of such indicators and the means of gathering the requisite information.

It seems to me that, instead of each working in splendid isolation, a more stable and closer relationship between IEA and OECD would be most beneficial. IEA very likely will be the source of much of the cognitive outcomes data the OECD will want to gather, and it certainly would be a waste of resources if OECD were to replicate what IEA is already doing. Organizationally, OECD represents governments, while IEA members represent mostly research organizations, universities, foundations, and some governments. OECD is a fairly stable enterprise, financed largely through fixed annual payments from member countries. IEA is rather more fragile, highly dependent on private philanthropy and volunteer labor. There is, to my mind anyway, an obvious symbiosis of interests, capabilities, and strengths waiting to be developed.

Although the single most important yardstick of the performance of any education system is student learning, we also need data, research, and indices of many other dimensions—features such as curriculum content, school organization, attendance, parent involvement, homework policies, retention rates, enrollments in advanced courses and in college, how education revenues are used, and performance in higher education. We need to gauge the gaps between the intended curriculum, the implemented curriculum, and the attained curriculum. And we need to look for links between student outcomes and the variables that parents and educators can control. I'm thinking here of Harold Stevenson's[1] finding that American parents tend to attribute their child's learning to innate ability whereas Asian parents more often believe it to be mainly a result of hard work. This, of course, is part of the reason Asian students learn more.

Studies such as Stevenson's are doubly important because they indicate not only differences in student achievement, but also factors that help explain the reasons for those differences. Stevenson's study points to a variable that lies within human control: attitudes are something we can work to change.

I'd like to see more studies such as this one and more analyses of data—across institutions, regions, nations, and other geographic and cultural boundaries. Such analyses can reveal weaknesses in an education system. They can also highlight features most susceptible to improvement, that is, those most likely, if modified, to yield increases in student outcomes.

Of course, not everything we need to find out is amenable to measurement and certainly not to testing. Psychometrics cannot capture or gauge subtle qualities of mind and spirit. But much that we want to know can be quantified and measured, and it should be! The whole field of testing and measurement is making great strides and will make more.

[1] Harold W. Stevenson, "America's Math Problems," *Educational Leadership* 45, 2 (October 1987): 4-10.

Although we must not forget the immeasurables, we must not slight what can be measured.

Today the work of the IEA comprises the single most valuable asset in the world with respect to the cross-national appraisal and analysis of education outcomes. It is far from a perfect organization, but our vow must be to strengthen it, for only with the kind of information it provides will we finally know whether our current reform efforts are productive and which areas need our attention next.

Association for Supervision and Curriculum Development
Alexandria, Virginia