NATIONAL CENTER FOR EDUCATION STATISTICS

Research and Development Report January 1992

International Mathematics and Science Assessment: What Have We Learned?

Elliott A. Medrich
Senuir Research Associate
MPR Associates, Inc.

Jeanne E. Griffith
Associate Commissioner for Data Development
National Center for Education Statistics

U. S. Department of Education
Office of Educational Research and Improvement NCES 92-011

U.S. Department of Education
Lamar Alexander
Secretary

Office of Educational Research and Improvement
Diane Ravitch
Assistant *Secretary*

National Center for Education Statistics
Emerson J. Elliott
Acting Commissioner

National Center for Education Statistics
"The purpose of the Center shall be to collect, and analyze, and disseminate statistics and other data related to education in the United States and in other nations."—Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221 e-l).

February 1992

Contact:
Jeanne E. Griffith (202)

Forward

As international economic pressures increase demans for a well-educated work force, Americans expect more from the Nation's schools. Over the past 25 years a series of international studies has focused attention on how elementary and secondary students from the United States perform in mathematics and science as compared with students from other countries. Results from the international surveys have been a matter of intense interest and debate. On the one hand, they have drawn attention to the apparently mediocre performance of American students, as well as to curriculum and instructional practices that have raised questions about our own. On the other hand, a variety of technical issues concerning the nature of the surveys, the comparability of the populations tested, and the quality of the data have led to some questions about the reality of the findings.

This report addresses two related issues. First, it summarizes the past international studies of mathematics and science, describing each study and its primary results. In place of country-by-country performance rankings, the report presents the average performance for each country accompanied by an estimate of the statistical error circumscribing the limits of meaningful country-to-country comparisons. Second, the report draws together critical and heretofore inaccessible documentation-information that scientists require to evaluate the quality of the surveys. For example, information on cross-national differences in response rates are presented in every case where these data were available. At the same time, the authors point to other non sampling errors that may affect the data reliability and validity as well, but about which we do not have sufficient information to quantify.

Despite these data-related concerns, the international surveys-which have been done at different times and in different ways--come to some similar conclusions. This pattern of consistency suggests that the overall results are powerful and cannot be discounted. Learning about teaching and learning processes in other countries can lead to enhanced student performance in American schools. Only by addressing the data-related problems that hamper international studies will the potential for this kind of research be fully realized. We hope that the insights in this report will continue to improve the planning and execution of future studies.

NCES, jointly with the National Science Foundation, has been striving in recent years to strengthen the quality and generalizability of international assessments. We believe that considerable improvements will soon be evident in reports from recent assessments of science and mathematics and also of reading literacy. Further improvements are being incorporated into the design of a new study of mathematics and science achievement scheduled in 1994 and 1998 that the United States will use in monitoring progress toward achieving the fourth National Education goal, which states that "By the year 2000, U.S. students will be fwst in the world in science and mathematics achievement."

Emerson J. Elliott
Acting Commissioner of Education Statistics

Acknowledgments

A number of individuals with significant commitments to enhancing the quality and utility of international achievement studies helped organize and execute this report.

Larry Suter (formerly at the National Center for Education Statistics and now at the National Science Foundation) proposed and initiated the project. Before the work began, he organized a one-day conference to define some of the issues that would be addressed in this report. Attending the conference were Senta Raizen (The National Center for Improving Science Education), Rarnsay Selden (Council of Chief State School Officers), Constance Sorrentino (U.S. Department of Labor, Bureau of Labor Statistics), Harold Stevenson (University of Michigan), and several representatives from the National Center for Education Statistics. The conference proved invaluable to the authors and considerably focused the project agenda..

Several individuals at the National Center for Education Statistics offered considerable assistance and guidance as the work progressed. John Ralph and Lois Peak provided insightful comments on the linkages between international issues and questions of policy in the United States, and Mary Frase and Sue Ahmed assured that statistical issues were appropriately addressed. Dawn Nelson assumed responsibility for managing the project in the winter of 1991, and did everything possible to facilitate the work and to ensure a high-quality product.

At MPR Associates, Gary Hoachlander provided the support essential to meeting the needs of the project, Philip Kaufman helped resolve a variety of thorny statistical problems, and Christina Chang ably assisted as a research intern.

National Center for Education Statistics
Research and Development Reports

The Research and Development (R&D) series of reports has been initiated:

1) To share studies and research that are developmental in nature. The results of such studies may be revised as the work continues and additional data become available.

2) To share results of studies that are, to some extent, on the “cutting-edge” of methodological developments. Emerging analytical approaches and new computer software development often permit new, and sometimes controversial, analysis to be done. By participating in “frontier research,” we hope to contribute to the resolution of issues and improved analysis.

3) To participate in discussion of emerging issues of interest to educational researchers, statisticians, and the Federal statistical community in general. Such reports may document workshops and symposiums sponsored by NCES that address methodological and analytical issues, may summarize or synthesize a body of quantitative research, or may share and discuss issues regarding NCES practice, procedures, and standards.

The common theme in all three goals is that these reports present results or discussion that do not reach definitive conclusions at this point in time, either because the data are tentative, the methodology is new and developing, or the topic is one on which there are divergent views. Therefore the techniques and inferences made from the data are tentative and are subject to revision. To facilitate the process of closure on the issues, we invite comment, criticism, and alternatives to what we have done. Such responses should be directed to:

Roger A. Herriot
Associate Commissioner for Statistical Standards
and Methodology National Center for
Education Statistics
555 New Jersey Avenue, NW
Washington, DC 20208-5654

Executive Summary

The changing world economic order, foreshadowing new demands on the labor force and workplace, highlights the larger international context within which American education must be viewed. In January 1990, President Bush and the Nation's Governors recognized these evolving needs and established a specific goal for mathematics and science education-two subject areas critical to successful competition among highly technological societies: "By the year 2000, U.S. students will be first in the world in science and mathematics achievement. " To measure progress toward this objective, there is increasing interest in the periodic international assessments of student performance in mathematics and science.

Over the past quarter century, there have been five major international studies of science and mathematics achievement at the elementary, middle, and secondary school levels. The studies have been conducted under the auspices of two different nongovernmental research consortia. More than 30 countries have participated in at least one of the surveys. The United States has been involved in every one. A great variety of findings have resulted from this work, and these studies represent valuable contributions to the ways in which schooling inputs and outcomes are understood. The research has challenged participating countries to examine the structure, practices, and curricula of their educational systems, and as a consequence, to envision the possibility of rethinking curriculum content and the ways in which students are taught.

This report provides a description of the international assessments and some of their findings, and addresses issues surrounding the collection and analysis of these data. Further, it offers suggestions about ways in which new data collection standards could improve the quality of the surveys and the utility of the reports in the future.

Three Mathematics Surveys

- The First International Mathematics Study, conducted in the 1960s, involved 13 year-old students from 10 countries and students in their last year of secondary ‘ school from 10 countries.

- The Second International Mathematics Study, performed in the early 1980s, involved 13-year-old students from 18 countries and students in their last year of secondary school from 13 countries.

- The First International Assessment of Educational Progress, carried out in 1988, involved 13-year-olds from six countries.

Three Science Surveys

- The First International Science Study, conducted between 1966 and 1973, involved 10-year-old students from 16 countries, 14-year-old students from 18 countries, and students in their last year of secondary school from 18 countries.

- The Second International Science Study, performed between 1983 and 1986, involved 10-year-old students from 15 countries, 14-year-old students from 17 countries, and students in their last year of secondary school from 13 countries.

- The First International Assessment of Educational Progress, carried out in 1988, involved 13-year-old students from six countries.

The evidence suggests, in general, that students from the United States have fared quite poorly on these assessments, with .tieir scores lagging behind those of students from other developed countries. This finding is based largely on analyses of mean achievement scores and related rankings of countries participating in each survey. Understanding that large-scale surveys pose a variety of analytical constraints and profit when complemented by more intensive case studies of particular findings, the international assessments do not explain why students from some countries perform better than their American counterparts. In fact, regular and systematic patterns of differences are absent. For example, while students from some countries may do better on some or most of the achievement tests than students from other countries, the findings are age-group and subject-matter specific. Hence, they are very difficult to generalize since they are not the product of a single set of related, overriding school or institutional factors. Even so, across the studies certain trends appear to be clear:

- The more students are taught, the more they learn, and the better they perform on the tests. There are significant differences in the content of instruction among countries at common levels of schooling.
- Use of a differentiated curriculum based on tracking is negatively associated with student performance on the international assessments and also reduces opportunities for some students to be exposed to more advanced curriculum.
- The school affects learning in some subject areas more than in others.
- Countries committed to keeping students enrolled in secondary school score less well on the international surveys, but they spread more knowledge across a larger population. Japan is an exception. Even with high retention rates at the secondary level, Japanese students perform very well on the mathematics and science achievement surveys.
- Generally the “best students” in the United States do less well on the international surveys when compared with the “best students” from other countries.

A number of technical considerations inhibit generalizing many other findings, The surveys have not achieved high degrees of statistical reliability across the age groups sampled and among all of the participating countries. Thus, from a statistical point of view, there is considerable uncertainty as to the magnitude of measured differences in achievement. Inconsistencies in sample design and sampling procedures, the nature of the samples and their outcomes, and other problems have undermined data quality. But despite their shortcomings, international achievement surveys now provide valued ways of documenting differences and investigating issues in student performance cross-nationally. The challenge in the future will be to make certain that these surveys meet quality technical standards.

From all indications, the various international testing authorities and consortia are moving expeditiously toward improving the quality of the surveys and upgrading their statistical reliability before the next rounds of international mathematics and science studies. Among the important tasks that lie ahead are strengthening the comparability of samples from country to country and developing new ways of reporting international achievement scores that will meet a variety of requirements and interests. It is noted that a considerable need also exists for small-scale case studies. These studies achieve in depth what they lack in breadth and help researchers understand the circumstances contributing to differences in performance among systems of education.

The report concludes by suggesting that there is a need for more deliberate consideration of policy concerns in the design of international assessments. This, in turn, may provide opportunities for policymakers and education practitioners to apply what is learned about cross-national differences in achievement to curriculum development and programming.

Table of Contents

Page

Appendices

List of Tables

Table
Page

Appendix Tables

Appendix B

List of Figures

Figure **Page**

Chapter I
Student Achievement in an International Context

As we enter the last decade of the 20th century, extraordinary changes in the shape of the world foreshadow equally important changes in the marketplace and in the workplace. The demands on the rising generation will be formidable.

For educators and education policymakers the implications of these changes have been clear for some time. As early as 1983, the National Commission on Excellence in Education cast special urgency on the matter of schooling and international competition in their landmark report, *A Nation at Risk.*

> Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world. . . .What was unimaginable a generation ago has begun to occur-others are matching and surpassing our educational attainments.[1]

In January 1990, 7 years after the National Commission's report, President Bush and the Nation's Governors highlighted the larger international context within which American education must be viewed:[2]

> Our people must be as knowledgeable, as well trained, as competent, and as inventive as those in any other nation. All our people, not just a few, must be able to think for a living, adapt to changing environments, and to understand the world around them. They must understand and accept the responsibilities and obligations of citizenship. They must continually learn and develop new skills throughout their lives.

Addressing the intense technology-based environment within which the United States must compete, the President and the Governors defined a specific objective in the areas of mathematics and science education. They proposed that by the year 2000, U.S. students should rank first in the world in science and mathematics achievement.

Policymakers, business leaders, educators, and citizens all note a perceived link between the future for a strong America and a well-educated labor force, capable of adjusting to the demands of a society in which technology and information hold the key to competitiveness. It is not that the Nation wants our education system to be driven by labor markets-for education defines the essence of our democracy and plays a much broader role in developing a responsible citizenry—but voices from all sectors point to the need for linkage. In a study by the Organization for Economic Cooperation and Development (OECD), it is

---

[1]U.S. Department of Education, National Commission on Excellence in Education, *A Nation at Risk* (Washington, DC, 1983), 5.
[2]U.S. Department of Education, *National Goals for Education* (Washington, DC, July 1990), 1.

> . . our societies are going through a period of rapid and far-reaching change. The signs of this are manifold... .Technological progress, international trade, the speed of communications, world competition...these are just some aspects of the change which is
> posing crucial questions for our societies, structures and habits . . . . The analyses undertakeil in the OECD, as elsewhere, in order to assess the effect of structural changes on economic performance all point to the decisive and fundamental importance of education systems. It is they that hold the key to possible progress and that determine each country's medium and long-term prospects in world competition.[3]

This is the challenge. And this is one reason why it is so important to understand how American youth compare with those of other countries on educational performance, and what factors in social, economic, and educational policies and programs are associated with different levels of achievement.

Since the early 1960s, cross-national studies of student achievement have become one way of evaluating the product of the educational enterprise. While objectives governing the design of these studies have been many and varied, as often as not, public attention has focused single-mindedly on how students score on the performance tests and how countries rank against one another—as though the surveys represented a kind of international intellectual olympics.

In fact, international studies of student achievement are useful for many reasons other than performance comparisons. The most important benefit to the United States of participating in the international assessments is that understanding is gained of a much wider variety of education policies, programs, and practices that can help us improve our own educational system. The National Research Council's Board on International Comparative Studies in Education (BICSE), which is sponsored by the National Center for Education Statistics, the National Science Foundation, and the Department of Defense, defines a broad set of objectives:

> . . comparative research on education.. increases the range of experience necessary to improve the measurement of educational achievement; it enhances confidence in the generalizability of studies that explain the factors important in educational achievement; it increases the probability of dissemination of new ideas to improve the design or management of schools and classrooms; and it increases the research capacity of the United States as well as tit of other countries. Finally, it provides an opportunity to chronicle practices and policies worthy of note in their own nght.[4]

While some believe that the American values of equality, practicality, and individualism combined with issues of local control of education may limit the possibility

[3]0rganization for Economic Cooperation and Development *Education and the Economy in a Changing Society* (Paris: OECD, 1989), 7.

[4]Norman M. Bradbum and Dorothy M. Gilford, eds., "A Framework and Principles for International Comparative Studies in Education" (Washington, DC: National Academy Press, 1990) 4.

of educational borrowing, there is clear evidence that all of this is changing. It is now understood that international achievement studies can influence and help improve education policy and programs in the United States and that these surveys represent important opportunities to think about and examine many aspects of schooling in America by means of comparison. On balance, too many of the most widely publicized summaries of the surveys obscure rather than illuminate their meaning, and draw conclusions inappropriate to their content and scope. This undermines many serious efforts to examine what these studies really say about the skills and capabilities of American students, as compared with those from other countries. Moreover, it diminishes efforts to describe what can be learned about teaching methods, classroom processes, and curriculum in other countries that might enhance schooling outcomes in the United States.

Objectives of This Report

By providing a summary of the results of a select group of cross-national surveys, this report turns its attention away from the newspaper headlines. A special effort will be made to understand the meaning of and import of the achievement test scores, recognizing that this is just one aspect of the research. This synthesis has four objectives:

- To summarize and describe the international mathematics and science surveys and survey samples;
- To understand what the test scores and associated findings do and do not say;
- To explore some important issues of study design and data presentation that may help researchers in preparing for similar studies in the future; and
- To suggest some strategies for upgrading data quality in future studies.

Comparative international achievement represents anew set of issues for the National Center for Education Statistics, and this report is written to meet several needs. First, NCES receives an increasing number of inquiries from Congress, the Executive Branch, and others who are interested in various issues addressed by the international achievement surveys and want to know more about what these data say. This report should be useful to those who require a general overview of these studies. Second, since NCES is now sponsoring international assessments,G it is impmtant to ascertain how the data measure up to NCES standards for data collection efforts. NCES is now being asked by policy makers to stand behind these studies. Can the data upon which the educational performance of U.S. students is compared with the performance of students from other countries meet the standards NCES applies to its own databases before release? This report describes a variety of data-related problems that deserve attention so that the quality of future surveys can be strengthened and their increasing use in the policy arena can be supported by this agency.

Despite data-related problems, the past international studies collectively have generated important findings and hypotheses in education research. These, too, are

---

[5]B. Bum and C. Hum, “An Analytic Comparison of Education Systems” (paper prepared for the U.S. Department of Education, National Commission on Excellence in Education, 1983).

[6]As this is written, two sets are in progress: one is near completion, the other is planned and scheduled. The Educational Testing Semite will publish results of the 1991 International Assessment of Educational Progress (in the winter of 1992); and the International Association for the Evaluation of Educational Progress will undertake the Third International Mathematics and Science Study in two phases, one in 1994, the other in 1998.

summarized in this report to demonstrate some of the strengths of comparative research. The findings stand out beyond the flaws for one or more of the following reasons: 1) they are consistent across many studies and test populations; 2) they are identified by analyzing relationships within the &ta that are less subject to the technical problems identified, 3) they are important in that they identified important hypotheses that appear to be supported crossnationally but may need further evaluations; or 4) they corroborate education and social theory tested in other national studies.

Thus, this study attempts to objectively present both the technical problems and the substantive strengths of these international assessments.

Organization of This Report

This report focuses on five studies of science and mathematics at elementary, middle, and secondary school levels—curriculum areas that, in the more developed countries at least, tend to involve instruction in somewhat similar subject matter covered at about the same grade ranges. Constraining the synthesis in this way provides an opportunity to look more closely at two areas of instruction that the Nation has associated with international competitive issues and our Nation’s capacity to move toward the emerging 21st century economy.

More than 30 countries have participated in one or more of the studies discussed in this report. Four grade levels have been tested in at least one subject area (mathematics and/or science) at least once. The United States is unique in its commitment to international testing. No other country has been involved in as many studies at as many grade levels.

Chapter I establishes the context for this synthesis. Chapter II summarizes the large scale international mathematics and science surveys that have been conducted over the past quarter century and explores general issues of data quality. Chapters III and IV look at the achievement scores and some of the key findings of the studies. These chapters should be read along with the accompanying appendices that bring together, for the first time in a single source, much of the basic data needed to understand and summarize the surveys. Chapter V looks ahead, raising some of the data-related issues that could be addressed and that might improve future international surveys. With a new round of studies underway, this is an appropriate moment to review some of the results of past research, and to look at what these studies report and on what basis.

Chapter II

International Achievement Surveys of Mathematics and Science: An Overview

International studies of student achievement are extraordinarily complex research projects that are difficult to organize, administer, and analyze. To appreciate their strengths and weaknesses, they must be understood against the backdrop of the research tradition that has defined their objectives and shaped their analytical focus.

The Comparative Education Research Tradition

Until the late 1950s most comparative education research was aimed at describing the mandate, structure, and support base for schooling within countries-types of schools, level and sources of fiscal support, curriculum, teaching methods, enrollments, and so forth. Little attention was paid to outcomes, other questions of performance, or student achievement.

In 1959 this situation changed dramatically. That year a number of researchers, committed to understanding not only the nature of schooling across nations but also the quality of the educational product, founded the Council for the International Evaluation of Educational Achievement, subsequently known as the International Association for the Evaluation of Educational Achievement (lEA).

Since its inception, the IEA has significantly influenced the direction of comparative education research by focusing its attention on the relationship between schooling inputs and processes and student performance. T. Neville Postlethwaite, one of the IEA's founders, described four objectives for comparative studies of this type:[7]

- Identifying what is happening in different countries that might help improve education systems and outcomes, such as philosophy of education, curriculum, resources, the organization of schools, teaching methods, qd so on;
- Describing similarities and differences among systems of education and interpreting them in terms of educational outcomes;
- Estimating the relative effects of variables that are thought to be detemninants of educational outcomes (both within and among systems of education); and
- Understanding why certain phenomena or practices appear to be important in some systems of education but not in others.

Comparative studies now subsume a large literature that, as Postlethwaite writes, "When done well.. can deepen our understanding of our own education and society.. can be of assistance to policymakers and administrators and.. can be a valuable component of

[7] T. Neville Postlethwaite, "Preface," ed. T. Neville Postlewaite, *Encyclopedia of Comparative Education and National Systems of Education* (Oxford: Pergamon, 1988), xvii-xxvi.

teacher education programs."[8] But despite the variety of stated objectives, among all the products of comparative education research, cross-national comparisons of student achievement have attracted the most attention. Interest in such comparisons is ubiquitous, and Americans, ever sensitive to issues of performance, are especially concerned with "where we stand." Although there may be many reasons to resist simple comparisons of student achievement, international studies rest uncomfortably between the world of the researcher, committed to using comparative data to enrich the ways of understanding how schools work, and the world of the policy maker and the educator, who must use student outcome data to help decide how to allocate scarce resources among programs and to defend the results of funded programs. International achievement comparisons represent an uneasy bridge between these two worlds.

The strength of the international surveys of student achievement, as with other surveys, rests on the quality of the study and sample design and its implementation. If these data are to represent real performance differences across countries, a necessary but not sufficient condition is that the samples must meet reasonable standards of cross-national comparability. From the perspective of policymakers and practitioners, the issue of sampling outcomes is far from academic, given the level of interest in the achievement scores and the potential bias that can be introduced by selective or nonrepresentative samples.

Evaluating Data Quality and Defining a Field Outcome Standard

This report attempts to evaluate some very selected technical aspects of the international mathematics and science studies with a view toward understanding where future improvements are indicated to support broader policy use of the results. International achievement surveys are based on samples; hence, the data are susceptible to both sampling and nonsampling errors. Sampling errors occur because estimates are based on samples of students, not on entire student populations. Nonsampling errors may be caused by many factors, among them an inability to obtain complete and correct information from and about participants and nonparticipants; non-response; mistakes in recording or coding data; and errors in collecting, processing, sampling, and estimating missing data. In international studies, the special problem of differences in meaning introduced in the translation of test instruments into different languages is an important non-sampling issue. Non-sampling errors are difficult to estimate, but they may result in bias and non-reliability of the data themselves. Weights were used in each study to account for the sampling design and to compensate for non-response; however, it was not possible to analyze weighting schemes and their impact on data in this report.

Response rates offer important information on the technical quality of each international survey sample. The response rate is the ratio of those who actually participated in a survey compared with those selected to participate in the sample. While there is no formal statistical basis for defining adequacy of response rates, the National Center for Education Statistics (NCES) has established its own standards, and these shall be adopted for this discussion.[9] The NCES standard establishes

---

[8] Ibid., xix.

[9] 'See U.S. Department of Education, National Center for Education Statistics, *Standards and Policies, March 16,1987,* CES Standard 87-03-04.

> minimum levels for performance in surveys and studies conducted by the Center. The levels of data completeness and minimum levels of data required for processing procedures and analysis are established to ensure that researchers and users will have confidence in the quality of the data. . .. The overall survey target response
> rate... should be at least 85 percent for cross-sectional surveys. In the case where the sample is selected hierarchically (e.g. schools, and then teachers within schools) these rates apply to each hierarchy….[10]

The NCES standard represents an effort to define an adequate "field outcome" for purposes of evaluating the quality of its own data programs and determining adequacy for release, which, in turn, provides one way of describing data quality. While differences of opinion exist regarding the definition of an acceptable response rate, the NCES standard is a rigorous target. Although the international surveys were neither organized nor funded to achieve such high response rates, high levels of non-response may have a significant impact on the findings and how they can be interpreted. In fact, a lower response rate might be acceptable if it could be shown that non-response bias was minimal or randomly distributed. However, for future studies that NCES is heavily involved in funding, adequate response levels will have to be attained for the agency to be able to stand behind the results. Therefore the NCES standard represents one way of evaluating the likelihood of non-response bias in the absence of any other test. To the extent that data fall short of the NCES standard, they may be more likely to be biased because it is not known if the nonresponse is proportionately distributed across the sample target population. Since non response was not analyzed in the technical reports supporting the surveys, the concern here about response rates is reasonable and survey results must be viewed with caution.

Assessing the adequacy of samples also requires examining the extent to which the
samples meet study design requirements, understanding how countries defined sample eligibility, and describing how refusals to participate were handled. These questions underlie larger issues of survey design and administration and may be as important sources of non-sampling error as are response rates. The data needed to analyze these matters were often not available for the international studies, and therefore, in this report they are discussed with reference to some studies and not others. The general issue of study design requirements and the international achievement surveys will be considered in the concluding chapter. [11]

The Five International Studies

This report focuses on five studies of mathematics and science achievement that were conducted over a 25-year period. They represent a range of test types and organizing procedures, and most important, they are arguably the most competently executed, largescale international surveys of their type. Figure II. 1 describes the basic elements of each study. As suggested by the figure, the ways in which participating entities defined themselves does not make for simple country-to-country comparisons. In many countries, sub-populations administered by autonomous educational authorities participated in the surveys independent of one another. In other words,
"whole" countries were not always
sampled. (For example, in some studies a number of Canadian provinces tested separately in French and English, as was also the case in French and Flemish Belgium.) Hence, it is

---

[10]Ibid., 15.

[11]Appendix A presents the response rates for the five studies discussed in this report

worth emphasizing that in every study there are more educational *systems* participating than *countries* participating. These distinctions, in turn, inhibit deriving and comparing national estimates.

The remainder of this chapter describes the surveys, their target populations and samples, the survey response rates, the content of each achievement test, and related data collection issues. The material is drawn from published sources, which presents a special problem. Many technical reports and strategic bulletins were produced in conjunction with the various studies after the surveys were completed, but most were not made available to the larger research community. As a result, while a great deal may be known about the samples by individuals directly involved in this research, much of what is required for evaluating the quality of the data is not available (e.g., reports describing sample execution from country to country are not available or accessible years after the studies were completed). Table II. 1 summarizes the response rates based on the NCES 85 percent standard mentioned above. Note particularly how few countries achieve the 85 percent goal, and that the United States reaches this level only on one study.

Figure II.1—Participants in the international achievement studies of mathematics and science

	Australia	Belgium	Belgium (Fl)	Belgium (Fr)	Canada (Eng)	British Columbia	New Brunswick (Eng)	New Brunswick (Fr)	Ontario (Eng)	Ontario (Fr)	Quebec (Eng)	Quebec (Fr)	Chile	England	Fed Rep of Germany	Finland	France	Hong Kong	Hungary
IEA																			
First Mathematics (1963-64)																			
Age 13	X	X												X	X	X	X		
Last-year secondary (mathematics)	X	X												X	X	X	X		
Last-year secondary (non-math)		X												X	X	X	X		
Second Mathematics (1980-82)																			
Age 13			X	X		X			X[1]					X[3]		X	X	X	X
Last-year secondary			X	X		X			X[1]					X[3]		X		X	X
IEA																			
First Science (1966-73)																			
Age 10			X	X										X	X	X			X
Age 14	X		X	X										X	X	X			X
Last-year secondary	X		X	X										X	X	X	X		X
Second Science (1983-86)																			
Age 10	X				X									X		X		X	X
Age 14	X				X									X		X		X	X
Last-year secondary	X				X									X		X		X[2]	X
IAEP																			
Mathematics and Science (1988)																			
Age 13						X	X	X	X	X	X	X		X[4]					

	India	Iran	Ireland	Israel	Italy	Japan	Korea	Luxembourg	Mexico	Netherlands	New Zealand	Nigeria	Norway	Philippines	Poland	Scotland	Singapore	Spain	Swaziland	Sweden	Taiwan	Thailand	USA
IEA																							
First Mathematics (1963-64)																							
Age 13				X		X				X						X				X			X
Last-year secondary (mathematics)				X		X				X						X				X			X
Last-year secondary (non-math)						X				X						X				X			X
Second Mathematics (1980-82)																							
Age 13				X		X		X		X	X	X				X			X	X		X	X
Last-year secondary				X		X					X					X				X		X	X
IEA																							
First Science (1966-73)																							
Age 10					X	X				X						X				X			X
Age 14					X	X				X	X					X				X			X
Last-year secondary					X					X	X					X				X			X
Second Science (1983-86)																							
Age 10					X	X	X						X	X	X		X			X			X
Age 14					X	X	X			X			X	X	X		X			X		X	X
Last-year secondary					X	X							X		X		X			X			X
IAEP																							
Mathematics and Science (1988)																							
Age 13			X				X											X					X

[1] English and French Ontario participated together as one system.
[2] Two grade levels, Hong Kong form 6 and form 7, participated separately as two systems.
[3] England and Wales.
[4] United Kingdom.

Table II.1—Number of participating systems known to achieve 85 percent response rate at each level of sampling

	Age 10		Age 13		Age 14		Last-year secondary	
	Total participating	Known to achieve 85% criterion	Total participating	Known to achieve 85% criterion	Total participating	Known to achieve 85% criterion	Total participating	Known to achieve 85% criterion
First Mathematics Study[1]	—	—	12	0	—	—	12 10	0 (Math students) 0 (Non-Math students)
Second Mathematics Study	—	—	20	4[2]	—	—	15	6[3]
First Science Study	17	6[4]	—	—	19	7[5]	19	4[6]
Second Science Study	15	8[7]	—	—	17	10[8]	14 14 14	1 (Biology)[9] 1 (Chemistry)[10] 2 (Physics)[11]
IAEP	—	—	12 12	10 (Mathematics)[12] 10 (Science)[13]	—	—		—

—Not sampled.
SOURCE: See Appendix A.

[1] No data available on response rates.
[2] Hungary, Luxembourg, Sweden, Thailand. No data, or insufficient data available for 12 systems. Four provided data and did not meet criteria. Hence 4 of 20 known to achieve standard.
[3] Finland, Sweden, Hungary, Japan, New Zealand, Thailand. No data or insufficient data available for 7 systems. Two provided data and did not meet criteria. Hence 6 of 15 known to achieve standard.
[4] Finland, Hungary, Israel, Japan, Scotland, Sweden. No data available for 1 system. Ten provided data and did not meet criteria. Hence 6 of 17 known to achieve standard.
[5] Australia, Finland, Hungary, Scotland, Japan, New Zealand, Sweden. No data available for 1 system. Eleven provided data and did not meet criteria. Hence 7 of 19 known to achieve standard.
[6] Australia, France, Hungary, Sweden. No data available for 1 system. Fourteen provided data and did not meet criteria. Hence 4 of 19 known to achieve standard.
[7] Singapore, Finland, Hong Kong, Hungary, Japan, Korea, Phillipines, Poland. Seven provided data and did not meet criteria. Hence 8 of 15 known to achieve standard.
[8] Finland, Hong Kong, Hungary, Japan, Korea, Netherlands, Phillipines, Thailand, Poland, Singapore, Thailand. Seven provided data and did not meet criteria. Hence 10 of 17 known to achieve standard.
[9] Japan. No data available for 7 systems. Six provided data and did not meet criteria. Hence 1 of 14 known to achieve standard.
[10] Japan. No data available for 7 systems. Six provided data and did not meet criteria. Hence 1 of 14 known to achieve standard.
[11] Japan, Poland. No data available for 7 systems. Five provided data and did not meet criteria. Hence 2 of 14 known to achieve standard.
[12] All except United Kingdom and Canada (New Brunswick: French). Hence 10 of 12 known to achieve standard.
[13] All except United Kingdom and Canada (New Brunswick: French). Hence 10 of 12 known to achieve standard.

The IEA Studies

Four IEA studies dating back to the mid- 1960s are reviewed here, representing the historic core of international surveys of student achievement in mathematics and science. Another IEA mathematics and science study (The Third International Mathematics and Science Study) is to be fielded in 1994 and 1998. Other cross-national IEA research not discussed here include studies of reading literacy, reading comprehension, literature, French, English, early childhood education, computer use, and civics and classroom teaching practice.

The lEA holds a unique leadership role in the international testing community. IEA was the first entity to develop and administer student achievement tests in more than one country. These studies have attempted to explore almost every aspect of the elementary and secondary school curriculum. The surveys have led to important improvements in largescale international sampling methodology, conceptual design, test administration, and data analysis. Because the surveys were developed as research projects, typically without clear financial support, they were consistently underfunded and even completing the achievement testing process required extraordinary effort and commitment on the part of the IEA researchers. The studies were originally designed to support comparative intimation al research, and while there was an interest in linkages to policy, the work did not explicitl y serve the diverse needs of policy makers, Since attention was drawn to the surveys, however, in *A Nation at Risk,*[12] enormous policy attention has focused on them.

The IEA is an independent international cooperative, funded through a variety of public and nonprofit sources with the participation of education research centers in near] y 50 developed and developing countries. Organized as a consortium of Ministries of Education, university education departments, and research institutes, projects are undertaken by international coordinating centers around the world, and are coordinated by IEA's small central staff. Most activities are undertaken on a highly decentralized basis with modest institutional oversight. The agenda of the IEA is to study systems of education from an international comparative perspective, focusing on five key issues: [13]

1. The curriculum and its effects on education outcomes;
2. School and classroom organization and its effects on education outcomes;
3. The relationship between achievement and attitudes;
4. Educational attainment among special populations; and
5. The relationship between changing demography and changing student achievement levels.

In addition, the lEA provides technical assistance to developing countries attempting to improve their educational research capabilities.

While IEA studies were not originally designed for or intended to be used specifically for purposes of ranking student achievement cross-nationally, collecting data from many

[12]citing the work of Barbara Lerner, *A Nation at Risk* described how poorly American students had performed on international achievement surveys.

[13]As described in T.N. postlethwaite, "Introduction," *Comparative Education Review* 1 (1987), 7-9; and T.N. Postlethwaite, "Comparative Educational Achievement Research Can It Be Improved?" *Comparative Education Review* 1 (1987), 150-58.

educational systems with identical test instruments has inevitably intensified interest in comparisons of the relative performance of one nation's students with the students of other nations. Perhaps unintentionally, issues of country rank have come to dominate discussions of the IEA survey results. Further, given the increasing interest in matters of international economic competitiveness in the United States, attention to this aspect of the IEA agenda continues to grow. At this point such comparisons are unavoidable, and IEA researchers now recognize that comparisons of achievement and country rankings are fundamental to their work. However, they continue to promote efforts to better understand many other factors affecting student performance.

The First International Mathematics Study *(FIMS)*

Purpose. Conducted in the mid- 1960s, the First International Mathematics Study was the IEA's initial attempt to identify factors associated with differences in student achievement. "The main objective of the study [was] to investigate the 'outcomes' of various school systems by relating as many of the relevant input variables as possible.. .to the output assessed by international test instruments."[14]

Mathematics was selected as a first area of study by the IEA because it was recognized as central to every nation's curriculum. Further, "most of the countries involved in the project were concerned with improving their scientific and technical education, at the basis of which lies the learning of mathematics."[15] Lasfly, the IEA felt that mathematics was a logical first subject area for study because it seemed "less difficult" to achieve agreement on the nature of the curriculum appropriate to examine and to develop acceptable test instruments in a cross-national setting.

Participants and survey *content.* Two age groups were surveyed:[16] *students at the grade level at which the majority of pupils were age 13* (U.S. 8th grade) from 12 educational systems; and *students in the last year of secondary education* (U.S. 12th grade) from 12 educational systems. At the secondary level, studies were conducted of students taking mathematics (from 11 systems) and students not taking mathematics (from 10 systems). More than 133,000 students, 18,500 teachers and head teachers, and 5,450 schools in 12 countries participated in the study.

Thirteen-year-olds were tested in the following areas: [17] basic arithmetic, advanced arithmetic, elementary algebra, intermediate algebra, Euclidean geometry, analytic geometry, sets, and affine geometry.

Two tests were derived for the last-year secondary population, one for those studying mathematics, and another for those not studying mathematics during the year of testing, Both groups were tested in the following areas: basic mathematics, advanced mathematics, elementary algebra, intermediate algebra, Euclidean geometry, analytic geometry, trigonometric and circular functions, analysis, probability, and logic. Those studying mathematics were also tested in calculus.

---

[14]Torsten Husen, *International Study of Achievement in Mathematics: A Comparison of Twelve Countries,* Vol. I (New York: Wiley, 1967), 30.
[15]T.N. Postlethwaite, "International Association for the Evaluation of Educational Achievement—The Mathematics Study," *JournaIforResearch in Mathematics Education 2 (1971): 70.*
[16]See Figure Figure II. 1.
[17]For a complete description of each content area, see T.N. Postlethwaite, 105–7.

The instruments consisted of 10 versions of a l-hour test. Each version included a subset of items from a pool of 174 mostly multiple choice items, graded in difficulty. Supplemental questiomaires were developed to explore student views of teaching practice and instruction in mathematics (22 items) and effective outcomes (43 items). Separate questionnaires for teachers and school administrators examined characteristics of the teaching environment at each school surveyed and those of the general educational program.[18]

Sample design andjield outcomes. Each participating entity established a center that was responsible for deriving a sampling procedure in accord with IEA guidelines and that met the approval of an international referee. 19 Two- or three-stage stratified probability samples were drawn in which schools were first stratified by type, and in some countries by geographical or administrative area (e.g., U.S. school districts) and rural–urban location.[20]

The First International Mathematics Study represents the early legacy of the IEA survey experiment. Published material reflects the monumental effort required to organize and accomplish the research and to develop an analytical model. However, the details of the sample procedures and execution results are sparse. Data on the sample design were largely unavailable in any of the published sources, and response rates are unknown (see Appendix tables A. 1–A.3). In addition, descriptions of sample exclusions and the effects of exclusions or refusals on the sample are unknown. Husen flags a serious problem, for example, but does not elaborate: "There are several cases in which the number of schools and pupils is quite small, and the results should very possibly be discounted. In the terminal mathematics group, there were only 222 pupils from France and 146 from Israel, two of the four countries with the highest means."[21]

It is possible that response rates were calculated in unpublished work associated with the study (especially individual country reports). However, the FIMS scores and rankings must be read with caution because the field outcomes cannot be examined and the quality of the data cannot be assessed.

The First International Science Study (FISS)

Purpose. The First International Science Study was one part of a larger research project formally known as the "Six Subject Survey" conducted by the IEA from 1966 through 1973. (The six curriculum areas were science, literature, reading comprehension, English as a foreign language, French as a foreign language, and civics.) The purpose of the Science Study was to assess students' scientific knowledge and to measure their ability to understand the nature and methods of science. The IEA had hoped to evaluate science curriculum reform (that is, the effects of innovative science programs) on achievement in science (especially the impact of "active learning" related to school science laboratory work). However, because it proved difficult to design instruments to evaluate laboratory.

---

[18]Husen , *Achievement in Mathematics,* VOL II, 47-50.
[19]Husen, *Achievement in Mathematics,* Vol. I, 40.
[20]Ibid., Chapter 9.
[2 l]Torsten Husen, *International study of Achievement in Mathematics: A Comparison of Twelve Countries,* Vol. II (New York Wiley, 1967), 27.

skills, most of the analyses focused on understanding the impact of home background, school, and attitudinal variables on achievement.[22]

Participants and survey content. Three populations were surveyed:[23] *students at age 10* (U.S. 5th grade) from 17 educati~al systems; *students at age 14 (U.S.* 9th grade) from 19 systems; and *students in the last year of secondary education (U.S. 12th* grade) from 19 systems. There were 137,000 students and 26,000 teachers from 6,900 schools participating in the First International Science Study.

The following content areas were tested: earth science (only 10-year-olds were tested in this subject); biology; physics; chemistry; nature and methods of science; and understanding science (only 14-year-olds and last-year secondary students were tested in the last two subject areas).

Students also completed attitudinal surveys, Younger students were asked about their interest in science. Middle- and secondary-level students were asked more comprehensive batteries concerning their interest in science, attitudes toward school science, attitudes toward science in the world, description of science teaching from textbooks, and description of science teaching in the laboratory. Teacher and administrator surveys explored cumiculum coverage and teaching practice.

Across the sampled populations, the tests and surveys varied in design.

- Tests for the 10-year-olds (two versions that were randomly assigned to test takers) ran for 30 minutes and consisted of 20 items each. Most of the items did not involve questions specific to science instruction, and 11 items overlapped with those administered to the 14-year-olds.

- Tests for the 14-year-olds (also in two versions that were randomly assigned to test-takers) ran 60 minutes and consisted of 40 items. Eleven items overlapped with those administered to the younger population, and 20 with those administered to the older population.

- Tests for those in the last year of secondary school were subject specific (biology, chemistry, and physics), ran for 60 minutes, and consisted of 40 questions each.

- Attitudinal surveys included 22 items for the youngest population and 48 for the two older groups.[24]

Sample design and field outcomes. As with the First International Mathematics Study, each participating country established a national center responsible for sampling and testing. An international referee approved each country's sampling plan. Depending on the size of the country's school-age population, two- or three-stage stratified probability samples 'were drawn. IEA did not have funds with which to monitor sampling programs,

---

[22] David Walker, *The IEA Six Subject Survey: An Empirical Study of Education in Twenty-One Countries (New* York: John Wiley, 1976), 22; L.C. Comber and J.P. Keeves, *Science Education in Nineteen Countries (New* York John Wiley, 1976), 286.
[23]See Figure II.1
[24]for a substantial review of instruments and procedures, see Comber abd Keeves, *Science Education* Chapter 2.

so it is not possible to determine whether all countries adhered to established procedures, except insofar as particular nations reported deviations.[25]

The Fimt International Science Study offers a relatively complete description of field outcomes. With regard to response rates (Appendix tables A.6-A.8), if the NCES response rate guidelines were applied to the survey of 10-year-olds, 10 of 17 educational systems reported response rates below the 85 percent response criterion (1 of these 10 did not provide sampling information), including 2 among the 5 educational systems with the highest mean scores.[26] Among the 14-year-olds, 11 of 19 educational systems fell short of the criterion (1 of these 10 did not provide sampling information), including 1 among the 5 participating systems with the highest mean scores.[27] Among those in the last year of secondary education, 14 of 19 systems reporting response rates fell below the NCES guideline (1 of these 14 did not provide sampling information), including 3 among the 5 systems with the highest mean scores.[28] In no case did the U.S. samples meet the guideline.

Other aspects of the sample were problematic. Ten- and 14-year-olds were not sampled in the same way in every country. Some countries sampled by grade, finding it too difficult and too costly to sample by age. As a result, some significant differences existed in the construction of individual country samples in terms of the proportion of the target *age group* effectively excluded by grade sampling.[29]

A more complicated problem arose in the sample of students in the last year of secondary school. Participating systems agreed that only those enrolled in school when the survey was administered would be tested and that no attempt would be made to test those who, for whatever reason, were not attending school. This has precipitated an ongoing debate over the import of student retention practices in relation to the high school samples and survey achievement scores. These retention rates varied dramatically from country to country at the time of each of the four IEA studies, especially the First International Science and Mathematics Studies (see Appendix C). An important aspect of the data in this appendix is the sharp increase in retention rates among many countries over the time span of these international assessments.

Documentation on the First Science Study sample affords a clearer picture of the sampling process and the difficulties encountered in trying to establish common sampling practices across participating countries; in trying to define a target population in a way that enables each country to successfully design and execute comparable samples; and, perhaps most important, in trying to persuade schools to participate in this type of voluntary testing program.

---

[25]Walker, *Six Subject Survey, 26.*
[26]Belgium (Flemish), United StateS.
[27]Federal Republic of Germany.
[28]Federal Republic of Germany, Netherlands, Scotland.
[29] Some cpuntries excluded students who were 1 or more years behind in grade for their age (e.g., Chile, Hungary, and Italy for 10-year-olds and Chile and Hungary for 14-year-olds); India only sampled the six
states in which Hindi is the official language; Israel excluded 14-year-olds not attending school and all Arabic-speaking students; Belgium excluded students at the secondary level attending vocational schools; and Thailand only sampled the area around Bangkok.

The Second International Mathematics Study (SIMS)

Purpose. As compared with its predecessor, the Second International Mathematics Study (SIMS) was a more ambitious and complicated project, reflecting a significant amount of learning about the possibilities of large-scale, cross-national achievement surveys. Conducted during the 198 l-8%chool year, the purpose of the project was

> **to** compare and contrast, in an international context, the varieties of curricula, instructional practices and student outcomes, both attitudinal and cognitive. By portraying the mathematics program and outcomes of each participating system against a cross-national backdrop, each system is afforded an opportunity to understand better the relative strengths and shortcomings of its own endeavors in mathematics education.[30]

Participants and survey content. Two groups Were surveyed:[31] *students at age 13* (U.S. 8th grade) from 20 educational systems; and students "who are in the normally accepted *terminal grade of the secondary education* system *and* who are studying mathematics as a substantial part [approximately 5 hours per week] of their academic program" from 15 systems. The United States, along with a smaller subsample of 8 systems, also participated in a longitudinal study designed to assess growth in skills during the course of the school year.[32] To enable attribution of particular outcomes to teacher practices and classroom processes, the longitudinal study pre-tested students early in the school year, post-tested them at the end of the school year, and asked teachers to complete comprehensive process questionnaires during the year.[33]

Thirteen-year-olds were tested in five content areas: arithmetic, algebra, geometry, measurement, and statistics. Content areas for the last year secondary tests were sets and relations, number systems, algebra, geometry, functions and calculus, and probability and statistics.

All 13-year-olds were administered the same 40-item core test and also one of four other tests consisting of 34 (or 35) items. A total of 176 items were available. Students in the last year of secondary education were administered two of eight tests of 17 items each from a set of 136 items. In both samples, items from the available pool were randomly assigned within content areas of each version, and test versions were randomly assigned to students.

In addition to the achievement tests, three other questionnaires were included in the cross-sectional survey:

- Student Background Questionnaire: gathering information about parents (e.g., education and occupation) and about the students' attitudes toward mathematics;

---

[30]David Robitaille and Robert Garden, *The IEA Study of Mathematics II: Contexts and Outcomes of School Madwmucics* (Oxford Pergamon Press, 1989).

[31]See Figure II.1.

[32]Participating countries in the longitudinal study of the younger population were Belgium (Flemish), Canada (British Columbia and Ontario), France, Israel, New Zealand, Thailand and the United States. In addition, Canada (British Columbia and Ontario) and the United States participated in the longitudinal study of the older population.

[33] Findings of the longitudinal study are forthcoming in L. Burstein, ***The IEA** Study of Mathematics III* (Oxford Pergamon Press, 1992).

- Teacher Questionnaire: gathering information about teaching experience, training, qualifications, and attitudes (the longitudinal study also explored instructional techniques); and

- School Questionnaire (completed by the school administrator): concerned with student demographics, teaching staff background, the mathematics curriculum, and aspects of mathematics instruction.

Taken together, these supplemental questionnaires were designed to provide an enhanced contextual analytical base.

Sample design and field outcomes. From the standpoint of sample quality, the Second International Mathematics Study has probably received more attention than any of the other international surveys. A published report by Robert Garden summarizes the samples and sampling procedures in considerable detail,[34] discusses a variety of technical problems with the data, and identifies gaps in the information available.

Appendix tables A.4 and A.5 summarize the response rates. For the 13-year-olds, 12 systems did not provide complete sampling information and 4 others, which did supply outcomes, did not meet the NCES 85 percent response rate standard-for a total of 16 of the 20 participating systems. In other words, 16 of the 20 participating educational systems (including the United States) were either unable to provide response rates at all stages of the sampling process, or had a response rate of less than 85 percent at one or more stages.[35] For example, if one were to apply the NCES response rate standard to the 13-year-old algebra sample, it would raise questions about data from 4 of the 5 systems with the highest mean scores.[36] Among students in the last year of secondary schools sample, 9 of 15 systems reported response rates that fell below the standard, or failed to provide complete sampling infornnation.[37] Looking at one last-year secondary testing area-number systems—2 of the 5 systems with the highest mean scores did not provide sampling information.[38] The U.S. sample had low response rates to the SIMS, although it was evidently better than the previous studies, especially at the school district level (48 percent). Bock and Spencer[39] argue that actual U.S. response rates for samples for both public and private school strata were under 35 percent, when the combined effect of district, school, and class response rates are calculated.[40]

Beyond the issue of response rates, documentation indicated some significant deviations from the definitions of the target populations in different countries.[4]1 From country to country, the age of sampled students also varied considerably. Further, the

---

[34]U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study Sampling Report* (Washington, DC, March 1987).

[35]Belgium (Flemish), Belgium (French), Canada (Ontario), Canada (British Columbia), England and Wales, Finlaml France, Hong Kong, Israel, Japan, New Zealand, Nigeria, Scotland, Swaziland United States.

[36]Japan, Canada, Belgium (Flemish), France.

[37]Belgium (Flemish), Belgium (French), Canada (British Columbia), Canada (Ontario), Hong Kong, Israel, Scotland, Sweden, United States.

[38]Hong Kong, England and Wales.

[39]R. Darrell Block and Bruce Spencer, "On Statistical Standards of the Second International Mathematics Study" (unpublished report, September 1985).

[40]1bid., 26.

[41]For example, the Netherlands did not include 20 percent of the grade 8 equivalent in the sample for various reasons; Nigeria sampled only 8 of 20 states in the country; Hungary inclu&d a broader population

complicated definition of the secondary-level target resulted in substantial sampling inconsistencies across participating entities. (In effect, different countries established different targets within the proposed target description.)

The Second International Science Study (S1SS)

Purpose. Like the Second Mathematics Study, the Second International Science Study (S1SS) was an effort to build on successful aspects of earlier work as well as to expand the scope of the research. The objectives of the study were to describe science education in the participating countries; to examine between-country achievement differences and, where possible, to explain their sources; to attempt to explain differences in achievement between students within countries 42 and to examine changes in achievement outcomes between the two science studies.[43] The study was intended to derive results that could be reliably compared across countries, but there was also a strong commitment to collecting information that would help analysts whp were particularly interested in the status of the science curriculum within countries.

Participants and survey content. Three populations were surveyed:[44] *10-year-olds (U.S.* grade 5) from 15 educational systems; *14-year-olds* (U.S. grade 9) from 17 systems; and *students in the last year of secondary education* (U.S. grade 12) from 14 systems. The eldest population was divided into four subgroups for testing purposes: those studying biology, those studying chemistry, those studying physics, and those not studying a science subject during the test year. Across all educational systems, a total of 260,830 students, 22,612 teachers, and 9,578 schools participated in the study.

For the 10-year-olds, the achievement tests consisted of 24 core items administered to all students, and four tests of 8 items randomly assigned among those taking the test. The achievement test for the 14-year-olds included 30 core items and four groups of 10 items each randomly assigned. For those in the last year of secondary school, there were specialized tests involving 39 items in biology, 39 in chemistry, and 38 in physics. A high proportion of items were taken from the First Science Study.

The Second Science Study included five instruments in addition to the achievement Test:

- Student Questionnaire: gathering basic information including sex, age, grade level, family background, time spent in class on science, and time spent on science homework;
- Attitude Questiomaire and Other Scales: measuring students' perception of science teaching, and verbal and quantitative skills;
- Process Exercises: an optional instrument measuring students' ability to handle science equipment, design experiments, and make observations;

for the terminal year of high school than was called for by the definition; and Scotland sampled two grades, either of which could be considered the terminal year of secondary school.

[42]T. Neville Postlethwaite, *Second International Science Study,* Vol. II Draft (Hamburg, July 1990), 11.

[43]Malcolm J. Rosier, "The Second International Science Study," *Comparative Education Review 31* (1) (1987): 107.

[44]See Figure II.1.

- Teacher Questionnaire (given to those who taught science to the students in the sample): to obtain information on teachers' qualifications and to rate opportunity to learn; and

- School Questionnaire (completed by school principals): profiling student demographics and teaching staff background.

Sample design andjield outcomes. As with the other IEA surveys, depending on the size of the target population, two- or three-stage probability samples were drawn. Appendix tables A.9–A. 13 summarize the sample response rates. All participating systems provided data.

For the 10-year-olds, 7 of the 15 participating educational systems failed to achieve 85 percent response rates at each stage of the sampling process, including 1 system among the 5 with the highest mean scores.[45] Among the 14-year-olds, 7 of the 17 systems failed to achieve the response rate guideline, including 1 system among the 5 with the highest mean scores.[46] The U.S. sample did not achieve the NCES response rate guideline. The last-year secondary samples were extremely complex to draw-involving specialized curricula with little information available as to the number of students or classes making up the targets. About one-half of the countries were unable to provide complete information on each step of the sampling process. Furthermore, samples at the last-year secondary level became very small, and in some cases response rates were exceptionally low. Some countries sampled selected intact classes, and some selected students within classes. Using the biology test as an example, only 7 of the 14 participating educational systems even provided complete information on the sample, and of these only 1 met the response rate standard, thus including only 1 system among the 5 with the highest mean scores. In general, the U.S. sample sizes-of both schools and students—were very small and did not achieve 85 percent response rates.

As reported by Postlethwaite,[47] exclusions were also significant. Less developed countries had very high levels of exclusion, often reflecting the small proportion of children past elementary age who were still enrolled in school. Other countries excluded small schools or tested only in the national language, which were factors likely to influence mean score performance. At the secondmy level, enrollment in school, in science, and in certain science subjects varied dramatically from country to country, making it virtually impossible to ascertain comparability of targets or samples.

From the inception of the survey process, the construction of the U.S. sample was problematic. Sampling lists were available only 6 months before testing. Since time was short, the decision was made not to follow a replacement strategy of drawing parallel sets of schools. Instead, a group of schools was selected that was twice the designed sample size. In other words, since the plan called for a sample of 125 schools, in order to assure an adequate sample size, 250 schools were asked to participate. Such an approach, while safeguarding the final *size* of the sample, does not reduce the problems introduced by selective nonresponse.

The last-year secondary response rates during the first year of testing were very low, and the following year another sample of schools was drawn and tested. Analysis of the U.S. data, however, showed that when items that were on both the first and second

---

[45]Sweden.
[46]Canada (English).
[47]T. Neville postlewaite, *The Second Jnternational Sciene Study,* Vol. II Draft (Hamburg, 1990).

international science study tests were compared, second science students significantly outperformed first science students. This seemed questionable since the results of the National Assessment of Educational Progress showed no comparable improvement in perfommnce over the same time intervai (roughly 1970 through 1983–84). The conclusion was that in the United States the Second International Science Study sample had underrepresented schools in which there would be larger proportions of "poor" performers. An entirely new data collection effort was mounted three years later, based on a completely new sample drawn in 1986. The objective **was to correct** for the underrepresented populations. This "phase two" sample became the official U.S. data set. All the reported scores were based on the phase two test results. The fact that the U.S. data were collected on two occasions raises questions about their utility.

The IAEP Study

The First International Assessment of Educational Progress (IAEP-1): Mathematics and Science

Purpose. The International Assessment of Educational Progress (IAEP-1) is related to another research program-the National Assessment of Educational Progress (NAEP), which has been conducted in the United States periodically since 1969. The initial IAEP, administered in February 1988, was designed to be exploratory in nature (although the results are often discussed as though they were definitive).[48] The IAEP had two objectives: to examine the feasibility of reducing the time and money spent on international comparative studies by capitalizing on design, materials, and procedures developed for the U.S. NAEP; and to permit interested countries to experiment with NAEP technologies to see whether or not they were appropriate for local evaluation projects.[49] Within this framework, the Educational Testing Service argued that the study should be used to "provide teachers, school administrators, policymakers, and taxpayers with information that helps to define the characteristics of successful student performance and suggests areas for possible improvement and change."[50]

Participants and survey content. Six countries (12 educational systems) participated in the study.sl The target population was defined as all students born during the calendar year 197&that is, students ranging in age from 13 years, 1 month to 14 years, 1 month at the time of testing.

The tests were organized around the following topics:

Mathematics: numbers and operations, relations and functions, geometry, measurement, data organization, and logic and problem solving.

[48]A second IAEP study was conducted in the fall 1990 and winter 1991. Findings are to be published in early 1992. IAEP-11 tested mathematics, science and geograph y proficiency among 9- and 13-year-olds. For the 9-year-olds, 18 systems participated in the mathematics **and** science assessment. For the 13-year-olds, 30 systems participated in the mathematics assessmen~ 29 in the science assessment, and 17 in geography.

[49]Benjamin F. King, *A World of Differences: Technical Report, Part I* (Princeton: Educational Testing Service, 1989), 2.

[50]Archie E. Lapointe, Nancy A. Mead, and Gary W. Phillips, *A World of Differences* (Princeton: Educational Testing Service, 1989), 7.

[51]See Figure 11.1.

Science: life science, physics, chemistry, earth and space science, and nature of science.

Test items were drawn from the 1986 NAEP. There were 63 mathematics questions selected from a poolof281 questions and 60 science questions chosen from a pool of 188. All science questions were multiple choice, and 14 of the mathematics questions were open-ended. Each test was 45 minutes in length.

Score comparisons were made on the basis of scales representing levels of proficiency, set to a mean of 500 and a standard deviation of 100. Hence, the study was designed to measure relative levels of competency, in contrast with the IEA research, which did not propose any proficiency measurement scales.

Sample design and field outcomes. The sampling plan called for a multi-stage cluster of 50 pairs of schools, a total of 100 schools, and a sample of about 20 students per school, or about 2,000 students per country. (Small schools were combined with adjacent schools to create "superschools" for sampling purposes.) The general sampling strategy involved two or three stages of selection, with a total among all countries of 24,000 students participating in the study.

Ten of the 12 participating systems achieved 85 percent response rates at eaeh stage in both mathematics and science (Appendix tables A. 14 and A.15).

General Perspective on Samples and Sample Quality

As described in this chapter, representative sampling on past international achievement surveys has been an elusive goal. Cursory review of field outcomes, using information published in conjunction with each of the sampling plans and real shortcomings in field execution. Utilizing the NCES mponse four IEA studies, for instance, suggests that there have been significant deviations from rate guidelines, U.S. data would be excluded from every IEA study at each grade level. The guideline should not be viewed as unreasonable, however. With care in administration and adequate resources, it is achieved regularly on a variety of voluntary, large-scale surveys in the United States. The International Assessment of Educational Progress, in contrast, achieved higher quality field outcomes than the IEA, but the samples were small and few countries participated in the study.

Four conclusions are inescapable:

1. Few educational systems participating in the IEA studies achieved response rates approaching the NCES guideline. Since studies of non-response were not published (and little research on this matter was conducted), the impact of nonresponse on the survey results represents a signiilcant concern.

2. It is not clear that comparable populations have been tested across participating countries.

3. From study to study, country to country, and age group to age group, there is considerable variability in sample quality. As yet, no single standard has been established as a basis against which samples are assessed before data analysis. Some of the variation in quality has to do with the execution of the sampling process, and some is a result of differences in the basic character of the target populations, particularly at the secondary level.

4. In many cases, sample sizes were very small. This should have influenced the design of the analyses and the results reported.

Many countries, including the United States, have had real difficulty achieving high response rates, thereby raising questions about sample representativeness. Until such problems are resolved, interpreting results of the international achievement surveys, the subject of the next chapter, requires caution.

Summary

The five studies described in this chapter are the core group of international achievement surveys of mathematics and science. Their objectives and scope set them in sharp contrast to small-scale studies, or case studies of selected populations or particular communities. Taken together, they represent a significant effort to develop ways of measuring and comparing the determinants of educational outcomes and the performance of educational systems, using modem survey and data processing techniques. Given questions of data quality raised in this chapter, key results of these studies, discussed in the following two chapters, should be viewed cautiously because they are more likely indicative of achievement-related trends and patterns, rather than definitive and conclusive.

Chapter III

The International Achievement Studies: Mathematics and Science Scores

The achievement studies described in Chapter II are undertakings of unusual complexity and scope (some surveys involving more than 100,000 students) anda test of the methodological capabilities of even the most sophisticated researchers. Over a period of 25 years, extraordinary talent, considerable time, and substantial resources have been brought to bear on this relatively new field of study. But the data have proven difficult to analyze and still harder to interpret. Taken together, the body of research is so large that it is hardly amenable to a brief overview of results. In fact, this represents both a strength and a weakness of the international achievement surveys.

Despite the technical issues cited in the preceding chapter, the international surveys do help determine "where we stand" in mathematics and science achievement—that is, the performance of American students as compared with students from other countries. The studies also suggest some of the possible reasons why these differences in performance occur. The studies are useful because of the consistency of many of their findings and the internal relationships identified, and because they frequently corroborate education and social theory. The focus of this chapter is on where we stand, while the next chapter discusses other key results.

Many of the results are study-specific (i.e., not corroborated for the same subject area in other studies and substantiated by another study's findings only occasionally). Even within a single study, findings for one population may be unique. As discussed in the next chapter, there are many reasons why this may be so; nevertheless, this fact constrains the way in which the material is reviewed here.

Beyond recognizing that the same educational systems did not participate in each survey, and that the sample targets and survey objectives distinguish these studies from each other, several caveats—ones that dictate against simple interpretations of the mathematics and science scores—should be mentioned.

- Participating educational systems were self-selected. International achievement studies do not offer comparisons of students from the same educatiomd systems or comparably aged students from survey to survey. Since participation in each study was voluntary, the reported rankings do not represent the U.S. standing among all nations of the world or even among all developed nations, but only among those who chose to participate in each study.

- Sample quality has much to do with the level of confidence one can have in the scores reported. As described in Chapter II, much of the data is technically problematic; hence, the scores must be viewed with caution.

- As noted in the preceding chapter, there is no consistency in how the sampled populations were defined. Different studies tested students of different ages, and participating educational systems did not consistently apply uniform sampling criteria. It is not always possible to characterize performance of “an age group” or “a grade level” across (or even within) studies.[52]

- Educational researchers, cufliculum specialists, and psychometrician have devoted extraordinary effort to developing instruments that could be used in every country participating in the international assessments. Countries are not scored against their own curriculum, and the scores are not adjusted based on differences in curriculum. The approach that has been agreed upon has the advantage of comparing performance against a common standard, derived through consensus and not bound by national curricular differences. But this procedure raises some questions. Even a cursory review of IEA national committee reports[53] indicates that in each country there are some categories of items tested that are not taught at all; some that are of low priority; and some that are entirely outside the instructional objectives for a particular age or grade grQup. It would have been helpful to policymakers if mean performance scores had also been measured and repofled against national curriculum (called the “intended” curriculum by the IEA). This approach would document how each country’s results measure against its own instructional objectives. Presentation of results in these two ways would have answered the dual issues ofi 1) how well do students perform (clearly affected by the differences in the curriculum); and 2) how well do students learn what they have been taught.

- The international testing community has devoted considerable attention to ascertaining curriculum differences (the “implemented” curriculum by the IEA) among countries participating in the achievement studies. A persistent problem, however, is how to account for these differences in the reporting of test scores. Kenneth Travers, for example, discusses the extent to which items on the Second Mathematics Study are reportedly taught in each participating country.[54] For the 13-year-olds, on a topic-by-topic basis,[55] “opportunity to learn”[56] for items on each test ranged from 31 percent of the tested items in some countries to 95 percent in others (the U.S. range was 44 percent for iterns tested in geometry to 87 percent for arithmetic). For the last-year secondary sample,[57] “opportunity to learn” ranged from 29 percent of items in some countries to 100 percent in others (the U.S. range was 46 percent of items on the probability and statistics test to 88 percent on the algebra test). The Travers findings signal a critical issue. The theory of opportunity to learn is a major contribution of the IEA research, but it is not taken into account in the summarized presentations of mean scores and country

---

[2]For an effort in this direction, see John Keeves, cd., *The IEA Study of Science III: Changes in Science Educatwn and Achievement 1970 to 1984* (Oxford Pergamon Press, 1991).

[53]See Chapter 2 for a description of the procechues used by the IEA to &fine test content.

[54]Kenneth Travers, “The Second International Mathematics Study: Overview of Major Findings,” (unpublished paper, Urbana-Champaign: University of Illinois, 1986). All the other IEA studies, as well as the IAEP, grapple with the problem to one degree or another.

[55]Ibid., 36.

[56]“Opportunity to learn” is an issue of considerable concern to those who attempt tO &velop achievement tests for cross-national studies. The concept means attempting to recognize differences between test content and curriculum (especially difficult to estimate for the United States, which has no centralized education authority). In the IEA studies every question on each achievement testis evaluated by a sample of teachers, who are asked to rate the probability with which students taking the test will have been taught the material necessary to answer the question correctly. It is assumed that teachers are in a position to know. (In fact, teachers in one grade might not know with assurance the substance of coursework from other grades.)

[57]Ibid., 45.

rankings. To do so could possibly affect country rankings and provide a counterpoint to measured mean scores.

These notes, along with the sampling issues discussed in Chapter II, are essential to the overview of test results that follows. The mean scores must be viewed with caution. While the scores offer a general perspective on the performance of American students in comparison with students from other countries on mathematics and science achievement, they must be carefully qualified.

International Achievement Test Scores: Interpreting the Results

While researchers have argued that the international achievement surveys are not designed to be an academic olympics, the general public has been exposed to little more than the test scores and country mnkings, leading to an inevitable overinterpretation of their meaning and import. A1though the test results may be viewed as a kind of indicator, this cannot be done responsible y without also understanding the methods used to collect and report the data and the degree to which samples of students either represent or fail to represent sampled cohorts.

Appendices B and E summarize the achievement scores survey by survey. Scores are reported in two ways, Appendix B shows, in tabular form, the measured means and comparisons of the mean scores of each participating educational system with the U.S. score. Appendix E calculates the confidence intervals[58] for all countries and graphically shows comparisons with the United States.[59] While the confidence intervals lack the precision implied by the means, they represent a reasonable reporting framework since they presume no greater accuracy than the data permit. The groupings in Appendix B show those participating educatiomd systems whose mean scores are higher, lower, or within the same range as the U.S. score. These comparisons are summarized in tables III.1 and 111.2.

For those unfamiliar with the statistical issues underlying the presentation of these tables, a brief note may be helpful. Because sampling techniques are used, it is not always possible to say whether the actual mean scores for some educational systems differ statistically from those of the United States. Even though the measured mean of one country may be higher or lower than that of the United States, the difference may not be statistically significant. As a result, the rank ordering of the United States could be different from that which is suggested by the measured mean score. Thus, if there are several countries whose measured scores are not significantly different statistically, this suggests that the sample size was not large enough to know for certain that the actual scores are diffenmt; any one country might actually have the highest or lowest score. For example, in looking at Appendix B, table 4 (from the Second Mathematics Study), the measured scores of seven other countries are not significantly different statistically from the United States, when the U.S. mean score is compared with the scores of other countries. While the measured mean score suggests that the United States "ranks 10th," statistically spedcing,

---

[58]Contidence intervals are estimated by mean ±1.96 x SE, except for the First Mathematics Study as noted in Appendix B. Standard errors are drawn from the study reports themselves. Methods of calculation were not always reported.

[59]Using Bonferroni adjustments, countries were compared with the United States, and scores were categorized as higher, the same, or lower than the United States, based on the results oft-tests at a 5 percent significance level. In Appendix E, in general terms, based on sampling error estimates, 95 percent of the time this range will include the actual country mean score between the upper and lower end of the range defined in the figures. Exactly where the actual mean score for the population falls in the range is not known, although the measured mean for the sample is shown.

Table III.1—International achievement test scores summary

	Number of participating educational systems[1]	Number of participating countries	U.S. rank by measured mean score	Number of participants significantly higher than U.S.	Number of participants not significantly different from U.S.	Number of participants significantly lower than U.S.
First Mathematics Study (IEA)						
Age 13–Core test	12	12	11	9	1	1
Last-year secondary–Math students	12	11	12	11	0	0
Last-year secondary–Non-Math students	10	10	10	7[2]	0	0
Second Mathematics Study (IEA)						
Age 13–Arithmetic	20	18	10	5	7	7
Age 13–Algebra	20	18	12	7	8	4
Age 13–Geometry	20	18	16	10	5	3
Age 13–Measurement	20	18	18	17	0	2
Age 13–Statistics	20	18	8	2	12	5
Last-year secondary–Number systems	15	13	12	9	3	2
Last-year secondary–Algebra	15	13	14	11	3	0
Last-year secondary–Geometry	15	13	12	10	4	0
Last-year secondary–Calculus	15	13	12	10	3	1
First Science Study (IEA)						
Age 10–Core test	17[4]	16	4	1	5	5
Age 14–Core test	19[4]	18	7	5	5	3
Last-year secondary–Core test	19[4]	18	14	10	3	0
Second Science Study (IEA)						
Age 10–Core test	15	15	8	5	4	5
Age 14–Core test	17	17	14	10	5	1
Last-year secondary–Biology	14[3]	13	14	12[3]	0	0
Last-year secondary–Chemistry	14[3]	13	12	9[3]	2	1
Last-year secondary–Physics	14[2]	13	10	7[2]	1	3
International Assessment (IAEP)						
Age 13–Mathematics	12	6	12	10	1	0
Age 13–Science	12	6	9	8	3	0

[1]In some countries more than one province participated, or more than one language group in the same country participated as a separate testing entity.
[2]Data not available for 2 participating educational systems. See Appendix B, E.
[3]Data not available for 1 participating educational system. See Appendix B, E.
[4]Data not available for 5 participating educational systems. See Appendix B, E.

SOURCE: See Appendix B.

Table III.2—Number of other participating systems scoring significantly higher than the United States by age or grade and number of participating systems

	Age 10	Age 13	Age 14	Last-year secondary
First Math Study		9 of 11 (Core test)		11 of 11 (Math students) 7 of 9 (Non-math students)[1]
Second Math Study		5 of 19 (Arithmetic) 7 of 19 (Algebra) 10 of 19 (Geometry) 17 of 19 (Measurement) 2 of 19 (Statistics)		9 of 14 (Number systems) 11 of 14 (Algebra) 10 of 14 (Geometry) 10 of 14 (Calculus)
First Science Study	1 of 11 (Core test)[2]		5 of 13 (Core test)	10 of 13 (Core test)
Second Science Study	5 of 14 (Core test)		10 of 16 (Core test)	12 of 12 (Biology)[3] 9 of 12 (Chemistry)[3] 7 of 11 (Physics)[1]
IAEP		10 of 11 (Mathematics) 8 of 11 (Science)		

[1]Data not available for 2 additional participating educational systems.
[2]Data not available for 5 additional participating educational systems.
[3]Data not available for 1 additional participating educational system.

SOURCE: See Appendix B.

the real picture is less definitive. The United States could actually rank anywhere from 6th to 13th.

In examining each study’s results, note again that the surveys did not test the same subject matter in the same way from 6ne study to another. This precludes representing trends beyond the very general. Nor were the same age and grade levels tested from study to study, further inhibiting across-study comparisons.

Among countries participating in the international studies reviewed here (the United States is the only country to have participated in them all), there is considerable movement in mean score and rank within age/grades and between subjects. Japan is a rare exception, ranking at or very near the top in almost every test. In some cases, U.S. performance is clearly low relative to that of other educational systems, but it is sometimes near the top or in the middle relative to other participants (Appendices B and E).

Summary

This brief description of results on the international mathematics and science surveys is not intended to obscure the general point that students from the United States have not performed very well on any international achievement study. At the same time, the reality is somewhat less clear than the picture that has been conveyed in the media. Generally, across the surveys, younger American students seem to perform better, relative to their international peers, than those enrolled in the last year of secondary school. Even here, however, using caution is essential, because the second~ school populations upon which the survey samples are based differ dramatically from country to country.

The next chapter describes findings associated with achievement that hold across the international surveys and also identifies other findings linked to achievement that are unique to a single subject area, age group, or study.

Chapter IV

What We Know About the Achievement Scores and Country Rankings: A Summary of Selected Results and Hypotheses from the International Surveys

The five mathematics and science studies have involved students from many countries at several grade levels. Taken together, these surveys offer an important perspective on differences in achievement across educational systems.

There is one consistent message. Students from the United States, regardless of grade level, generally lag behind many of their counterparts from other developed countries in both mathematics and science achievement. That, perhaps, is the only consistent message. But caution is necessary. Chapter II identified a variety of technical problems that raise questions about the achievement survey data and make it difficult to know the degree to which sampling and non-sampling errors may bias the results reported. The discussion in the preceding chapters, the standard error tables in Appendix B, and the achievement scorerelated confidence intervals in Appendix E all demonstrate how problematic it is to attempt country achievement score ranking comparisons. However, the consistency of the results across studies and populations suggests that there is an important underlying theme of lagging U.S. performance.

Although a number of hypotheses have been offered, international surveys have been far less successful at explaining why particular groups of students achieve as they do in comparison with students of the same age or comparable grade level from other countries.[60] These studies have not led to consistent conclusions as to why students from other countries perform better academically than their American counterparts, and there are few powerful correlates associated with the overall pattern of achievement across the populations participating in the international surveys.

Despite the technical flaws of the international studies, this chapter examines a number of explanatory issues that have contributed uniquely to our understanding of comparative achievement results. The findings seem to supersede the technical flaws for several reasons. First, as noted above, is the consistency of some findings across studies and age groups with different shortcomings. Second, some of the findings are based on internal relationships identified in the data that are less affected by sampling issues. Third, some of the findings are important because they identify important hypotheses that appear to be supported cross-nationally but may need further exploration. Finally, some of the findings corroborate education and social theory that has been developed based on national studies, thus supporting the basis for these inferences.

In fact, the international data inform a variety of issues, which are not specifically related to the achievement scores and country rankings. To that end, this chapter pursues two lines of inquiry:

[60]See, for example: Curtis M~Knight, Fe Joe Crosswhite, John A. Dossey, Edward Kifer, Jane O. Swafford, Kenneth J. Travers, and Thomas J. Cooney, *The Underachieving Curriculum: Assessing U.S. Mathematics from an International Perspective* (Champaign: Stipes, 1989); and John Keeves, cd., *The IEA Study of Science III: Changes in Science Education and Achievement 1970-84* (Oxford: Pergamon Press, 1991).

- The first looks across the studies and asks, “What results do the surveys report in common?” At this general level, to the extent that there are any cornmonalities, they mostly describe differences in the way schools are organized and differences in mtional education policy and objectives.

- The second explores subject- and age-specific results and hypotheses. Here mathematics and science studies are discussed separately. A certain number of cross-national, cross-test hypotheses regarding the correlates of achievement emerge, although most are subject- and grade-level specific.

Dividing the chapter in this manner provides **a** broader perspective on the international achievement surveys as a body of work, and it raises important tensions in the literature. At the level of subject and grade, there are innumerable interesting, and probably productive, avenues of investigation from the perspectives of rese~her, policy maker, and practitioner alike. At the same time, since the studies have not been conducted with-a consistent focus on a common set of issues, many of the results reported remain uncorroborated across surveys.

The results and hypotheses summarized in this chapter are drawn from published papers. The data have not been analyzed independently. An effort, however, has been made to report those results that have gained general acceptance (or are the focus of ongoing analysis) within the research community.

Results Reported Across the IEA Studies

Across the IEA mathematics and science achievement studies, some systematic patterns of differences have been observed.

1. The more content students are taught, the more they learn, and the better they perJorm on the achievement tests.

While this point may seem obvious, it reflects some important differences crossnationally. From country to country the mathematics and science curricula vary considerably; as a result, students at the same grade level may be taught more or less, and may be taught more or less intensively in a particular subject area. The result is more or less breadth and depth in learning. This proposition represents a theme woven through the IEA research. For instance, it has been shown that, in comparison with higher achieving countries, the American mathematics curriculum tends to be relatively shallow and narrow. A great deal of time is devoted to review and repetition, the work is generally less demanding, and teachers have lower expectations of students.[61] Students learn what they

---

[61]See discussions in Curtis McKnight et al., *The Underachieving Curriculum*; Charles Finn, “Afterword A World of Assessment, A Universe of Data” in *International Comparisons, ed.* Alan Purves (Alexandria, VA. Association for Supervision and Curriculum Development, 1989), 74–81; and R.A. Garden, “The Second IEA Mathematics Study,” *Comparative Education Review* 31 (February 1987): 47–58. Low coverage, measured by “oppxtunity to learn” scores, emerges throughout the IEA data in discussions of differences in achievement among students from various countries. One paper that looks at this issue from a policy perspective is Marshall Smith, “A First Look at the Policy Implications of the Findings of the Second Mathematics Study of the IEA” (paper presented at the National Conference on the Teaching and Learning of Mathematics in the United States, Champaign-Urbana, University of Illinois, 24 September 1984). See also Lain Anderson and T. Neville Postlethwaite, “What IEA Studies Say about Teachers and Teaching,” in *International Comparisons and Education Reform,* ed. Alan Purves (Alexandria, VA: Association for Supervision and Curriculum Development 1989), 74-81.

are taught, and there are significant differences in the content of instruction among countries at common levels of schooling.

2. Although international studies suggest that tracking as practiced in the United States seems to be negatively associated with student performance and student exposure to challenging coursework, some other countries have stronger forms of ability grouping that positively influence their assessment results.

Tracking, or some type of classroom ability grouping, is standard practice in many American schools. Other countries also define the mix of students in schools and classrooms, but often this is not called "tracking." For instance, at the secondary level in some countries, national selection and placement policies filter students into ability groups or otherwise determine which students have access to college preparatory academic programs. This may not represent tracking in the American sense, but it has a simi Iar effect, although *within these* highly selective systems there may be little or no tracking. Nevertheless, all countries with nearly universal secondary school enrollment practice some form of "tracking," whether it is of students into schools, or of students within schools into classrooms. The importance of the distinction is that in countries where "tracking" is into schools, tracking is associated with higher performance levels. But in countries where "tracking" is *within* schools into classrooms, it is associated with lower performance levels.

In terms of the international achievement surveys, tracking and selection practices affect the "pool" of students participating in the international surveys. At the secondary level, survey targets from selective systems tend to be from academic programs. So it is perhaps not *surprising* that highly selective educational systems (which do not track students in the way that American schools do) tend to produce students who perform better on average in the international surveys than students from countries that do track.[62] These studies have not investigated the effects of school and classroom tracking on students performing at lower achievement levels.

Circumstances are different at presecondary levels before selection policies are in evidence. Here students from systems that do not ability group tend to perform better in the aggregate on the international achievement tests.[63] It has been hypothesized that students from some of these countries perform well because there is significant cultural and social homogeneity. *However, data from the international surveys do not enable analysis of this notion, except in very general terms.* In Japan, for example, it has been noted that at the presecondary level virtually all students are exposed to the entire mathematics curriculum, and there is no evidence that students have been sorted.[64]

Curriculum exposure, which is related to tracking, shares a common consequence, as Kifer writes:

---

[62]See Edward Kifer, "what IEA Studies Say about curriculum and Schwl Organization" in *International Comparisons and Education Reform,* ed. AlarI Purves (Alexan&ia, VA. Association for Supervision and Curriculum Development 1989), 71.

[63]William Platt "policymaking and Intermational Studies in Educational Evaluation," eds. Alan Purves and Daniel Levine, *Educational Policy and International Assessments: Implications of the IEA Surveys of Achievement* (Berkeley: McCutchan, 1975); also Kifer, "Curriculum and School Organization,"71.

[64]Leigh Burstein, cd., *The Second International Mathematics Study,* Vol. III, Draft (April 1990), chapters 11 and 13.

. . 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 United States, for instance, less than 15 percent of the students arein a track that will require them 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.[65]

While it is not possible to estimate what proportions of the student population get to experience different types of mathematics instruction, country by country, McKnight notes that eighth-grade American students are found in one of four types of mathematics classes—’’remedial,” “typical,” “enriched,” and “algebra.” Those in remedial classes were taught only about one-third of the algebra on the Second Mathematics Achievement Test, while those in algebra classes were taught almost all of the algebra on this test. More extensive differentiation in curriculum was found at this level of schooling in the United States than in any other country participating in the study.[66]

Tracking is an issue of special interest to American policy makers and educational practitioners. Carefully controlled longitudinal studies in the United States have found a modest to non-significant relationship between tracking and student performance once preexisting differences in student ability and background are held constant.[67] At the secondary level, the apparent negative association between tracking and international performance is obscured by the fact that there are so many different types of policies for dealing with ability differences that the definition of tracking is problematic. At the presecondary level, tracking in the American sense appears to be most directly related to exposure to a particular curriculum. Lack of a common definition of the term *tracking,* applied uniformly across all of the countries participating in each achievement survey, suggests that these results must be viewed cautiously.

3. *The schooling experience affects learning more in some subject areas than in others.*

Certain subjects appear to be school intensive—i.e., more learning and mastery goes on in the classroom than outside of it. Among the many subjects that have been examined by the IEA, some appear to be more closely associated with school exposure than others.[68] The import of schooling appears to be strongest for subjects such as science and much weaker for subjects such as foreign languages. Walker[69] hypothesizes that this might also hold for mathematics, a curriculum in which parents are not necessarily knowledgeable, thereby increasing the school effects.

4. *To the extent that family background characteristics have been captured in the international surveys, they have been shown to have explanatory power crossnationally.*

---

[65]Edward K,ifer, “What IEA Studies Say,”71.

[66]McKnight Underachieving Curriculum, 106.

[67]K.L. Alexander and M.A. Cook, “Curricula and Coursework A Surprising Ending to a Familiar Story,” *American Sociological Review (47) 1982,636.*

[68]See, for example, Anderson and Postlethwaite, ‘Teachers and Teaching.”

[69] David A, Walker, *The IEA Six subject survey: An Empirical Study* of Education in *Twenty-One Countries (New* York Wiley, 1976), “228.

Confirming a well-documented finding in the United States, cross-national studies have demonstrated relationships between family background and achievement.[70] To Americans this point may be well understood. Even though particular background variables may mean different things in various countries, multivariate analyses in several IEA studies show some associations. In the between-schools analysis of all schools in the First Science Study, 33 percent of the explained variance was accounted for by home background variables for the 10-year-olds; 45 percent of the explained variance for the 14-year-olds; and 44 percent for students in the last year of secondary education.[71] In the First Mathematics Survey, a smaller proportion of variance was explained by home background because "opportunity to learn" variables were extremely powerful in that study's multivariate analysis (suggesting the importance of including a broader set of explanatory factors in the model).[72] These data suggest that among more developed countries at least, home background shows some relation to achievement patterns cross-nationally and that this is not uniquely a U.S. phenomenon.

5. *Educational systems committed to keeping students enrolled in school score less well on the international surveys, but they formally educate a larger population. Japan is an important but lone exception to this proposition, calling the simplicity of this link into question.*

Over the three decades of IEA research, the impact of secondary school enrollment policies on achievement patterns has received considerable attention. Data from the studies suggest that countries retaining a large proportion of the eligible age group in seconda~ school (e.g., the United States and Sweden, which both have high levels of school "retention") tend to perform less well cm the secondary-level achievement tests in part because a greater range of student skills and capabilities are represented in the student population.[73] According to this argument, countries with higher rates of student retention are producing more knowledge across a larger population base.[74] The issue of retention

---

[70]Family background variables are discussed in each of the IEA study reports and are often treated **as** groups of variables in multiple regression analyses. See Purves and Postlethwaite, "Teachers and Teaching" and Anderson and Postlethwaite, "What IEA Studies Say." While this conclusion may hold generally across more developed countries, Heyneman and others have analyzed the IEA data along with data from other sources. They report, among other things, that at the country level, the lower the income of the country, the weaker the influence of pupils' social status on achievement and "...conversely, in low-income countries, the effect of school and teacher quality on academic achievement in primary school is comparatively greater." See Steven Heyneman and William A. Loxley, "The Effect of Primary School Quality on Academic Achievement across 29 High- and Low-Income Countries," *American Journal of Sociology 88 (6)* (May 1983), 1162–94; and Steven P. Heyneman,"The Search for School Effects in Developing Countries" (Seminar Paper No. 33, Washington, DC: The World Bank Economic Development Institute, 1986).

[71]See David Walker, *The IEA Six Subject Survey,* 96-97. Four sew of vtiables were included in the analysis: home and background (including proxy SES measures and parents' education and occupation); school type and program (including class size and "opportunity to learn"); and learning conditions and "kindred variables" (attitudes, interests, motivation, out-of-school time use, and so forth).

[72]Torsten Husen, International Study *of Achievement in Mathematics: A* cO??pZriSOn *of Twelve Countries*, Vol. 2 (New York John Wiley, 1967), 286.

[73] The magnitude of negative relationships varies considerably from survey to survey, ranging from marginal to substantial depending on the kinds of analyses undertaken.

[74]For excellent discussions, see David Robitaille and Kenneth Travers, "IntematiOnal studies in Mathematics Education," forthcoming; see also David A. Walker, The IEA Six Subject Survey, 279; M. David Miller and Robert L. Linn, "Cross National Achievement with Differential Retention Rates," *Journal for Research in Mathematics Education 20* (1) (1989), 28-40; Ian Westbury, 'The Problem of Comparing Curriculums across Educational Systems," in *International Comparisons and Education Reform,* ed. Alan

tends to reflect broader educational and social policy objectives. As more countries increase their student retention rates at the secondary level, the issue may lose its power. Or, scores among some countries that previously had highly selective secondary systems may decline relatively as they retain a greater number and variety of students in school. But this is not always the case. In Japan, for instance, secondary school retention has increased dramatically over the past two decades, but Japanese students continue to perform near the highest level on both the mathematics and science surveys. This result appears to be unique to Japan. No other country with rapidly rising rates of student retention exhibits a comparable pattern. Further, these data are difficult to evaluate over time, and there is no way of knowing whether the high levels of performance among Japanese students in the 1980s is as "high" as it was at an earlier point in time when retention rates were lower. However, this significant exception suggests the importance of further research on the issue of whether factors other than breadth of retention have a greater effect on student performance. The issue of school retention and selectivity highlights one of the areas in which sampling age-level cohorts might offer more representative national achievement estimates than grade-level cohorts, at least among sectmdary students. .

6. *Generally, the "best students" in the United States do less well on the international achievement surveys when compared with the "best students" from other countries.*

Although this may reflect the nature of the school population (which is less selective in the United States), it deserves consideration. For example, on the algebra subtest in the Second Mathematics Study, achievement among the top 1 percent of U.S. 12th-grade students was lower than achievement among the top 1 percent of any other country. On functions and calculus, the top 5 percent of U.S. students scored in a lower range than the top 5 percent of students from almost every other participating system.[75] Linn and Miller argue that while retention rates on the Second Mathematics Survey overall accounted for some achievement differences for the more able students, variables such as opportunity to learn were more important in explaining differences in achievement scores across participating systems.[76] On the IAEP, where 9 percent of American 13-year-olds performed at the second highest mathematics proficiency level, 40 percent of Koreans performed at that level.

7. *Students from less developed countries do less well on tests of achievement than students from more developed countries.*

As participation in the IEA studies has increased over the past two decades, the differences between more and less developed countries have become verv clear. This seems

Purves (Alexandria, VA Association for Supervision and Curriculum Development, 1989), 31. In his analysis of data from the first mathematics study, Husen hypothesizes that "higher levels of mathematical achievement will be attained by a smaller proportion of those still in school, but by a larger proportion of the total age group. " Torsten Husen, *International Study of* Achievement *in Mathematics II: A Comparison of Twelve Countries* (Oxford: Pergamon Press, 1967), 128.

[75]McKnight, *The Underachieving Curriculum*, 26, 27. Although it is difficult to control for the effects of selectivity, McKnight writes: "In order to control [for the] selection effects, an analysis was made of the average achievement in algebra of the top l$ZO and top 5% of the age group in each country. The results showed that the U.S. came out as the lowest of any country for which data were available." Miller and Linn, following Husen's analysis in the first mathematics study, developed the procedure for defining and calculating achievement scores of the 1 percent and 5 percent cohort. See M. David Miller and Robert L. Linn, "Cross-National Achievement."

[76]M. David Miller and Robert L. Linn, "Cross. National Achievement" 38-40).

to be associated, at least, with differences between more and less developed countries in curriculum content and in the grade level at which some subject matter is taught. In general, regardless of subject matter or age group, students from less developed countries do not perform well on the achievement tests, even though the in-school population in some of these less developed countries is a small fraction of the age cohort and typically comes from high-status farnilies.[77]

Selected Subject and Age Group Results and Hypotheses from Individual IEA Studies: Linkages to Achievement

Since most of the international surveys were designed and organized by loose administrative consortia, data analyses have not been closely coordinated. As a result, researchers have pursued different agendas, and while there may have been opportunities to corroborate results across studies, this has not often occurred. Hence, many results and hypotheses appear to be associated with a particular study, subject area, or age or grade group-which may or may not be the case—simply because the groups were only tested one time. As the following discussion indicates, it seems important to encourage analytical replication in the future, so that hypotheses will be tested overtime and across subject areas and age or grade groups. This section describes results associated with achievement and three types of variables in individual studies: curriculum, teaching, and instructional methods; student characteristics and family background, and organization of schools and instructional programs.

Curriculum, Teaching, Instructional Methods, and Achievement

The international surveys have identified a number of linkages of performance and curriculum and teaching methods.

The First Mathematics Study found considerable variation in curricula across systems, especially in the timing of instruction in particular topics and concepts. In some countries topics were taught much earlier than in others. (This was particularly true in highly selective enrollment systems.) Consistent with this proposition, the Second Mathematics Study concluded that students learned what they were taught, and those from countries with more demandmg curriculum learned more of the kinds of items tested in the survey, and performed better. In other words, "... achievement follows content . ..." The study also revealed something that many Amerlcans had not supposed possible—that students can be taught complex mathematics at a relatively early age.[78]

In the First Mathematics Study, the "opportunity to learn" variable emerged as an important indicator of performance, especially at the secondary level.[79] In the Second Mathematics Study, "opportunity to learn" was also closely associated with achievement. Among 13-year-olds, American students were more lklv to have had an arithmetic-based curriculum. In other countries the curriculum was more likely to be based on algebra and

---

[77]For a discussion, see D. Spearitt, "Evaluation of National Comparisons," in The *lnternational Encyclopedia of Educational Evaluation,* cd., H.J. Walberg and G.D. Haertel, (Oxford Pergamon Press, 1990), 5 1–59; and A. Inkeles, "National Differences in Scholastic Performance," *Comparative Education Review 23 (1979): 21 1–229.*

[78]*Finn,* "Afterword," 113.

[79]Richwd M. Wolf, *Achievement in America: National Report of the U.S. for the International Education Achievement Project (New* York Teachers College, 1977).

geometry.[80] As a consequence, where some countries were pressing forward with conceptually advanced curricula for students at an early age, Amencan stuctents were still focused on elementary mathematics skill-building. The American cumculum more closely resembled elementary school as compared with the better performing countries in which the curriculum was more like our earlv high school.[81]

In the First Science Study, "opportunity to learn" proved central to understanding achievement score differences, especially for secondary-level students. But no relationship was found for younger students. Even so, it was noted that in some countries, including the United States, at the early and middle grade levels science was often not taught as a separate subject. Instead, science instruction was conducted by regular classroom teachers, many of whom may not have been equipped to build a foundation for specialized science learning in future years.
that success in science was distinctly related to the quality and extent of instruction. Different branches of science instruction were stressed to greater or lesser degrees from country to coun sclentdlc subjects and Ideas when they leave school is… to a large extent m the hands of those who cteslgn the cumcula for the subject.[82]

In the Second Science Study, as with mathematics, science cumicula were distinguished by differences in timing—that is, when particular subjects were offered to students. At the lower and middle grade levels, there was substantial variation in the degree to which specific courses were available in each branch of science. Lower and middle grade students from systems in which more specific scientific instruction was provided performed better on the achievement tests.

Distinctions in curriculum and instructional methods were characterized in other ways. The Second Mathematics Study documented differences in the level of difficulty of the program. Students in the last year of secondary school from systems with higher retention rates, like the United States, were more hkely to be studying algebra or trigonometry and 1ess likely to be studying more comrdex subjects like calculus. Program relatecl differences were also found in the Second Science Studv. In some countries students were *required* to take courses in *each* branch of science *separately;* in other countries they were only required to take general science courses, or they could choose a few courses among those available [limited requirements). When there were separately; in other countries they were only *required* to take courses in *each* branch of science *separetely*; in other countries they were only required to take general science courses, or they could choose a few courses among those available (limited requirements). When there were separate science course requirements, subject matter demands were greater, students were taught morem and they performed better on the achievement tests.

Teacher preparation time also has been examined in relation to achievement. Several studies showed some relationship of teaching time and teacher preparation timact_nevement. 1n both the Second Mathematics and First Science Stuches, the amount of instructlonal preparation time for teachers in and outside of school was related to student achievement. Teachers in the United States and some other countries had little time available during the scnool day to plan for classes, and they did not spend propornonately more ume preparing materials after school.

[80]ORobitaille and Garden, *Mathematics, 238.*
[81]McKnight et al., *Mathematics.*
[82]Walker, *The IEA Six Subject Survey, 232.*

Student and Fmily Background Characteristics and Achievement

Although the data were not consistent with regard to student characteristics, gender was generally associated with achievement across the international studies. Gender differences in performance have attracted attention since the First Mathematics Study, which defined a number of issues. At the most basic level, the proportions of girls enrolled in mathematics courses varied considerably. Further, among the younger population and also by and large at the pm-university level, boys expressed more interest in mathematics than girls.[83] Finally, substantial differences in performance across gender were found among the younger sample population in all countries except the United States and Sweden, where the differences in performance between boys and girls were much smaller.

In the First Mathematics Study, gender differences in achievement scores at the secondary level were greatest in countries with large proportions of single-sex, as opposed to coeducational, schools. At the same time, interest in irmthematics among girls was higher in systems with large numbers of single-sex schools. But, while interest levels may have been higher in these schools, performance was not enhanced.

Gender-related patterns were not consistent, even between the two mathematics studies. In the Second Mathematics Study, among the older cohort, there were many more male than female mathematics students across all participating systems, and boys almost always outperformed girls. Among the younger cohort, girls outperformed boys in some topic areas.

The First Science Study identified still another sex-related pattern. Boys showed a greater interest in science than did girls, a phenomenon that increased with the older cohorts. Similarly, with reference to total science score by educational system, boys outperformed girls at all levels: at age 10 (by about one-quarter of a standard deviation); at age 14 (by about one-half of a standard deviation); and at the last year of secondary school (by about three-quarters of a standard deviation).[84] In subject interests, boys were more likely to be enrolled in physical sciences courses, and girls in biological sciences. By the last year of secondary school, boys generally outperformed girls in all science subjects, but the gap was considerably less in the biological sciences. In the Second Science Study, boys scored higher than girls, and the differences increased from elementary to middle school. In the IAEP, among 13-year-olds, boys and girls performed at about the same level in mathematics; however, this was not the case in science. Except in the United States and the United Kingdom, boys systematically performed better than girls.

Beyond the question of gender and performance, some aspects of family background should also be mentioned. In the First Mathematics Study, student and family background were associated with performance to a greater degree in the United States than in other counties. Particularly among the eighth-grade sample, scores were related to parents' education and father's occupation. In countries other than the United States, the import of these background variables declined at the secondary level, perhaps because enrollment selection policies homogenized student profiles in the later years of school.[85] The IAEP found another relationship among family life, activities outside of school, and science

[83] Husen, *Mathematics,* Vol. 2,305.
[84]Comber and Keeves, *Science Education, 139-53.*
[85]Ibid.. 303.

performance: doing better in science was associated with such variables as parents' talking with childnm about science topics at home.[86]

Organization of Schools and of Instruction and Achievement

One other set of issues, bridging several studies, concerns achievement and the structure ,of schools. While the classroom was not explored in a consistent fashion in relation to achievement variables in the First Mathematics Study, class size did not show a systematic tdationship to performance. (In fact, some of the counties with the largest class sizes produced some of the highest achieving students.) In the Second Science Study, except for the youngest students, class size was not related to achievement. Again, counties with the largest classes tended to have the best scores.[87]

Similarly in the Second Science Study, across countries, school organization variables including total hours of school each week, of mathematics'instruction each week, of homework, and of mathematics homework showed virtually no relationship to the scores of the 13-year-old sample. For students in the last year of secondary school, hours of mathematics instruction and of mathematics homework showed a small positive relationship to achievement.[88]

Summary

This chapter has selectively summarized results and hypotheses associated with the international achievement test scores. The fmt section described na.dts that held across the studies at a general level, while the second section focused on lines of inquiry related to individual international achievement studies. To a degree, the studies are so different in analytical focus that the results reported seem rather eclectic. To the extent that these results can be pursued systematically in future research, policymakers may be able to find more ways of applying international findings on curriculum, instruction, and organization, and achievement to issues of schooling in America.

---

[86]Lapointe, *A World of Differences, 45.*
[87]Post1ethwaite, *Science,* chapters *8* and
9. [88]Husen, *Malhemutics,* 300.

Chapter V

Looking Ahead: Toward Future International Achievement Surveys

After three decades of research, the comparative education community has produced a series of important studies examining differences in mathematics and science achievement among students of different ages from a number of developed and less developed countries worldwide. Differences in achievement have been observed, but their magnitude is uncertain. Because of inconsistencies in sample design and sampling procedures, the nature of the samples and their outcomes, and other factors, it is difficult to know the degree to which the past surveys accurately measure student performance across comparable populations from country to country.

Despite their shortcomings, international achievement surveys are now highly valued, providing a way to explore the import of many schooling inputs and processes that can best be observed cross-nationally. While results of the international assessments document many differences in the nature and organization of educational systems, achievement scores and country-by-country performance rankings have received the most attention. The interest in scores and rankings demands that the data used by U.S. policy makers and educators meet high technical standards.

There is considerable evidence that the various international testing authorities and consortia are moving expeditiously toward improving the quality of the surveys and upgrading their statistical reliability. The National Center for Education Statistics (NCES) supports and encourages these efforts because they serve to enhance the utility of these data for policy makers and education practitioners. Recognizing that the comparative achievement scores and country rankings are likely to become even more visible in the future, it is essential for the design of new international studies to reflect lessons learned from the past.

This concluding chapter discusses a number of issues raised by the National Research Council's Board on International Comparative Studies in Education (BICSE),[89] and it also focuses in greater detail on matters related to the design of international achievement surveys, the ways the results of the international assessments are reported, and the nature of the reporting process itself. Where BICSE outlines broad strategies for the assessment process, this chapter offers a more strategic look at issues concerned with the data.

Areas of Improvement: Sample Comparability and Reporting International Achievement Scores

Congress, the Executive Branch, the media, and much of the general public continue to focus attention on test scores and country rankings described in the international surveys. This poses a real challenge to the comparative education community. Researchers must continue to elaborate and refine the ways in which they measure cross-national achievement and must continue in their efforts to describe why differences occur. But to the extent that the international achievement scores serve as visible "leading indicators" of

[89]See Norman M. Bradbm and Dorothy M. Gilford, eds.,"A Framework and Principles for International Comparative Studies in Education" (Washington, DC: National Academy Press, 1990).

educational competitiveness, it is essential that reports accurately reflect *real* differences in achievement among sampled populations.

Early international achievement survey results were compiled from country data of widely varying quality. If cross-national comparisons are to be scientifically credible and if policymakers are to rely on the findings, then stringent data collection standards must be established and achieved by all participating entities. Two ways of improving data quality in future studies are strengthening sample comparability and adding some adjusted international score reports that might make it easier for audiences without technical backgrounds to accurately interpret findings.

Sample Comparability

A standards' review procedure could help assure that reported findings are based on accurate, representative sample estimates. To that end, it would be useful to examine sampling outcomes from the standpoint of *compatibility* and *represgntativeness* before and after data are collected and before extensive analysis. This would enable researchers to ascertain the degree to which samples represent targets and it would encourage participants to devote more attention to sampling issues during development of the survey process in each country. At least six questions concerning samples and data quality have arisen from the international surveys to date:

1. To what extent did the samples meet the study design requirements?

2. Were there differences among countries in how the target populations and eligibles were defiied? Did each country follow identical procedures?

3. How were modifications to the sample handled? For instance, when countries legitimately sample target populations that are not thoroughly comparable with those of other nations participating in a study, were these noncomparable circumstances articulated, justified, and their implications discussed?[90]

4. Were the response rates adequate on a country-by-country, stratum-by-stratum basis?

5. Did the characteristics of those declining to participate (or excluded from testing) differ substantially from country to country? Within countries, did this affect the degree to which the achieved sample represented the target population? Were the characteristics of schools in the design sample but not in the achieved sample compared, and were the comparisons reported?

6. Did the age distributions of test samples differ substantially. and if so. what were the analytical implications?

Achieving sample comparability represents an important, but still eIusive, goal. As described in the following section, significant change in research designs are being made, and advances are evident on issues of comparability. But at this point, in some international surveys the composition of samples (and of the units sampled) differs substantially from

---

[90]Sampling differences may be important. In the IAEP, the Inner London Education Authority declined to pardcipatq so no testing took place in England's principal, and most demographically diverse, city. In the First International Science Study, only the six Indian states in which Hindi is the official language were

country to country. Two strategic issues underlie the question of comparability-age-level and grade-level sampling. If *age* is the basis for sampling, then all participating entities must assure representative samples of the same age cohort. This becomes complicated, of course, because from country to country one age group may bridge several schooling grades, or in a given country, one age group in school may be more or less representative of a national age cohort. The BICSE report emphasized the importance of defining sampled populations in similar ways and assuring “comparable coverage of the populations.”[91] Further, surveys should be able to “support reasonably accurate inferences about an age or grade cohort, and the proportion of each cohort covered should be carefully estimated and reported. The sample should be designed to ensure that it captures the range of individual, school, or classroom variation that exists in the nation sarnpled.”[92] Sampling age cohorts enables comparisons of particular age groups in each country, but it is more costly given the expense associated with finding and testing students who are out of school and those at different grade levels. In contrast, if *grade is* the basis for sampling, all participating nations should strive to provide comparative information about students who have been in school for the same number of years.[93] Further, grade sampling offers the opportunity to relate classroom characteristics (e.g., classroom processes and teacher practice) to student performance in ways that would not be possible with an age-based sample.

Solving problems like those associated with age versus grade cohort testing represents a signii3cant concern in terms of deriving samples that are analytically *equivalent* across all participating countries and meeting the intended purposes of the assessments. Evidence from the IEA Reading Literacy Study (see below) suggests that this dilemma is well recognized and that steps are being taken to improve prospects of sampling comparability y or, at least, assure that minimum standards are achieved on future IEA achievement surveys. In all cases, the objective should be to assure accurate comparisons of achievement between countries across school and age cohorts, even within the context of different policies for selecting and retaining students in school.[94]

---

sampled. In both cases, results were reported, and national data were used for international comparisons without a discussion of the analytical implications.

[91]Bradburn and Gilford, “Framework” 9.

[92]Ibid., 25.

[93]1n the BICSE report, the following is noted “. . .it is not clear whether students should be tested according to their age or their years in school. Children start school at different ages; first graders may be 5, 6, or 7 years old . . . . Grade progression also occurs at different rates across countries. Some of the Nordic countries have policies against repetition. Thus, if one were interested in evaluating achievement at about the transition between “lower” and “middle” school, should one test fourth graders or 9-year-olds? In comparing systems with different age rules for school entry, there may be quite large differences in the average age of students . ...”Bradbum and Gilfor~ “Framework,” 8.

[94]The ongoing Education Indicators Project at the OECD raises the question of comparability in data collection”. ..given the possibility of widespread system differences.” B y way of example, the comparability issue is clearly articulated in the following:

> Nations differ in the pattern and intensity of their science instruction. Some prefer exposing secondary students to an army of scientific subjects. Others choose to immerse secondary students in one or a relative few subjects. [Of course, some nations may offer no science at all.] An international comparison of secondary school biology or physics knowledge, in the absence of information regarding learning opportunities for students, might lead to inaccurate conclusions regarding a nation’s school effectiveness or students’ ability levels. Similarly, some nations delay the onset of formal instruction until a later chronological age than others. Hence, assessing reading ability at an early age may provide a misleading comparative picture of a particular nation’s educational achievement level. (Organization for Economic Cooperation and Development, “Assessing Assessments: Considerations in Selecting Cross-National Educational Performance Indicators,” draft report November 1990.)

Reporting Achievement Scores

Beyond comparability of data, there is the equally important issue of how data are reported and what is reported. Much that has been learned about the international achievement survey results suggests that mean score and country ranking reports require careful qualification, elaboration, and provision of context for interpretation.

Among other things, it is important to discuss factors likely to influence country

- Where *systems of education and fundamental nationul policies* affect mean scores and rankings, these differences should be accounted for in the reporting process. For instance, at the secondary level, national or local school retention policies in and of themselves are apparently related to achievement scores.

- Where *differences in curriculum* positively or negatively affect students who are taking the international tests, these differences must be articulated. The curricula in some systems are quite consistent with the elements tested on the surveys. In these instances, students are likely to answer more questions correctly. This means that *a priori"* students from some countries are likely to do better on the tests than their peers from other countries. Using the opportunity to learn indices, curriculum advantage should be reported along with test scores.

- Where the *test formats* themselves may affect outcomes, these need to be investigated and discussed. For example, students from some countries may be more or less familiar with achievement testing generally and with the particular formats used in comparative studies. To the extent that they are known, reports must account for such differences. Further, there are countries in which students are exposed to a great deal of testing, and, as a result, participating students may expend less effort on "low-stakes tests" that they believe do not affect their educational futures. Thus, in reporting data, researchers must consider the potential consequences of student indifference.[95] These kinds of issues could be addressed in the repofing process, clarifying fundamental differences across participating entities that may be associated with scores and rankings.

In addition, new data dissemination formats could be constructed as a way of moving beyond comparisons of measured mean scores. The following might be two options:

- Developing sets of scores that would reflect each system's achievement against items common to its curriculum.

- Developing sets of scores against some minimum or optimum performance standards, agreed upon by all participating educational systems for the purpose of defining the proportions of students achieving at or above that level for a given age or grade level. This approach would, admittedly, be more difficult because of the problem of achieving international consensus on such a matter, but BICSE supports alternatives of this sort with appropriate caution:

---

[95]Bradford Gilford 'Tramework" 31.

> Studies concerned with student achievement data can be enhanced considerably by reporting outcomes in terms of performance standards, for example, the percentage of students who know everyday science facts or who use scientific procedures and analyze scientific data. This can be difficult to accomplish, however, and there is a risk that arbitrarily established standards will lead to serious misinterpretations of achievement levels. If results are reported relative to specified performance levels (e.g., functional literacy), the basis for establishing these levels must be explicit, defensible, and responsive to the needs and contexts of all the nations involved.[96]

- Providing standard errors or confidence intervals for all estimates.

In general, as noted by BICSE, reporting should be

> sensitive to technical limitations on a study's interpretability. Limitations might include caveats about the comparability of national samples, the limited number of test items or range of content on which comparisons are based, differences in administration conditions from place to place, the match of tests to different curricula, the difficulty of translating exercises from one language to another, the limited precision of sample statistics, or other qualifications on study findings.[97]

A Place for Small-Scale Studies

Small-scale, intensive case studies can enrich the presentation and interpretation of data from the large-scale international achievement studies. Harold Stevenson's study of first and fifth graders in Minneapolis, Taipei, Taiwan, and Sendai, Japan exemplifies a way of describing the correlates of achievement and developing hypotheses that deserve special attention.[98] Intensive, small-scale projects have an important role to play in building the international information program-what they lack in breadth, they achieve in depth. At this micro-level, differences among systems of education can be examined in considerable detail. For instance, Stevenson's work suggests that some of the factors underlying differences in achievement between U. S., Japanese, and Taiwanese children are in evidence as early as the first grade. Clearly, if this kind of finding were sustained in other case studies, an important new dimension might be added to the international achievement debate.

Small-scale studies can serve a variety of purposes in the international arena:

1. They can be used to help identify issues and to develop measures for study in mom generalizable, large-scale sample surveys.

2. They can be used to identify hypotheses appropriate for measurement with large scale methodologies, or to study variables that may of interest to a few countries.

---

[96]Ibid., 33.
[97]Ibid., 31.
[98] Harold W. Stevenson et al., *Making the Grade in Mathematics* (Reston, VA: National Council Of Teachem of Mathematics, 1990).

3. They can be used to test new data collection methods.
4. They can be used to help ensure that data from large-scale studies are intemreted appropriately by provid;ng richer contextual description. When conducted'along with large-scale studies, detail is provided that is often missing from purely statistical presentations.

Many of the ways in which the findings of cross-national differences in large-scale studies come to be understood are based on analyses derived from small-scale studies. While they may not gain the visibility of the large-scale international surveys, these studies should receive adequate support and attention. Small-scale ethnographies, longitudinal studies, and case studies represent significant opportunities to quickly learn more about which exogenous variables and schooling inputs and processes are systematically linked to performance outcomes. In some instances detailed, purposeful studies of specific phenomena in a small number of comparable countries may be the best way to identify variables appropriate to test in a broader variety of settings. .

Evidence of Progress: IEA Reading Literacy Study, IAEP-11, and the Third International Mathematics and Science Study

Two recent international achievement surveys indicate that issues of sampIing and sample comparability are receiving more attention than has been so in the past. In 1990 and 1991, the IEA conducted a study of students from 43 educational systems for the purpose of 1) describing the types and levels of reading literacy; and 2) examining the impact of varying educational policies and programs as well as home influence on reading literacy. Two populations were sampled: students in the grade in which most 9-year-olds are enrolled; and students in the grade in which most 13-year-olds are enrolled. Investigators were specifically interested in comparing reading achievement among comparable samples of students in participating educational systems.[99]

The sampling procedures adopted for the Reading Literacy Study are to be replicated in the Third International Mathematics and Science Study (TIMSS).[100] Therefore, they are worth mentioning here. The Reading Literacy sampling manual, data collection procedures, and data coding and cleaning manual were all considerably improved over prior studies. Detailed field administration procedures were easier to follow, and across all participating countries, Education Ministry involvement was significantly increased to facilitate the process of drawing and executing the samples. The impact of these changes were dramatic, at least for the United States. At the fourth-grade level, school and student response rates were 87 percent without replacement, while at the ninth-grade level, response rates were 86 percent. With more attention to coordination and administrative detail, it appears that the overall quality of the data for many countries and the United States will improve. Equally important, it appears that the countries participating in the Reading Literacy Study were better able than they were previously to estimate financial need and generate sufficient support to enable higher quality data collection.

---

[99]International Association for the Evaluation of Educational Achievement *IEA Guidebook 1989* (*The* Hague: IEA, 1989), 30-34.

[100]International Association for the Evaluation of Educational Achievement, "A Brief Introduction to the Third International Mathematics and Science Study" (September 1990); and Jeanne Griffith, Eugene Owen, Lois Peak, and Elliott Medrich, "National Education Goals and the Third International Mathematics and Science," (paper presented,at the Annual Meeting of the American Statistical Association, Atlantzt August 1991).

IAEP-11, which tested 9- and 13-year-olds in mathematics, science, and geography (13-year-olds only) expanded participation beyond that of the first IAEP. In the 1990-91 test administration, 28 educational systems participated in one or more of the assessments. Student and school background questionnaires were expanded, although there was still considerable variation in the relative emphasis educational systems assigned to the various topics covered by the tests[101] Even so, with its careful sampling strategy, field design, and increased participation, IAEP-11 has addressed some important questions that were raised by the earlier study.

Both of these studies suggest that the quality of the international assessments is improving. When the results of these surveys become available, their value will be enhanced to the extent that there will be full and complete information on: 1) sampling procedures and field execution; 2) comparability of the samples across participating educational systems; and 3) issues that could affect how the data are interpreted, so that published comparisons are appropriate to the nature and quality of the data.

Utility of International Studies for the Policy Agenda

The bridge between descriptions of performance and matters of policy maybe among the least satisfying aspects of the international assessment survey literature. While some argue that these assessments should remain broadly focused on describing differences in achievement, there have been efforts to inform a policy research agenda —some planned, others post hoc. While some of the policy analysis has been provocative, often it is inconclusive because the research was never intended or designed to answer the questions posed. Within the constraints of large-scale survey methodology, efforts should be made by researchers to design studies and analyses that tap issues of special interest to the policy community. So, for example, if there is interest in looking strategically at the substance of successful programs—that is, policies or conditions that seem associated with superior performance outcomes—appropriate methods and questions must be built into the research design. Further, it is understood that successful practice in one country may not necessarily work in another. Differences in culture and approaches to schooling and teaching are powerful intervening factors. But the international surveys can and should help to isolate aspects of the teaching and learning process that are amenable to policy intervention and, therefore, of interest across national borders.

At the same time, there maybe issues of interest to subsets of countries only. Within limits, surveys should be flexible enough to enable substudies designed to explore questions of concern to groups of countries. This poses many problems, not the least of which have to do with resources and time. If each participating country were to pursue its own agenda within the context of the international surveys, the testing mechanism might collapse.

The first priority must be to ensure the quality of the common data program. An important step in this direction is to assure adequate planning and provision of sufficient resources to assure timely, high-quality completion of all phases of the study.

[10l]International Assessment of Education progress, Center for tie Assessment of EdllCatiOnal prOgmSS, “The 1991 IAEP Assessment Objectives for Mathematics, Science and Geography” (Princeton: Educational Testing Service, 1991).

Conclusion

Large-scale cross-sectional and cross-cultural studies of student achievement yield data that are increasingly important indicators of the success of national efforts to educate an accomplished citizenry and a Productive work force. The United States should participate in these studies enthusiastically to foster international cooperation and the sharing of information on all aspects of education, to enrich our understanding of our own system of education, and to help uncover practices in other educational systems that might help improve achievement among American students.

The objectives of this report have been to make available substantial documentation and description of the various international achievement surveys of mathematics and science; to identify aspects of the survey design, data collection processes, and reporting of results that could be improved; to synthesize some of the important findings or hypotheses generated by these studies; and to suggest some strategies for upgrading data quality in future studies. .

There are clear indications that many of the concerns discussed in this report will be addressed in international studies now being designed and implemented, such as the Third International Mathematics and Science Survey scheduled for 1993–94. As these kinds of studies attract more attention, it becomes essential that they meet high technical standards. It is also important that every effort be made to help those who find these data informative and useful to understand the possibilities and limitations of the survey results. Participating countries have learned an enormous amount about the challenges of conducting complex, large-scale international surveys. They have also learned a great deal about the problems associated with interpreting the results of these studies. The activities now in progress provide substantial evidence that the quality of these surveys will improve in the future.

Appendix A

Achieved Sample Size and Response Rates

Table A.1
Sample size and response rates—schools and students: First International Mathematics Study, 13-year-olds

Educational system	Schools: Achieved sample	Schools: Response rate (percent)	Students: Achieved sample	Students: Response rate (percent)
Australia	72	--	3,078	--
Belgium	61	--	2,645	--
England	182	--	3,179	--
Federal Republic of Germany	161	--	4,475	--
Finland	111	--	1,325	--
France	124	--	3,850	--
Israel	154	--	3,336	--
Japan	210	--	2,050	--
Netherlands	90	--	1,443	--
Scotland	73	--	5,949	--
Sweden	80	--	3,712	--
United States	395	--	6,733	--

--Not available.

NOTE: Data for grade level containing the most 13-year-olds (population 1b).

SOURCE: Data from T. Husen, *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*, Vol. I (New York: John Wiley & Sons, Inc., 1967), 158–61.

Table A.2
Sample size and response rates—schools and students: First International Mathematics Study, Last year of secondary school (mathematics students)

Educational system	Schools		Students	
	Achieved sample	Response rate (percent)	Achieved sample	Response rate (percent)
Australia	56	--	1,089	--
Belgium	30	--	519	--
England	77	--	1,031	--
Federal Republic of Germany	37	--	649	--
Finland	27	--	460	--
France	14	--	337	--
Israel	8	--	146	--
Japan	91	--	818	--
Netherlands	30	--	491	--
Scotland	63	--	1,422	--
Sweden	23	--	1,024	--
United States	149	--	1,660	--

--Not available.

NOTE: Data for mathematics students in last-year secondary (population 3A).

SOURCE: Data from T. Husen, *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*, Vol. I (New York: John Wiley & Sons, Inc., 1967), 158–61.

Table A.3
Sample size and response rates—schools and students: First International Mathematics Study, Last year of secondary school (non-mathematics students)

Educational system	Schools		Students	
	Achieved sample	Response rate (percent)	Achieved sample	Response rate (percent)
Belgium	43	--	1,004	--
England	84	--	1,906	--
Federal Republic of Germany	36	--	643	--
Finland	24	--	482	--
France	--	--	--	--
Japan	349	--	4,372	--
Netherlands	--	--	--	--
Scotland	64	--	2,123	--
Sweden	20	--	320	--
United States	155	--	2,152	--

--Not available.

NOTE: Data for non-mathematics students in last-year secondary (population 3b).

SOURCE: Data from T. Husen, *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*, Vol. I (New York: John Wiley & Sons, Inc., 1967), 158–61.

Table A.4
Sample size and response rates—schools and students: Second International Mathematics Study, 13-year-olds

Educational system	Schools		Students	
	Achieved sample	Response rate (percent)	Achieved sample	Response rate (percent)
Belgium (Flemish)	158	--	3,103	--
Belgium (French)	108	86	3,103	--
Canada (British Columbia)	89	--	2,228	--
Canada (Ontario)	112	86	5,013	--
England and Wales	94	82	2,678	84
Finland	98	95	4,484	--
France	187	99	8,889	--
Hong Kong	125	--	5,548	--
Hungary	70	100	1,754	95
Israel	81	82	3,819	78
Japan	213	97	8,091	--
Luxembourg	42	91	2,106	96
Netherlands	236	100	5,500	--
New Zealand	100	100	5,218	--
Nigeria	48	72	1,456	72
Scotland	76	--	1,356	67
Swaziland	25	100	904	--
Sweden	96	96	3,585	88
Thailand	99	99	4,023	95
United States (Districts)	93	50	(*)	(*)
United States	150	83	6,858	76

--Not available.
*Not applicable.

SOURCE: Data from U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), section 4.

Table A.5
Sample size and response rates—schools and students: Second International Mathematics Study, Last year of secondary school

Educational system	Schools: Achieved sample	Schools: Response rate (percent)	Students: Achieved sample	Students: Response rate (percent)
Belgium (Flemish)	131	87	2,859	--
Belgium (French)	87	77	2,062	--
Canada (British Columbia)	78	--	1,954	--
Canada (Ontario)	79	93	3,214	--
England and Wales	312	90	3,578	--
Finland	81	92	1,550	88
Hong Kong	112	--	3,294	--
Hungary	92	100	2,455	97
Israel	64	70	1,905	72
Japan	192	93	7,954	100
New Zealand	79	99	1,193	98
Scotland	54	81	1,501	--
Sweden	127	98	2,712	93
Thailand	64	100	3,747	90
United States (Districts)	93	48	(*)	(*)
United States	150	69	4,671	77

--Not available.
*Not applicable.

SOURCE: Data from U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), section 5.

Table A.6
Sample size and response rates—schools and students: First International Science Study, 10-year-olds

Educational system	Schools: Achieved sample	Schools: Response rate (percent)	Students: Achieved sample	Students: Response rate (percent)
Belgium (Flemish)	31	59	717	42
Belgium (French)	33	77	767	70
Chile*	81	82	1,470	80
England	162	79	3,573	73
Federal Republic of Germany	68	46	1,741	46
Finland	97	97	1,305	97
Hungary	152	99	4,858	95
India*	176	50	2,704	26
Iran*	53	--	1,640	--
Israel*	110	97	1,887	92
Italy	298	73	4,503	49
Japan	250	100	2,467	100
Netherlands	60	66	1,629	65
Scotland	105	98	2,169	92
Sweden	98	99	2,009	96
Thailand*	31	94	1,810	82
United States	239	68	5,479	64

--Not available.

*Mean achievement scores not calculated, or not published; or system did not participate in the achievement test survey.

SOURCE: Data from Gilbert Peaker, *An Empirical Study of Education in Twenty-One Countries: A Technical Report* (New York: John Wiley & Sons, Inc., 1975), 36, 37.

Table A.7
Sample size and response rate—students and schools: First International Science Study, 14-year-olds

Educational system	Schools: Achieved sample	Schools: Response rate (percent)	Students: Achieved sample	Students: Response rate (percent)
Australia	221	99	5,301	96
Belgium (Flemish)	31	53	699	39
Belgium (French)	21	34	564	22
Chile*	103	75	1,311	72
England	146	66	3,256	60
Federal Republic of Germany	83	59	2,233	56
Finland	77	100	2,325	98
Hungary	210	100	7,026	94
India*	155	44	2,931	40
Iran*	33	--	1,336	--
Israel*	125	91	1,958	80
Italy	343	86	7,383	83
Japan	196	98	1,945	98
Netherlands	50	52	1,236	49
New Zealand	74	100	1,974	91
Scotland	70	95	1,982	85
Sweden	95	96	2,475	91
Thailand*	29	91	1,932	81
United States	142	57	6,870	46

--Not available.

*Mean achievement scores not calculated, or not published; or system did not participate in the achievement test survey.

SOURCE: Data from Gilbert Peaker, *An Empirical Study of Education in Twenty-One Countries: A Technical Report* (New York: John Wiley & Sons, Inc., 1975), 36, 37.

Table A.8
Sample size and response rates—schools and students: First International Science Study, Last year of secondary school

Educational system	Schools		Students	
	Achieved sample	Response rate (percent)	Achieved sample	Response rate (percent)
Australia	194	99	4,197	92
Belgium (Flemish)	18	34	472	27
Belgium (French)	42	63	1,231	59
Chile*	73	76	2,052	74
England	70	32	2,274	27
Federal Republic of Germany	80	80	1,988	71
Finland	77	100	1,807	82
France	141	90	3,582	87
Hungary	39	100	2,855	98
India*	124	31	3,153	28
Iran*	34	--	1,435	--
Israel*	71	84	863	81
Italy	253	70	16,437	61
Netherlands	38	39	1,164	37
New Zealand	69	100	1,714	83
Scotland	69	88	1,328	80
Sweden	142	95	2,988	90
Thailand*	13	93	724	66
United States	114	43	5,200	35

--Not available.

*Mean achievement scores not calculated, or not published; or system did not participate in the achievement test survey.

SOURCE: Data from Gilbert Peaker, *An Empirical Study of Education in Twenty-One Countries: A Technical Report* (New York: John Wiley & Sons, Inc., 1975), 36, 37.

Table A.9
Sample size and response rates—schools and students: Second International Science Study, 10-year-olds

Educational system	Schools		Students	
	Achieved sample	Response rate (percent)	Achieved sample	Response rate (percent)
Australia	220	78	4,259	67
Canada (English)	215	69	5,104	67
England	181	66	3,748	62
Finland	106	96	1,600	86
Hong Kong*	146	99	5,342	96
Hungary	100	100	2,590	95
Italy	119	58	5,156	84
Japan	221	99	7,924	99
Korea	146	99	3,489	99
Norway	91	62	1,305	54
Philippines	463	93	16,851	92
Poland	199	100	4,390	93
Singapore	232	92	5,547	92
Sweden	64	70	1,449	74
United States	123	88	2,822	77

*Sampled classes, not schools.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 96, 81.

Table A.10
Sample size and response rates—schools and students: Second International Science Study, 14-year-olds

Educational system	Schools		Students	
	Achieved sample	Response rate (percent)	Achieved sample	Response rate (percent)
Australia	233	84	4,917	74
Canada (English)	209	66	5,543	66
England	147	60	3,118	53
Finland	90	97	2,546	90
Hong Kong*	132	99	4,973	95
Hungary	99	99	2,515	93
Italy	291	72	3,228	89
Japan	199	99	7,610	95
Korea	189	100	4,522	100
Netherlands	224	92	5,025	86
Norway	77	65	1,420	59
Philippines	269	90	10,874	88
Poland	201	100	4,520	95
Singapore	185	100	4,520	95
Sweden	69	60	1,461	50
Thailand	96	93	3,780	92
United States	119	85	2,519	69

*Sampled classes, not schools.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 96, 82.

Table A.11
Sample size and response rates—schools and students: Second International Science Study, Last year of secondary school (biology students)

Educational system	Schools: Achieved sample	Schools: Response rate (percent)	Students: Achieved sample	Students: Response rate (percent)
Australia	164	83	1,631	72
Canada (English)	187	65	3,254	53
England	123	--	884	--
Finland	43	94	1,652	84
Hong Kong (form 6)	158	--	5,960	--
Hong Kong (form 7)	114	--	3,614	--
Hungary	71	--	301	--
Italy	12	--	147	--
Japan*	38	95	1,212	90
Norway	52	--	276	--
Poland	71	100	764	45
Singapore	8	100	902	84
Sweden	119	--	1,232	--
United States	43	92	659	77

--Not available.
*Sampled classes, not schools.

NOTE: Australia tested 29 items; United States tested 25 items; others tested 30 items.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 96, 84.

Table A.12
Sample size and response rates—schools and students: Second International Science Study, Last year of secondary school (chemistry students)

Educational system	Schools: Achieved sample	Schools: Response rate (percent)	Students: Achieved sample	Students: Response rate (percent)
Australia	164	82	1,177	77
Canada (English)	179	60	2,923	51
England	123	--	892	--
Finland	44	96	971	83
Hong Kong (form 6)	158	--	6,018	--
Hong Kong (form 7)	114	--	3,670	--
Hungary	56	--	143	--
Italy	24	--	217	--
Japan*	43	100	1,468	93
Norway	46	--	283	--
Poland	71	100	765	45
Singapore	8	100	945	74
Sweden	119	--	1,172	--
United States	40	76	537	70

--Not available.
*Sampled classes, not schools.

NOTE: United States tested 25 items; others tested 30 items.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 85, 96.

Table A.13
Sample size and response rates—schools and students:
Second International Science Study,
Last year of secondary school (physics students)

Educational system	Schools: Achieved sample	Schools: Response rate (percent)	Students: Achieved sample	Students: Response rate (percent)
Australia	163	82	1,073	76
Canada (English)	181	64	2,766	54
England	125	--	917	--
Finland	42	91	810	83
Hong Kong (Form 6)	158	--	6,025	--
Hong Kong (Form 7)	114	--	3,679	--
Hungary	75	--	398	--
Italy	120	--	1,766	--
Japan*	36	92	1,187	89
Norway	55	--	443	--
Poland	79	100	1,716	91
Singapore	8	100	1,071	82
Sweden	119	--	1,156	--
United States	35	76	485	64

--Not available.
*Sampled classes, not schools.

NOTE: United States tested 26 items; Canada tested 29 items; others tested 30 items.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 96, 86.

Table A.14
Sample size and response rates—schools and students: International Assessment of Educational Progress, 13-year-olds (mathematics proficiency)

Educational system	Schools		Students	
	Achieved sample	Response rate (percent)	Achieved sample	Response rate (percent)
Canada (British Columbia)	--	100	3,025	98
Canada (New Brunswick: English)	--	95	2,047	92
Canada (New Brunswick: French)	--	91	1,548	74
Canada (Ontario: English)	--	96	2,008	94
Canada (Ontario: French)	--	97	2,075	96
Canada (Quebec: English)	--	97	2,090	98
Canada (Quebec: French)	--	95	2,186	97
Ireland	99	97	2,253	90
Korea	--	94	2,243	98
Spain	100	89	1,756	98
United Kingdom	85	70	2,202	94
United States	--	87	905	90

--Not available.

NOTE: British Columbia tested both public and private schools, but response rate for British Columbia reflects public schools only.

SOURCE: Data from Archie E. Lapointe, Nancy A. Mead, and Gary W. Phillips, *A World of Differences: An International Assessment of Mathematics and Science* (Princeton: Educational Testing Service, January 1989), 84, 85.

Table A.15
Sample size and response rates—schools and students: International Assessment of Educational Progress, 13-year-olds (science proficiency)

Educational system	Schools		Students	
	Achieved sample	Response rate (percent)	Achieved sample	Response rate (percent)
Canada (British Columbia)	--	100	3,025	96
Canada (New Brunswick: English)*	--	95	2,047	93
Canada (New Brunswick: French)	--	91	1,548	73
Canada (Ontario: English)	--	96	2,008	94
Canada (Ontario: French)	--	97	2,075	96
Canada (Quebec: English)	--	97	2,090	96
Canada (Quebec: French)	--	95	2,186	97
Ireland	--	97	2,253	90
Korea	--	94	2,243	98
Spain	--	89	1,756	98
United Kingdom	--	70	2,202	94
United States	--	87	859	90

--Not available.
*Sampled classes, not schools.

NOTE: British Columbia tested both public and private schools, but response rate for British Columbia reflects public schools only.

SOURCE: Data from Archie E. Lapointe, Nancy A. Mead, and Gary W. Phillips, *A World of Differences: An International Assessment of Mathematics and Science* (Princeton: Educational Testing Service, January 1989), 84, 85.

Appendix B

Mean Scores and
Means Compared with United States

Table B.1
Mean scores and means compared with United States:
First International Mathematics Study, 13-year-olds (70 items)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Israel	32.3	0.47	14.7	higher
Japan	32.2	0.52	16.9	higher
Belgium	30.4	0.53	13.7	higher
Finland	26.4	0.60	9.6	higher
Federal Republic of Germany	25.4	0.58	11.7	higher
England	23.8	0.57	18.5	higher
Scotland	22.3	0.64	15.7	higher
Netherlands	21.4	0.61	12.1	higher
France	21.0	0.70	13.2	higher
Australia	18.9	0.38	12.3	same
United States	17.8	0.28	13.3	U.S.
Sweden	15.3	0.51	10.8	lower

[1]Standard error calculated using design effect specified by country in Husen, 158–61.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from Torsten Husen, *International Study of Achievement in Mathematics*, Vol. II (New York: John Wiley & Sons, Inc., 1967), 23; Vol. I, 158–61.

Table B.2
Mean scores and means compared with United States:
First International Mathematics Study, Last year of secondary school (mathematics students—69 items)

Educational system and rank by mean	Number of items correct: Mean	Standard error of mean[1]	Standard deviation	Mean compared with U.S.[2]
Israel	36.4	1.35	8.6	higher
England	34.6	0.72	12.6	higher
Belgium	34.6	0.89	12.6	higher
France	33.4	0.80	10.8	higher
Netherlands	31.9	0.60	8.1	higher
Japan	31.4	1.00	8.6	higher
Federal Republic of Germany	28.8	0.50	9.8	higher
Sweden	27.3	0.68	11.9	higher
Scotland	25.5	0.44	10.4	higher
Finland	25.3	0.65	9.6	higher
Australia	21.6	0.64	10.5	higher
United States	13.8	0.51	12.6	U.S.

[1]Standard error calculated using design effect specified by country in Husen, 158–61.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from Torsten Husen, *International Study of Achievement in Mathematics*, Vol. II (New York: John Wiley & Sons, Inc., 1967), 23; Vol. I, 158–61.

Table B.3
Mean scores and means compared with United States:
First International Mathematics Study, Last year of secondary school (non-mathematics students—58 items)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Federal Republic of Germany	27.7	0.30	7.6	higher
France	26.2	--	9.5	--
Japan	25.3	0.43	14.3	higher
Netherlands	24.7	--	9.8	--
Belgium	24.2	0.57	9.5	higher
Finland	22.5	0.54	8.3	higher
England	21.4	0.31	10.0	higher
Scotland	20.7	0.37	9.5	higher
Sweden	12.6	0.38	6.2	higher
United States	8.3	0.36	9.0	U.S.

--Not available.
[1]Standard error calculated using design effect by country in Husen, 158–61.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from Torsten Husen, *International Study of Achievement in Mathematics*, Vol. II (New York: Wiley, 1967), 25; Vol. I, 158–61.

Table B.4
Mean scores and means compared with United States: Second International Mathematics Study, 13-year-olds (46 core items—arithmetic)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Japan	60.3	1.5	--	higher
Netherlands	59.3	1.1	--	higher
Canada (British Columbia)	58.0	1.3	--	higher
Belgium (Flemish)	58.0	1.4	--	higher
France	57.7	1.3	--	higher
Belgium (French)	57.0	1.8	--	higher
Hungary	56.8	1.5	--	same
Hong Kong	55.1	0.5	--	same
Canada (Ontario)	54.5	1.1	--	same
United States	51.4	1.2	--	same
Scotland	50.2	0.5	--	U.S.
Israel	49.9	1.5	--	same
England and Wales	48.2	0.9	--	same
New Zealand	45.6	1.2	--	same
Finland	45.5	1.3	--	lower
Luxembourg	45.4	0.4	--	lower
Thailand	43.1	1.3	--	lower
Sweden	40.6	0.9	--	lower
Nigeria	40.8	1.3	--	lower
Swaziland	32.3	1.4	--	lower

--Not available.
[1]Standard error calculated in Robitaille and Garden. Method not specified.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics,* Vol. II (Oxford: Pergamon Press, 1989), 105; U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), 115; U.S. Department of Education, National Center for Education Statistics, *Digest of Education Statistics* (Washington, DC, 1989), 389.

Table B.5
Mean scores and means compared with United States: Second International Mathematics Study, 13-year-olds (30 core items—algebra)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Japan	60.3	1.6	--	higher
France	55.0	0.9	--	higher
Belgium (Flemish)	52.9	1.7	--	higher
Netherlands	51.3	1.2	--	higher
Hungary	50.4	1.2	--	higher
Belgium (French)	49.1	2.0	--	higher
Canada (British Columbia)	47.9	1.4	--	higher
Israel	44.0	1.6	--	same
Finland	43.6	1.3	--	same
Hong Kong	43.2	0.8	--	same
Scotland	42.9	0.7	--	same
United States	42.1	1.2	--	U.S.
Canada (Ontario)	42.0	0.7	--	same
England and Wales	40.1	1.1	--	same
New Zealand	39.4	1.1	--	same
Thailand	37.7	1.0	--	same
Nigeria	32.4	1.7	--	lower
Sweden	32.3	0.8	--	lower
Luxembourg	31.2	0.5	--	lower
Swaziland	25.1	1.5	--	lower

--Not available.

[1]Standard error calculated in Robitaille and Garden. Method not specified.

[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics,* Vol. II (Oxford: Pergamon Press, 1989), 105; U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), 115; U.S. Department of Education, National Center for Education Statistics, *Digest of Education Statistics* (Washington, DC, 1989), 389.

Table B.6
Mean scores and means compared with United States: Second International Mathematics Study, 13-year-olds (39 core items—geometry)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Japan	57.6	1.3	--	higher
Hungary	53.4	1.0	--	higher
Netherlands	52.0	1.0	--	higher
Scotland	45.5	0.6	--	higher
England and Wales	44.8	0.8	--	higher
New Zealand	44.8	1.0	--	higher
Canada (Ontario)	43.2	0.7	--	higher
Finland	43.2	1.2	--	higher
Belgium (French)	42.8	1.5	--	same
Belgium (Flemish)	42.5	1.1	--	higher
Hong Kong	42.5	0.5	--	higher
Canada (British Columbia)	42.3	1.2	--	same
Sweden	39.4	0.8	--	same
Thailand	39.3	0.9	--	same
France	38.0	0.8	--	same
United States	37.8	0.9	--	U.S.
Israel	35.9	1.4	--	same
Swaziland	31.1	1.3	--	lower
Nigeria	26.2	0.8	--	lower
Luxembourg	25.3	0.4	--	lower

--Not available.
[1]Standard error calculated in Robitaille and Garden. Method not specified.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics*, Vol. II (Oxford: Pergamon Press, 1989), 105; U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), 115; U.S. Department of Education, National Center for Education Statistics, *Digest of Education Statistics* (Washington, DC, 1989), 389.

Table B.7
Mean scores and means compared with United States: Second International Mathematics Study, 13-year-olds (24 core items—measurement)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Japan	68.6	1.3	--	higher
Hungary	62.1	1.4	--	higher
Netherlands	61.9	1.0	--	higher
France	59.5	0.9	--	higher
Belgium (Flemish)	58.2	1.3	--	higher
Belgium (French)	56.8	1.5	--	higher
Hong Kong	52.6	0.4	--	higher
Canada (British Columbia)	51.9	1.3	--	higher
Finland	51.3	1.2	--	higher
Canada (Ontario)	50.8	0.9	--	higher
Luxembourg	50.1	0.4	--	higher
Sweden	48.7	1.0	--	higher
England and Wales	48.6	0.9	--	higher
Scotland	48.4	0.7	--	higher
Thailand	48.3	1.1	--	higher
Israel	46.4	1.2	--	higher
New Zealand	45.1	1.1	--	higher
United States	40.8	0.9	--	U.S.
Swaziland	35.2	1.3	--	lower
Nigeria	30.7	1.1	--	lower

--Not available.

[1]Standard error calculated in Robitaille and Garden. Method not specified.

[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics,* Vol. II (Oxford: Pergamon Press, 1989), 105; U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), 115; U.S. Department of Education, National Center for Education Statistics, *Digest of Education Statistics* (Washington, DC, 1989), 389.

Table B.8
Mean scores and means compared with United States: Second International Mathematics Study, 13-year-olds (18 core items—descriptive statistics)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Japan	70.9	1.5	--	higher
Netherlands	65.9	0.9	--	higher
Canada (British Columbia)	61.3	1.3	--	same
Hungary	60.4	1.4	--	same
England and Wales	60.2	0.9	--	same
Scotland	59.3	0.5	--	same
Belgium (Flemish)	58.2	1.5	--	same
United States	57.7	1.1	--	U.S.
Finland	57.6	1.1	--	same
France	57.4	1.0	--	same
New Zealand	57.3	1.1	--	same
Canada (Ontario)	57.0	1.0	--	same
Sweden	56.3	1.1	--	same
Hong Kong	55.9	0.6	--	same
Belgium (French)	52.0	1.7	--	same
Israel	51.9	1.3	--	lower
Thailand	45.3	1.0	--	lower
Luxembourg	37.3	0.4	--	lower
Nigeria	37.0	1.3	--	lower
Swaziland	36.0	1.7	--	lower

--Not available.
[1]Standard error calculated in Robitaille and Garden. Method not specified.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics,* Vol. II (Oxford: Pergamon Press, 1989), 105; U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), 115; U.S. Department of Education, National Center for Education Statistics, *Digest of Education Statistics* (Washington, DC, 1989), 389.

Table B.9
Mean scores and means compared with United States:
Second International Mathematics Study, Last year of secondary school
(17 items—number systems)

Educational system and rank by mean	Percent of items correct: Mean[1]	Standard error of mean[2]	Standard deviation	Mean compared with U.S.[3]
Hong Kong	78	1.5	--	higher
Japan	68	1.1	--	higher
Sweden	62	0.8	--	higher
England and Wales	59	0.8	--	higher
Finland	57	1.1	--	higher
New Zealand	51	1.4	--	higher
Belgium (Flemish)	48	1.1	--	higher
Canada (Ontario)	47	0.9	--	higher
Israel	46	1.5	--	higher
Belgium (French)	44	1.5	--	same
Canada (British Columbia)	43	1.3	--	same
United States	40	1.1	--	U.S.
Scotland	39	1.1	--	same
Thailand	33	1.2	--	lower
Hungary	28	1.3	--	lower

--Not available.
[1]Data available rounded to nearest whole number.
[2]Standard error calculated in Robitaille and Garden. Method not specified.
[3]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics*, Vol. II (Oxford: Pergamon Press, 1989), 130; U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), 120; Curtis C. McKnight et al., *The Underachieving Curriculum* (Champaign: Stipes, 1989), 125.

Table B.10
Mean scores and means compared with United States: Second International Mathematics Study, Last year of secondary school (26 items—algebra)

Educational system and rank by mean	Percent of items correct			Mean compared with U.S.[3]
	Mean[1]	Standard error of mean[2]	Standard deviation	
Hong Kong	78	1.4	--	higher
Japan	78	1.0	--	higher
Finland	69	0.8	--	higher
England and Wales	66	0.6	--	higher
Belgium (Flemish)	60	1.2	--	higher
Sweden	61	0.8	--	higher
Israel	60	1.5	--	higher
New Zealand	57	1.2	--	higher
Canada (Ontario)	57	1.0	--	higher
Belgium (French)	55	1.6	--	higher
Scotland	48	0.9	--	higher
Canada (British Columbia)	47	1.4	--	higher
Hungary	45	1.5	--	same
United States	43	1.2	--	same
Thailand	38	1.4	--	U.S.
				same

--Not available.

[1]Data available rounded to nearest whole number.

[2]Standard error calculated in Robitaille and Garden. Method not specified.

[3]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics*, Vol. II (Oxford: Pergamon Press, 1989), 130; U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), 120; Curtis C. McKnight et al., *The Underachieving Curriculum* (Champaign: Stipes, 1989), 125.

Table B.11
Mean scores and means compared with United States:
Second International Mathematics Study, Last year of secondary school
(26 items—geometry)

Educational system and rank by mean	Percent of items correct			Mean compared with U.S.[3]
	Mean[1]	Standard error of mean[2]	Standard deviation	
Hong Kong	65	1.4	--	higher
Japan	60	1.1	--	higher
England and Wales	51	0.5	--	higher
Sweden	49	0.5	--	higher
Finland	48	0.8	--	higher
New Zealand	43	1.0	--	higher
Scotland	42	0.8	--	higher
Canada (Ontario)	42	0.7	--	higher
Belgium (Flemish)	42	1.1	--	higher
Belgium (French)	38	1.3	--	higher
Israel	35	1.5	--	same
United States	31	1.0	--	U.S.
Hungary	30	1.1	--	same
Canada (British Columbia)	30	1.2	--	same
Thailand	28	0.9	--	same

--Not available.
[1]Data available rounded to nearest whole number.
[2]Standard error calculated in Robitaille and Garden. Method not specified.
[3]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics*, Vol. II (Oxford: Pergamon Press, 1989), 130; U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), 120; Curtis C. McKnight et al., *The Underachieving Curriculum* (Champaign: Stipes, 1989), 125.

Table B.12
Mean scores and means compared with United States: Second International Mathematics Study, Last year of secondary school (46 items—elementary functions and calculus)

Educational system and rank by mean	Percent of items correct			Mean compared with U.S.[3]
	Mean[1]	Standard error of mean[2]	Standard deviation	
Hong Kong	71	1.9	--	higher
Japan	66	1.5	--	higher
England and Wales	58	0.6	--	higher
Finland	55	1.1	--	higher
Sweden	51	0.8	--	higher
New Zealand	48	1.1	--	higher
Canada (Ontario)	46	1.0	--	higher
Belgium (Flemish)	46	1.1	--	higher
Israel	45	1.6	--	higher
Belgium (French)	43	1.4	--	higher
Scotland	32	0.9	--	higher
United States	29	1.2	--	same
Thailand	26	0.8	--	U.S.
Hungary	26	1.1	--	same
Canada (British Columbia)	21	1.0	--	same
				lower

--Not available.
[1]Data available rounded to nearest whole number.
[2]Standard error calculated in Robitaille and Garden. Method not specified.
[3]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics*, Vol. II (Oxford: Pergamon Press, 1989), 130; U.S. Department of Education, Center for Education Statistics, Robert A. Garden, *Second IEA Mathematics Study: Sampling Report* (Washington, DC, March 1987), 120; Curtis C. McKnight et al., *The Underachieving Curriculum* (Champaign: Stipes, 1989), 125.

Table B.13
Mean scores and means compared with United States:
First International Science Study, 10-year-olds (40 core items)

Educational system and rank by mean	Number of items correct: Mean	Standard error of mean[1]	Standard deviation	Mean compared with U.S.[2]
Japan	21.7	0.31	7.7	higher
Sweden	18.3	0.39	7.3	same
Belgium (Flemish)	17.9	0.65	7.3	same
United States	17.7	0.30	9.3	U.S.
Finland	17.5	0.54	8.2	same
Hungary	16.7	0.28	8.0	same
Italy	16.5	0.31	8.6	same
England	15.7	0.34	8.5	lower
Netherlands	15.3	0.45	7.6	lower
Federal Republic of Germany	14.9	0.43	7.5	lower
Scotland	14.0	0.43	8.4	lower
Belgium (French)	13.9	0.62	7.1	lower

[1]Standard error derived from Peaker ($\sqrt{DEFF} = 2.4$).
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

NOTE: Mean scores not available for five systems—Chile, India, Iran, Israel, Thailand. Score may not have been calculated, or score may not have been published, or system may not have participated in the achievement test survey.

SOURCE: Data from L.C. Comber and John P. Keeves, *Science Education in Nineteen Countries* (New York: John Wiley & Sons, Inc., 1973), 159; Gilbert F. Peaker, *An Empirical Study of Education in Twenty-One Countries: A Technical Report* (New York: John Wiley & Sons, Inc., 1975), 36, 37.

Table B.14
Mean scores and means compared with United States: First International Science Study, 14-year-olds (80 core items)

Educational system and rank by mean	Number of items correct: Mean	Standard error of mean[1]	Standard deviation	Mean compared with U.S.[2]
Japan	31.2	0.81	14.8	higher
Hungary	29.1	0.36	12.7	higher
Australia	24.6	0.44	13.4	higher
New Zealand	24.2	0.70	12.9	higher
Federal Republic of Germany	23.7	0.58	11.5	higher
Sweden	21.7	0.56	11.7	same
United States	21.6	0.34	11.6	U.S.
Scotland	21.4	0.77	14.2	same
England	21.3	0.59	14.1	same
Belgium (Flemish)	21.2	0.84	9.2	same
Finland	20.5	0.53	10.6	same
Italy	18.5	0.28	10.2	lower
Netherlands	17.8	0.68	10.0	lower
Belgium (French)	15.4	0.89	8.8	lower

[1]Standard error derived from Peaker ($\sqrt{DEFF} = 2.4$).
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

NOTE: Mean scores not available for five systems—Chile, India, Iran, Israel, Thailand. Score may not have been calculated, or score may not have been published, or system may not have participated in the achievement test survey.

SOURCE: Data from L.C. Comber and John P. Keeves, *Science Education in Nineteen Countries* (New York: John Wiley & Sons, Inc., 1973), 159; Gilbert F. Peaker, *An Empirical Study of Education in Twenty-One Countries: A Technical Report* (New York: John Wiley & Sons, Inc., 1975), 37.

Table B.15
Mean scores and means compared with United States:
First International Science Study, Last year of secondary school
(60 core items)

Educational system and rank by mean	Number of items correct: Mean	Standard error of mean[1]	Standard deviation	Mean compared with U.S.[2]
New Zealand	29.0	0.63	11.6	higher
Federal Republic of Germany	26.9	0.48	8.9	higher
Australia	24.7	0.40	10.7	higher
Netherlands	23.3	0.78	11.1	higher
Scotland	23.1	0.80	12.1	higher
England	23.1	0.58	11.5	higher
Hungary	23.0	0.40	9.0	higher
Finland	19.8	0.55	9.8	higher
Sweden	19.2	0.45	10.2	higher
France	18.3	0.35	8.7	higher
Belgium (Flemish)	17.4	0.90	8.1	same
Italy	15.9	0.16	8.8	same
Belgium (French)	15.3	0.54	7.9	same
United States	13.7	1.00	9.5	U.S.

[1]Standard error derived from Peaker ($\sqrt{DEFF} = 2.4$).
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

NOTE: Mean scores not available for five systems—Chile, India, Iran, Israel, Thailand. Score may not have been calculated, or score may not have been published, or system may not have participated in the achievement test survey.

SOURCE: Data from L.C. Comber and John P. Keeves, *Science Education in Nineteen Countries* (New York: John Wiley & Sons, Inc., 1973), 159; Gilbert F. Peaker, *An Empirical Study of Education in Twenty-One Countries: A Technical Report* (New York: John Wiley & Sons, Inc., 1975), 37.

Table B.16
Mean scores and means compared with United States: Second International Science Study, 10-year-olds (24 core items)

Educational system and rank by mean	Number of items correct: Mean	Number of items correct: Standard error of mean[1]	Number of items correct: Standard deviation	Mean compared with U.S.[2]
Japan	15.4	0.07	4.0	higher
Korea	15.4	0.16	4.2	higher
Finland	15.3	0.15	4.0	higher
Sweden	14.7	0.16	4.0	higher
Hungary	14.4	0.23	4.5	higher
Canada (English)	13.7	0.13	4.3	higher
Italy	13.4	0.26	4.7	same
United States	13.2	0.18	4.6	same
Australia	12.9	0.18	4.5	U.S.
Norway	12.7	0.30	4.1	same
Poland	11.9	0.16	4.5	same
England	11.7	0.17	4.5	lower
Singapore	11.2	0.18	4.1	lower
Hong Kong	11.2	0.20	4.2	lower
Philippines	9.5	0.16	4.5	lower
				lower

[1]Standard error jacknifed, see IEA, 23.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 81, 96.

Table B.17
Mean scores and means compared with United States: Second International Science Study, 14-year-olds (30 core items)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Hungary	21.7	0.26	4.7	higher
Japan	20.2	0.09	5.0	higher
Netherlands	19.8	0.26	5.1	higher
Canada (English)	18.6	0.17	4.7	higher
Finland	18.5	0.13	4.2	higher
Sweden	18.4	0.22	4.9	higher
Korea	18.1	0.15	4.6	higher
Poland	18.1	0.22	5.2	higher
Norway	17.9	0.16	4.7	higher
Australia	17.8	0.19	4.9	higher
England	16.7	0.22	4.9	same
Italy	16.7	0.28	5.0	same
Singapore	16.5	0.28	4.9	same
United States	16.5	0.27	5.0	U.S.
Thailand	16.5	0.22	4.1	same
Hong Kong	16.4	0.25	4.5	same
Philippines	11.5	0.20	4.6	lower

[1]Standard error jacknifed, see IEA, 23.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 32.

Table B.18
Mean scores and means compared with United States: Second International Science Study, Last year of secondary school (30 core items—biology)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Singapore	66.8	--	12.8	--
England	63.4	0.24	13.1	higher
Hungary	59.7	0.37	13.5	higher
Poland	56.9	0.29	12.9	higher
Hong Kong (form 7)	55.8	0.28	16.8	higher
Norway	54.8	0.33	15.0	higher
Finland	51.9	0.16	12.8	higher
Hong Kong (form 6)	50.8	0.27	14.8	higher
Sweden	48.5	0.23	15.8	higher
Australia	48.2	0.15	13.9	higher
Canada (English)	45.9	0.20	14.0	higher
Japan	46.2	0.48	15.1	higher
Italy	42.3	1.05	14.1	higher
United States[3]	37.9	0.41	15.4	U.S.

--Not available.
[1]Standard error jacknifed, see IEA, 23.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.
[3]United States test core 25 items, and Australia tested 29 items, as compared with 30 items in all other countries.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 51.

Table B.19
Mean scores and means compared with United States: Second International Science Study, Last year of secondary school (30 core items—chemistry)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Hong Kong (form 7)	77.0	0.43	17.4	higher
England	69.5	0.29	17.2	higher
Singapore	66.1	--	17.4	--
Hong Kong (form 6)	64.4	0.40	17.0	higher
Japan	51.9	0.86	22.0	higher
Hungary	47.7	0.70	18.3	higher
Australia	46.6	0.31	18.8	higher
Poland	44.6	0.44	17.1	higher
Norway	41.9	0.37	16.8	higher
Sweden	40.0	0.23	16.6	higher
Italy	38.0	1.45	23.4	same
United States[3]	37.7	0.67	18.3	U.S.
Canada (English)	36.9	0.31	16.0	same
Finland	33.3	0.26	13.7	lower

--Not available.
[1]Standard error jacknifed, see IEA, 23.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.
[3]United States test core 25 items, as compared with 30 items in other countries.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 52.

Table B.20
Mean scores and means compared with United States:
Second International Science Study, Last year of secondary school
(30 core items—physics)

Educational system and rank by mean	Number of items correct			Mean compared with U.S.[2]
	Mean	Standard error of mean[1]	Standard deviation	
Hong Kong (form 7)	69.9	0.38	14.4	higher
Hong Kong (form 6)	59.3	0.34	14.7	higher
England	58.3	0.20	14.9	higher
Hungary	56.5	0.50	17.2	higher
Japan	56.1	0.58	17.2	higher
Singapore	54.9	--	13.2	--
Norway	52.8	0.33	15.6	higher
Poland	51.5	--	17.2	--
Australia	48.5	0.21	15.1	higher
United States[3]	45.5	0.53	15.8	U.S.
Sweden	44.8	0.18	14.9	same
Canada (English)	39.6	0.20	14.6	lower
Finland	37.9	0.27	13.8	lower
Italy	28.0	0.25	12.9	lower

--Not available.
[1]Standard error jacknifed, see IEA, 23.
[2]Based on Bonferroni adjusted t-test for comparisons with the United States.
[3]United States test core 26 items, as against 30 items in other countries.

SOURCE: Data from International Association for the Evaluation of Educational Achievement, *Student Achievement in Seventeen Countries* (Oxford: Pergamon Press, 1988), 53.

Table B.21
Mean scores and means compared with United States: International Assessment of Educational Progress, 13-year-olds (63 items—mathematics proficiency)

Educational system and rank by mean	Proficiency score		Mean compared with U.S.[3]
	Mean[1]	Standard error[2]	
Korea	567.8	2.7	higher
Canada (Quebec: French)	543.0	3.1	higher
Canada (British Columbia)	539.8	2.2	higher
Canada (Quebec: English)	535.8	2.0	higher
Canada (New Brunswick: English)	529.0	2.6	higher
Canada (Ontario: English)	516.1	3.1	higher
Canada (New Brunswick: French)	514.2	3.3	higher
Spain	511.7	4.6	higher
United Kingdom	509.9	3.5	higher
Ireland	504.3	3.7	higher
Canada (Ontario: French)	481.5	2.7	same
United States	473.9	4.5	U.S.

[1]Score based on scale ranging from 0 to 1,000, with mean of 500 and standard deviation of 100.
[2]Standard error jacknifed.
[3]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from Archie E. Lapointe, Nancy A. Mead, and Gary W. Phillips, *A World of Differences: International Assessment of Educational Progress* (Princeton: Educational Testing Service, 1989), 14, 84.

Table B.22
Mean scores and means compared with United States: International Assessment of Educational Progress, 13-year-olds (60 items—science proficiency)

Educational system and rank by mean	Proficiency score Mean[1]	Standard error[2]	Mean compared with U.S.[3]
Canada (British Columbia)	551.3	2.1	higher
Korea	549.9	2.9	higher
United Kingdom	519.5	3.7	higher
Canada (Quebec: English)	515.3	2.8	higher
Canada (Ontario: English)	514.7	2.7	higher
Canada (Quebec: French)	513.4	3.3	higher
Canada (New Brunswick: English)	510.5	2.7	higher
Spain	503.9	4.3	higher
United States	478.5	3.5	U.S.
Ireland	469.3	3.5	same
Canada (Ontario: French)	468.3	2.2	same
Canada (New Brunswick: French)	468.1	3.9	same

[1]Score based on scale ranging from 0 to 1,000, with a mean of 500 and standard deviation of 100.
[2]Standard error jackknifed.
[3]Based on Bonferroni adjusted t-test for comparisons with the United States.

SOURCE: Data from Archie E. Lapointe, Nancy A. Mead, and Gary W. Phillips, *A World of Differences: International Assessment of Educational Progress* (Princeton: Educational Testing Service, 1989), 36, 84.

Appendix C

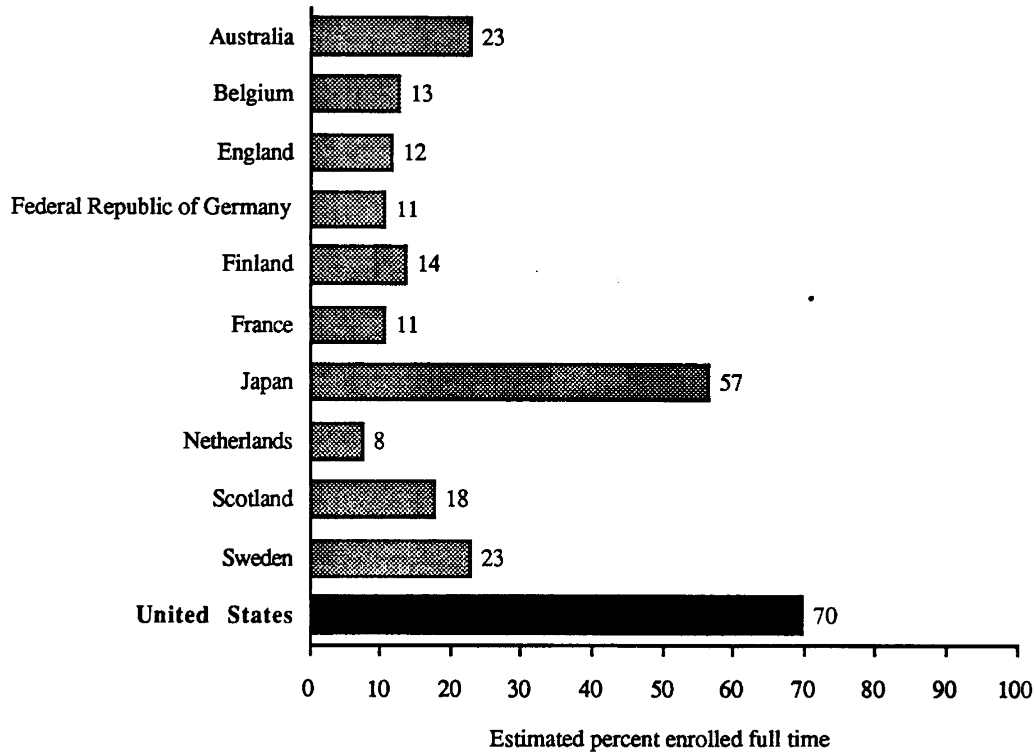
Secondary School Retention Rates

Secondary School Retention Rates

The countries participating in the international achievement studies have distinct national policies with regard to advancing students through the secondary educational system. The United States and several other countries attempt to enroll, retain, and graduate as many secondary school age students as possible. In this country, almost all secondary students attend comprehensive high schools. Although some other countries have high enrollments of this age group, students may attend any one of several types of learning institutions, only some of which are designed around academic curriculum. Still other countries significantly limit access to academic secondary schooling programs. For purposes of the international assessments, groups of students attending particular types of institutions may be excluded from the design sample, and therefore, countries may not be sampling comparable pools of students. The result is neither a representative sample of the age cohort, nor a representative sample of students in any kind of school during the last year of secondary school.

With the exception offigure C5, data in the following set offigures were drawn from the international survey reports themselves. They are not consistent because retention estimates may not have been calculated on the same basis from study to study or country to country. The most recent data (figure C.S), from the Organization for Economic Cooperation and Development (OECD), may more closely re~ect current trends. Generally, however, the data on student retention patterns must be viewed with considerable caution because the estimates from country to country and study to study may not be predicated on the same sets of assumptions.

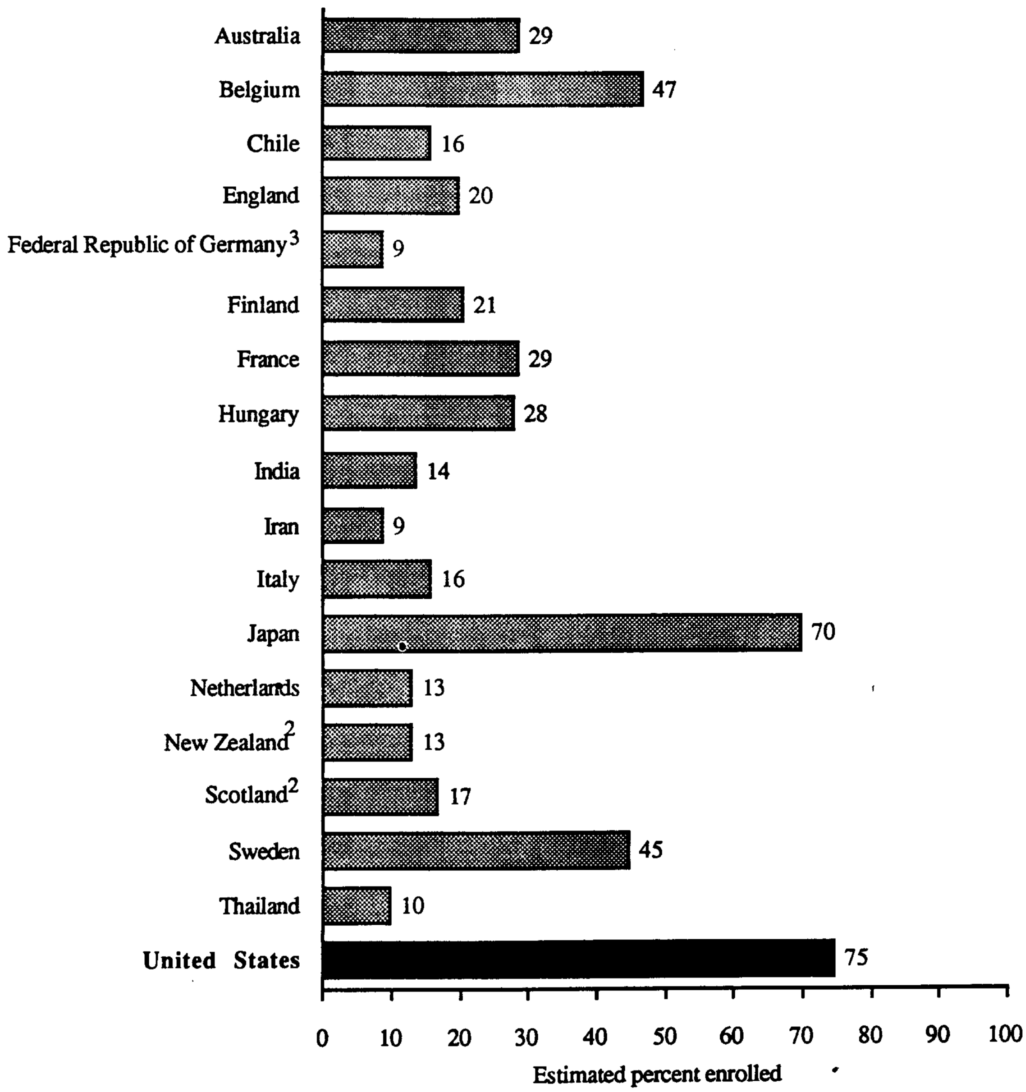
Figure C.1—Estimated percentage of age group[1] enrolled full time in the last year of secondary school: First International Mathematics Study, 1963–64



[1]The age at which students typically attain the last year of secondary school varies from country to country, ranging from 17 to 20 years.

SOURCE: Data from Torsten Husen, *International Study of Achievement in Mathematics*, Vol. II (New York: John Wiley & Sons, Inc., 1967), table 3.34.

Figure C.2—Estimated percentage of age group[1] enrolled in last year of secondary school: First International Science Study, 1969



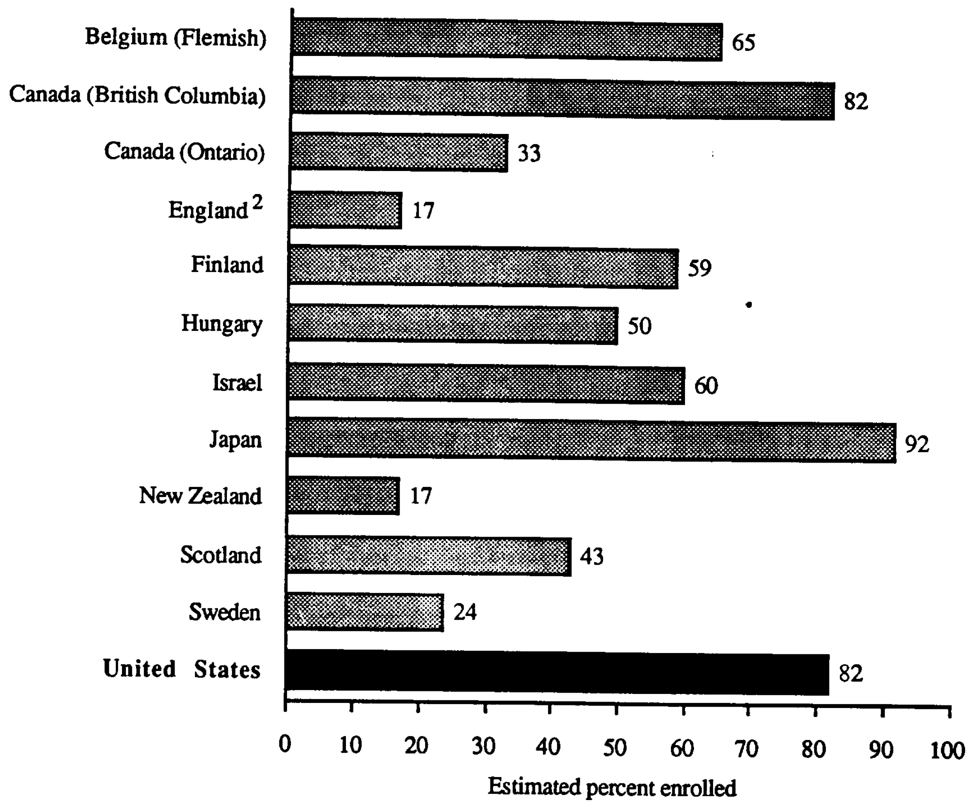
[1]The age at which students typically attain the last year of secondary school varies from country to country, ranging from 17 to 20 years.

[2]This figure is for the terminal grade of the secondary school system (i.e., the grade that was sampled). In both New Zealand and Scotland, it is possible to proceed to university from the previous grade.

[3]This figure is for the Gymnasia only; there are many students of this age in other schools such as higher technical schools.

SOURCE: Data from L.C. Comber and John P. Keeves, *Science Education in Nineteen Countries* (New York: John Wiley & Sons, Inc., 1973), table 4.1.

Figure C.3—Estimated percentage of age group[1] enrolled in last year of secondary school: Second International Mathematics Study, 1980–82

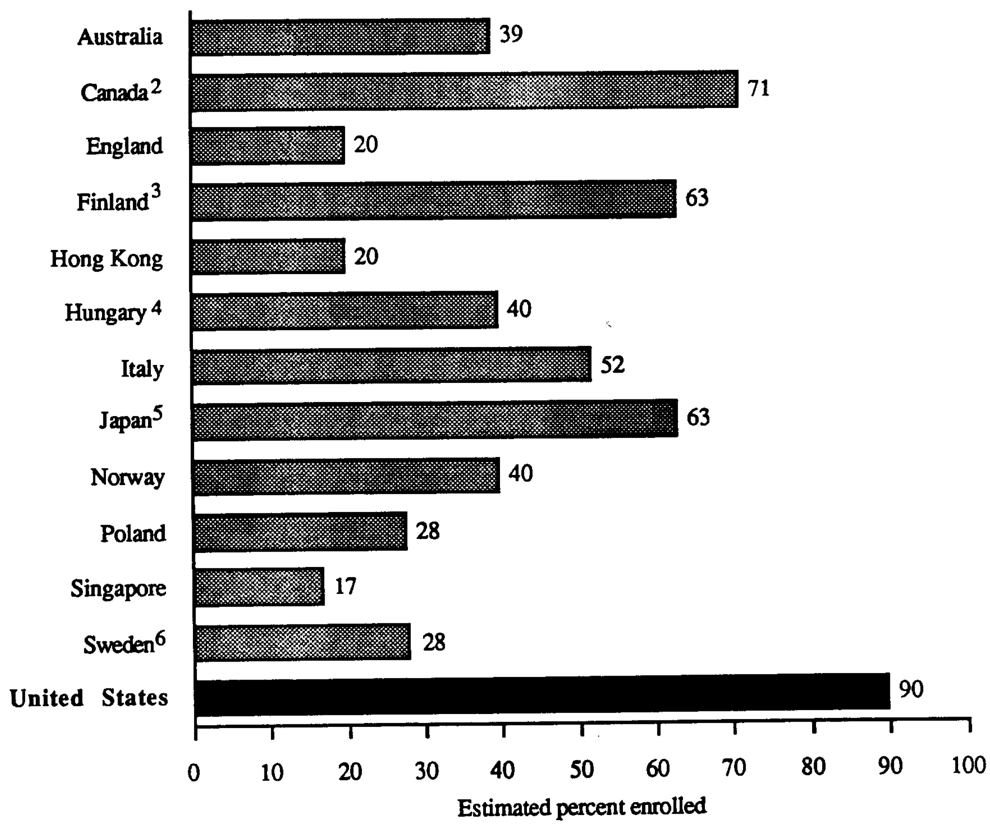


[1]The age at which students typically attain the last year of secondary school varies from country to country, ranging from 17 to 20 years.

[2]Includes data from Wales.

SOURCE: Data from Kenneth Travers, "The Second International Mathematics Study: Overview of Major Findings" (unpublished paper, Champaign-Urbana: University of Illinois, 1986), table 4.2.1.

Figure C.4—Estimated percentage of age group[1] in last year of secondary school: Second International Science Study, 1983–86



[1]The age at which students typically attain the last year of secondary school varies from country to country, ranging from 17 to 20 years.

[2]Data for English-speaking students only.

[3]Includes an estimated 18 percent of age group attending vocational schools.

[4]Includes an estimated 22 percent of age group attending vocational schools.

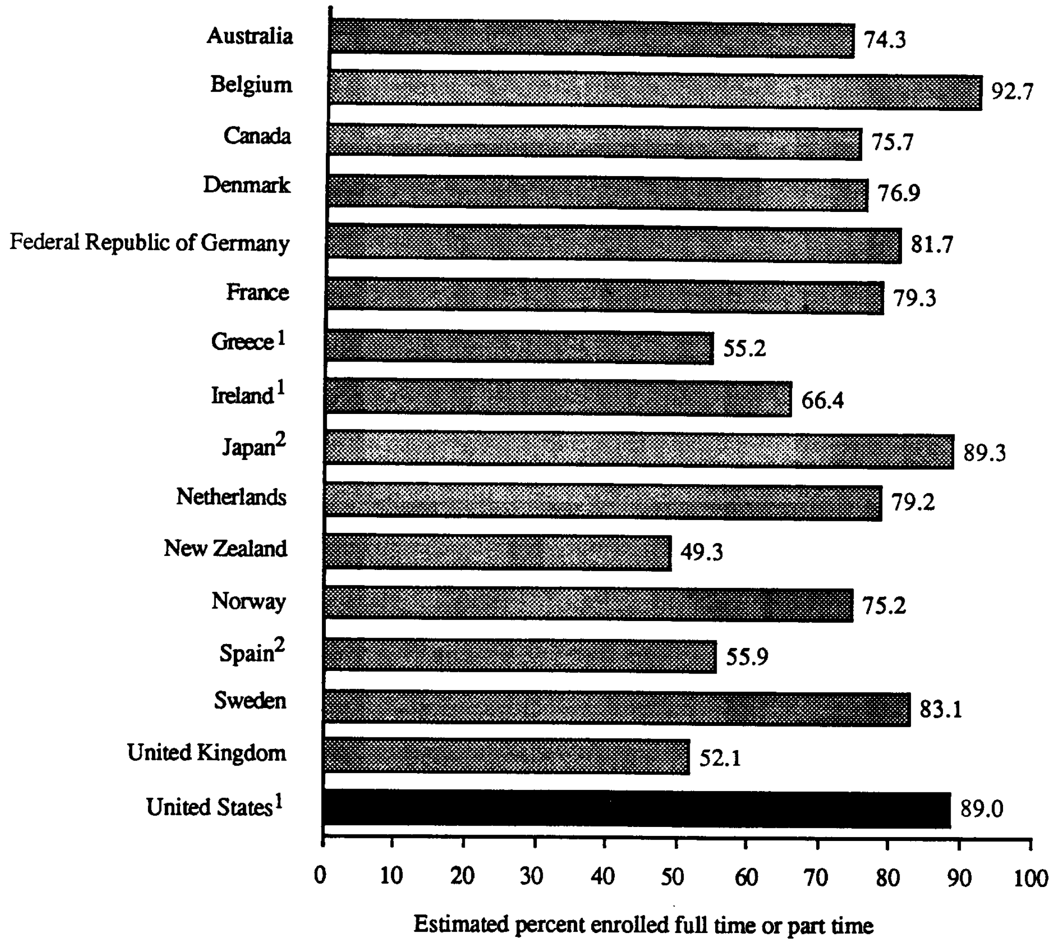
[5]Excludes vocational school enrollments.

[6]In Sweden, 90 percent of the age group are enrolled in upper secondary education (grades 10–12), and 80 percent complete upper secondary education (grade 11). Thirteen percent of the age group are enrolled in science tracks and 15 percent in non-science tracts in Grade 12. The remainder take a 2-year vocational or general track and leave school after Grade 11.

NOTES: No data available for Korea; Data in this figure may differ from table D.10 due to the preliminary nature of the publication (see source below) from which these estimates are drawn.

SOURCE: Data from International Association for the Evaluation of Education Achievement (IEA), *Science Achievement in Seventeen Countries: A Preliminary Report* (Oxford: Pergamon Press, 1988), table 1B.

Figure C.5—Estimated percentage of 17-year-olds enrolled in school full time or part time at the secondary level: 1987–88



[1]1986–87 data.

[2]Does not include 17-year-olds who happen to be enrolled in postsecondary studies.

SOURCE: Data from Organization for Economic Cooperation and Development (OECD), *Education in OECD Countries: 1987–88, A Compendium* (Paris: OECD, 1990), table 4.2.

Appendix D

Age Distributions and Related Characteristics of Test Takers

Table D.1—Mean age and standard deviation of the age distribution of 13-year-old-sample: First International Mathematics Study

Country	Mean age[1]	Standard deviation[2]
Australia	13:03	7.7
Belgium	14:00	8.8
England	14:04	4.2
Finland	13:11	7.3
France	13:07	7.8
Federal Republic of Germany	13:08	6.6
Netherlands	13:01	11.6
Israel	13:11	5.6
Japan	13:05	3.4
Scotland	14:00	5.4
Sweden	13:08	4.9
United States	14:00	6.8

[1]Mean age in years and months.
[2]Standard deviation in months.

SOURCE: Data from T. Husen, *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*, Vol. I (New York: John Wiley & Sons, Inc., 1967), 270–73.

Table D.2—Mean age and standard deviation of the age distribution of last-year secondary sample (mathematics students): First International Mathematics Study

Country	Mean age[1]	Standard deviation[2]
Australia	17:02	9.2
Belgium	18:01	11.6
England	17:11	7.5
Finland	19:01	10.6
France	18:07	13.7
Federal Republic of Germany	19:10	8.4
Netherlands	18:02	11.7
Israel	18:02	8.5
Japan	17:08	3.6
Scotland	17:06	8.0
Sweden	19:07	10.9
United States	17:09	6.3

[1]Mean age in years and months.
[2]Standard deviation in months.

SOURCE: Data from T. Husen, *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*, Vol. I (New York: John Wiley & Sons, Inc., 1967), 270–73.

Table D.3—Mean age and standard deviation of the age distribution of last-year secondary sample (non-mathematics students): First International Mathematics Study

Country	Mean age[1]	Standard deviation[2]
Belgium	18:00	11.2
England	17:11	6.8
Finland	19:02	10.8
France	18:09	12.8
Federal Republic of Germany	19:09	8.8
Netherlands	18:07	11.3
Japan	17:08	3.7
Scotland	17:01	6.2
Sweden	19:07	11.3
United States	17:10	7.3

[1]Mean age in years and months.
[2]Standard deviation in months.

SOURCE: Data from T. Husen, *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*, Vol. I (New York: John Wiley & Sons, Inc., 1967), 270–73.

Table D.4—Percentage of 10-year-old sample in different grades: First International Science Study

Country	Mean age[1]	Percent of sample in grade 3	Percent of sample in grade 4	Percent of sample in grade 5	Percent of sample in grade 6
Belgium (Flemish)	11:00	0	1	88	11
Belgium (French)	10:06	0	20	75	4
Chile	10:05	15	38	41	6
England	10:06	0	0	48	52
Federal Republic of Germany[2]	10:05	8	44	51	1
Finland	10:06	23	77	0	0
Hungary	10:07	0	71	29	0
India	10:07	19	30	30	13
Iran	10:03	3	78	19	0
Italy	10:07	0	4	96	0
Japan	10:05	0	0	100	0
Netherlands	10:06	5	37	58	0
Scotland	10:06	0	1	35	63
Sweden	10:05	47	53	0	0
Thailand	10:07	2	52	42	3
United States[2]	10:07	2	37	66	0

[1]Mean age in years and months.
[2]Data provided sums to over 100 percent.

NOTE: May not sum to 100 percent because some sampled students may fall outside grades reported in table.

SOURCE: Data from L.C. Comber and John P. Keeves, *Science Education in Nineteen Countries* (New York: John Wiley & Sons, 1973), 48.

Table D.5—Percentage of 14-year-old sample in different grades: First International Science Study

Country	Mean age*	Percent of sample in grade 6	Percent of sample in grade 7	Percent of sample in grade 8	Percent of sample in grade 9	Percent of sample in grade 10
Australia	14:05	0	0	5	41	52
Belgium (Flemish)	15:00	0	0	7	86	3
Belgium (French)	14:07	0	0	35	65	0
Chile	14:06	20	14	28	20	4
England	14:07	0	0	0	48	52
Federal Republic of Germany	14:05	2	9	44	45	0
Finland	14:06	4	36	60	0	0
Hungary	14:05	0	0	77	23	0
India	14:03	14	28	37	13	0
Iran	14:04	0	4	93	3	0
Italy	14:08	0	0	45	53	2
Japan	14:05	0	0	0	100	0
Netherlands	14:06	0	18	71	10	0
New Zealand	14:06	0	0	0	26	72
Scotland	14:07	0	0	1	45	54
Sweden	14:06	0	49	51	0	0
Thailand	14:06	1	4	33	55	6
United States	14:07	0	2	26	72	0

*Mean age in years and months.

NOTE: May not sum to 100 because some sampled students may fall outside grades reported in table.

SOURCE: Data from L.C. Comber and John P. Keeves, *Science Education in Nineteen Countries* (New York: John Wiley & Sons, Inc., 1973), 48.

Table D.6—Mean age of 13-year-old sample[1]: Second International Mathematics Study

Country	Mean age[2]	Standard deviation[3]
Belgium (Flemish)	14:02	8
Belgium (French)	14:08	11
Canada (British Columbia)	14:00	6
Canada (Ontario)	14:01	7
England and Wales	14:01	4
Finland	13:08	5
France	14:01	8
Hong Kong	13:02	11
Hungary	14:02	13
Israel	14:00	5
Japan	13:05	4
Luxembourg	14:05	9
Netherlands	14:04	8
Nigeria	16:07	38
New Zealand	14:00	5
Scotland	14:00	4
Swaziland	15:07	23
Sweden	13:09	4
Thailand	14:02	9
United States	14:01	6

[1]Students in grade where majority has attained age 13:00 to 13:11 years by the middle of the school year.
[2]Mean age in years and months.
[3]Standard deviation in months.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics*, Vol. II (Oxford: Pergamon Press, 1989), 64.

Table D.7—Mean age of last-year secondary sample: Second International Mathematics Study

Country	Mean age[1]	Standard deviation[2]
Belgium (Flemish)	18:01	10
Belgium (French)	18:04	11
Canada (British Columbia)	17:09	6
Canada (Ontario)	18:05	14
England and Wales	18:01	4
Finland	18:06	6
Hong Kong	18:05	12
Hungary	18:01	4
Israel	17:09	5
Japan	18:01	4
New Zealand	17:08	6
Scotland	16:09	7
Sweden	19:02	9
Thailand	18:02	9
United States	17:08	7

[1]Mean age in years and months.
[2]Standard deviation in months.

SOURCE: Data from David Robitaille and Robert Garden, eds., *The International Association for the Evaluation of Education Achievement (IEA) Study of Mathematics II: Contexts and Outcomes of School Mathematics*, Vol. II (Oxford: Pergamon Press, 1989), 64.

Table D.8—Mean age of 10-year-old sample and selected schooling characteristics: Second International Science Study

Country	Age entering formal school	Grade tested in study	Percent in school	Mean age[1]	Standard deviation[2]
Australia	6	4,5,6	99	10:06	3.3
Canada (English)	6	5	99	11:01	7.1
England	5	5	99	10:03	3.6
Finland	7	4	99	10:10	4.1
Hong Kong	6	4	99	10:05	9.8
Hungary	6	4	99	10:03	5.2
Italy	6	5	99	10:09	5.2
Japan	6	5	99	10:07	3.5
Korea	6	5	99	11:02	7.4
Norway	7	4	99	10:11	4.0
Philippines	7	5	97	11:01	11.3
Poland	7	4	99	10:11	5.4
Singapore	6	5	99	10:10	4.9
Sweden (A)	7	3	99	9:10	3.7
Sweden (B)	7	4	99	10:10	4.1
United States	6	5	99	11:03	6.9

[1]Mean age in years and months.
[2]Standard deviation in months.

SOURCE: Data from T. N. Postlethwaite, *Second International Science Study*, Vol. II Draft (Hamburg, July 1990), 6, 7.

Table D.9—Mean age of 14-year-old sample and selected schooling characteristics: Second International Science Study

Country	Grade tested	Percent in school	Mean age[1]	Standard deviation[2]
Australia	8,9,10	98	14:05	3.3
Canada (English)	9	99	15:00	6.1
England	9	98	14:02	3.6
Finland	8	99	14:10	4.1
Hong Kong	8	99	14:07	10.9
Hungary	8	92	14:03	4.7
Italy (A)	8	99	13:11	8.6
Italy (B)	9	72	14:08	3.2
Japan	9	99	14:07	3.5
Korea	9	99	15:00	7.2
Netherlands	9	99	15:06	12.5
Norway	9	99	15:10	4.0
Philippines	9	60	16:01	18.9
Poland	8	91	15:00	5.8
Singapore	9	91	15:03	9.2
Sweden (A)	7	99	13:10	4.8
Sweden (B)	8	99	14:10	3.8
Thailand	9	32	15:04	8.9
United States	9	99	15:03	9.1

[1]Mean age in years and months.
[2]Standard deviation in months.

SOURCE: Data from T.N. Postlethwaite, *Second International Science Study*, Vol. II Draft (Hamburg, July 1990), 6, 7.

Table D.10—Mean age of last-year secondary sample and selected schooling characteristics: Second International Science Study

Country	Grade tested	Percent in school	Mean age[1]	Standard deviation[2]
Australia	12	39	17:03	11
Canada (English)	12,13	68	18:03	11
England	13	20	18:00	7
Finland	12	41[3]	18:06	7
Hong Kong (form 6)	12	27	18:03	13
Hong Kong (form 7)	13	20	19:02	11
Hungary	12	18[3]	18:00	4
Italy	12,13	34	19:00	13
Japan	12	63[3]	18:02	4
Korea	12	38[3]	17:09	8
Norway	12	40	18:09	7
Poland	12	28	18:06	5
Singapore	12,13	17	18:01	8
Sweden	12	28[4]	19:00	11
United States	12	83	17:07	9

[1]Mean age in years and months.
[2]Standard deviation in months.
[3]Certain countries excluded vocational students from calculation of students in school:
Finland—63 percent inclusive of vocational, 41 percent exclusive.
Hungary—40 percent inclusive of vocational, 18 percent exclusive. In Hungary, 18 percent of the age group are in academic secondary schools studying science. Forty percent are actually in school.
Japan—89 percent inclusive of vocational, 63 percent exclusive.
Korea—83 percent inclusive of vocational, 38 percent exclusive.
[4]In Sweden, 90 percent of the age group are enrolled in upper secondary education (grades 10–12), and 80 percent complete upper secondary education (grade 11). Thirteen percent are enrolled in science tracks, and 15 percent in non-science tracks in grade 12; hence, the Second Science calculation of 28 percent.

NOTE: Data in this table does not precisely coincide with data in table C.4. Table C.4 is based on information from preliminary report, issued in 1988.

SOURCE: Data from T. N. Postlethwaite, *Second International Science Study*, Vol. II Draft (Hamburg, July 1990), 6, 7.

Table D.11—Mean age of last-year secondary sample (biology) and percentage of students in school taking biology: Second International Science Study

Country	Percent of those in school taking biology	Mean age*
Australia	18	17:03
Canada (English)	28	18:02
Canada (French)	7	17:02
England	4	18:00
Finland	41	18:07
Hong Kong (form 6)	12	18:05
Hong Kong (form 7)	7	19:02
Hungary	3	18:00
Israel	20	17:07
Italy	4	19:05
Japan	12	18:01
Korea	38	17:11
Norway	4	18:11
Poland	9	18:07
Singapore	3	18:00
Sweden	5	18:11
Thailand	7	18:03
United States	12	17:05

*Mean age in years and months.

SOURCE: Data from T. N. Postlethwaite, *Second International Science Study*, Vol. II Draft (Hamburg, July 1990), 6, 7.

Table D.12—Mean age of last-year secondary sample (chemistry) and percentage of students in school taking chemistry: Second International Science Study

Country	Percent of those in school taking chemistry	Mean age[1]
Australia	12	17:03
Canada (English)	25	18:04
Canada (French)	37	17:01
England	5	18:00
Finland	16	18:06
Hong Kong (form 6)	20	18:04
Hong Kong (form 7)	12	19:03
Hungary	1	18:01
Israel	8	17:07
Italy	1	19:02
Japan	16	18:02
Korea	37	17:10
Norway	6	18:11
Poland	9	18:07
Singapore	5	18:00
Sweden	6[2]	19:00
Thailand	7	18:03
United States	2	17:08

[1]Mean age in years and months.

[2]In Sweden, although only 6 percent of the age group studies chemistry, 13 percent were tested.

SOURCE: Data from T. N. Postlethwaite, *Second International Science Study,* Vol. II Draft (Hamburg, July 1990), 6, 7.

Table D.13—Mean age of last-year secondary sample (physics) and percentage of students in school taking physics: Second International Science Study

Country	Percent of those in school taking Physics	Mean age*
Australia	11	17:03
Canada (English)	18	18:04
Canada (French)	35	17:01
England	6	18:00
Finland	14	18:07
Hong Kong (form 6)	20	18:04
Hong Kong (form 7)	12	19:03
Hungary	4	18:00
Israel	12	17:07
Italy	13	19:02
Japan	11	18:02
Korea	14	17:11
Norway	10	18:11
Poland	9	18:07
Singapore	7	18:00
Sweden	13	19:00
Thailand	7	18:02
United States	1	17:10

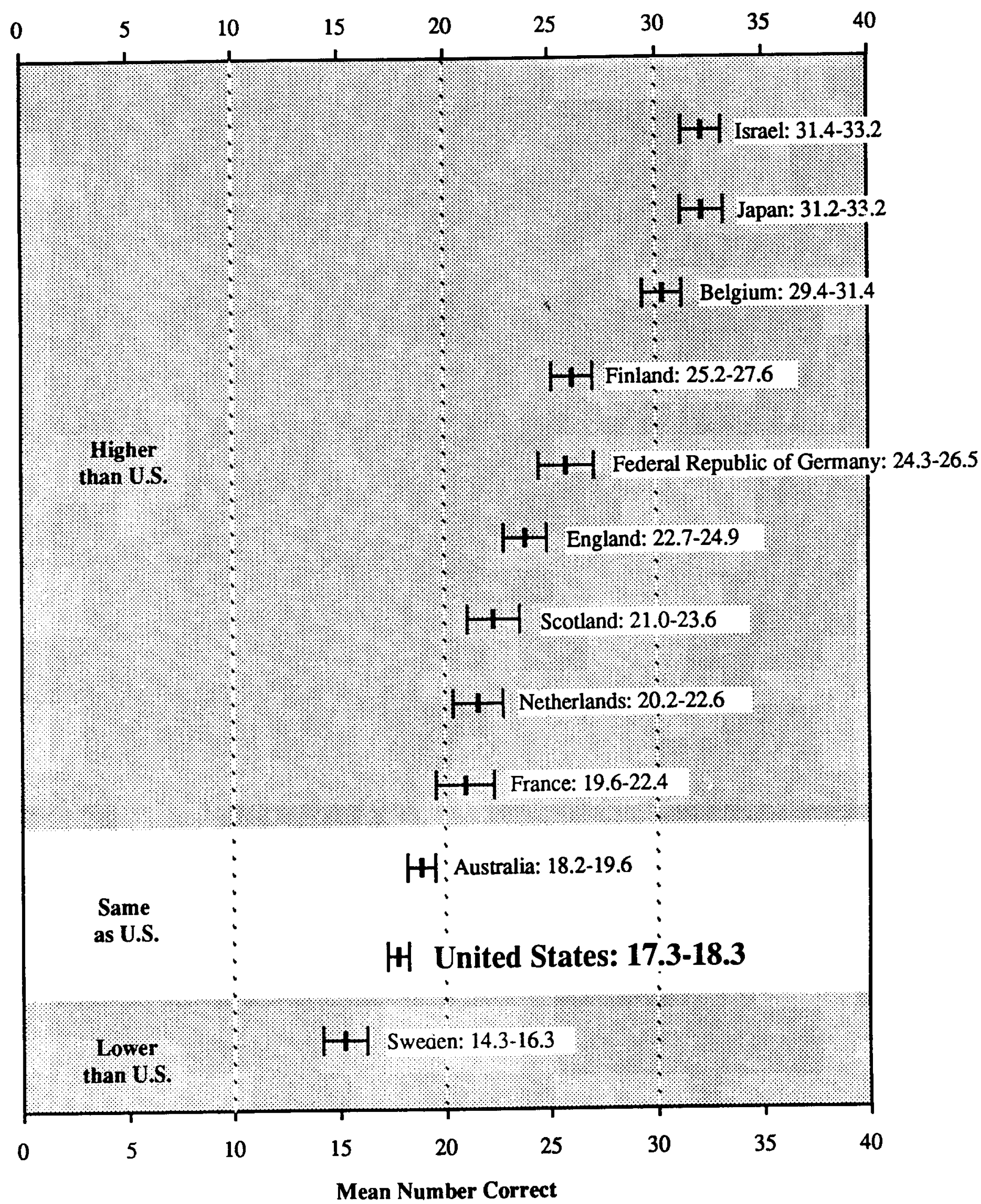
*Mean age in years and months.

SOURCE: Data from T. N. Postlethwaite, *Second International Science Study*, Vol. II Draft (Hamburg, July 1990), 6, 7.

Appendix E

Mean Scores and Confidence Intervals
for Participating Educational Systems

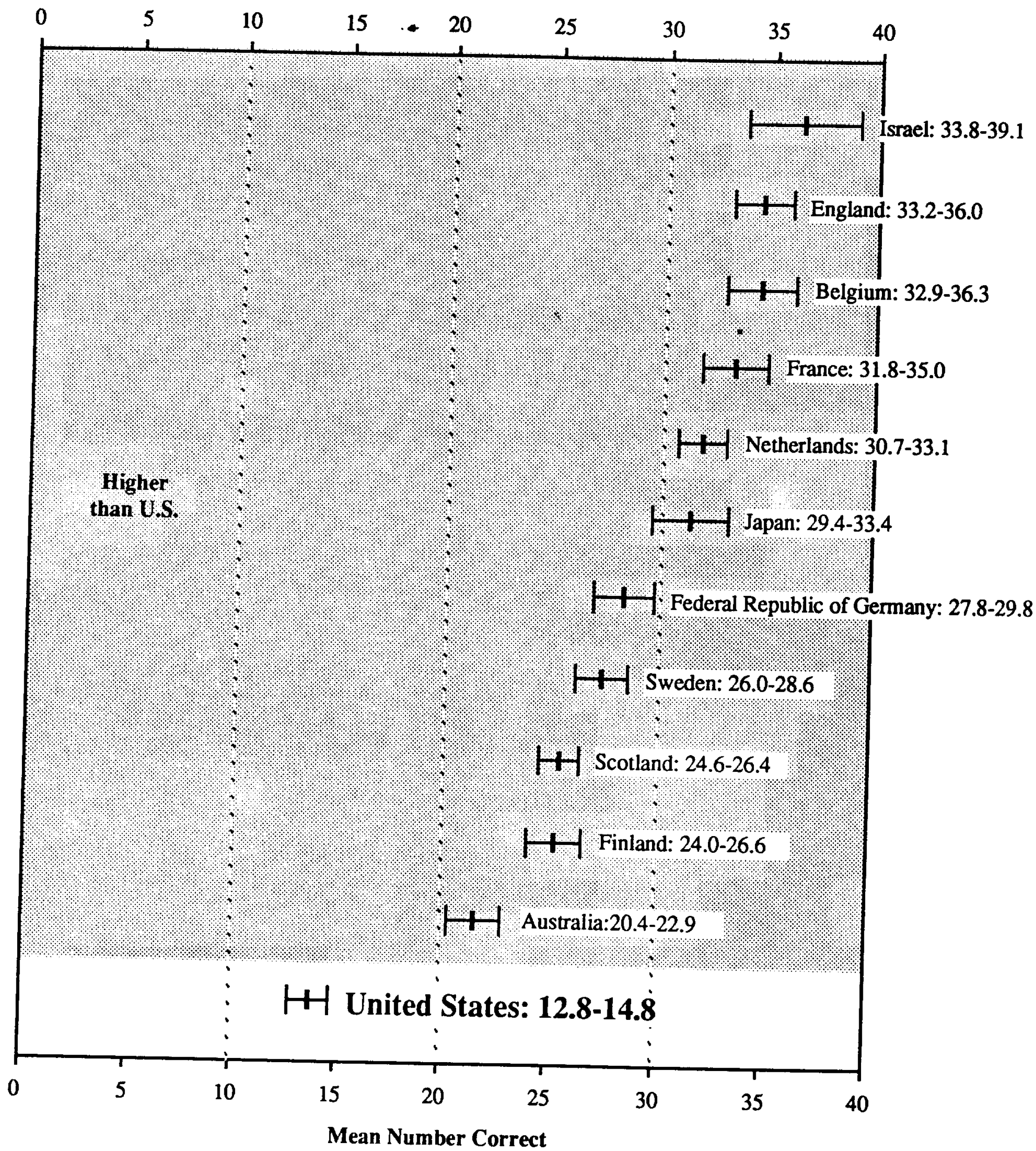
Figure E.1
Mean scores and confidence intervals for participating educational systems: First International Mathematics Study, 13-year-olds (70 items)



NOTE: Mean scores are denoted by the bold vertical (" ❙ "). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 11 comparisons with the United States.

SOURCE: See Appendix B.

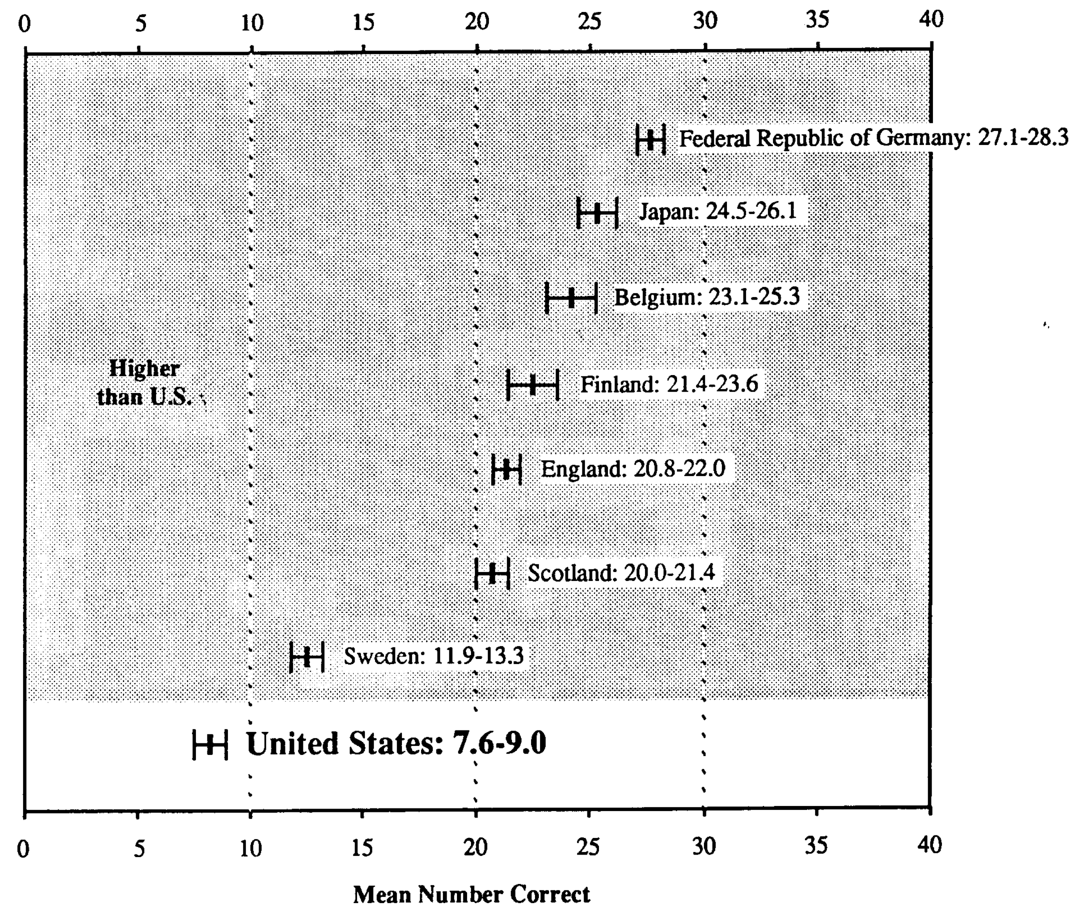
Figure E.2
Mean scores and confidence intervals for participating educational systems: First International Mathematics Study, Last year of secondary school (mathematics students—69 items)



NOTE: Mean scores are denoted by the bold vertical (" | "). The simple 95 percent confidence interval for each mean is denoted by ("|-|-|"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 11 comparisons with the United States.

SOURCE: See Appendix B.

Figure E.3
Mean scores and confidence intervals for participating educational systems: First International Mathematics Study, Last year of secondary school (non-mathematics students—58 items)

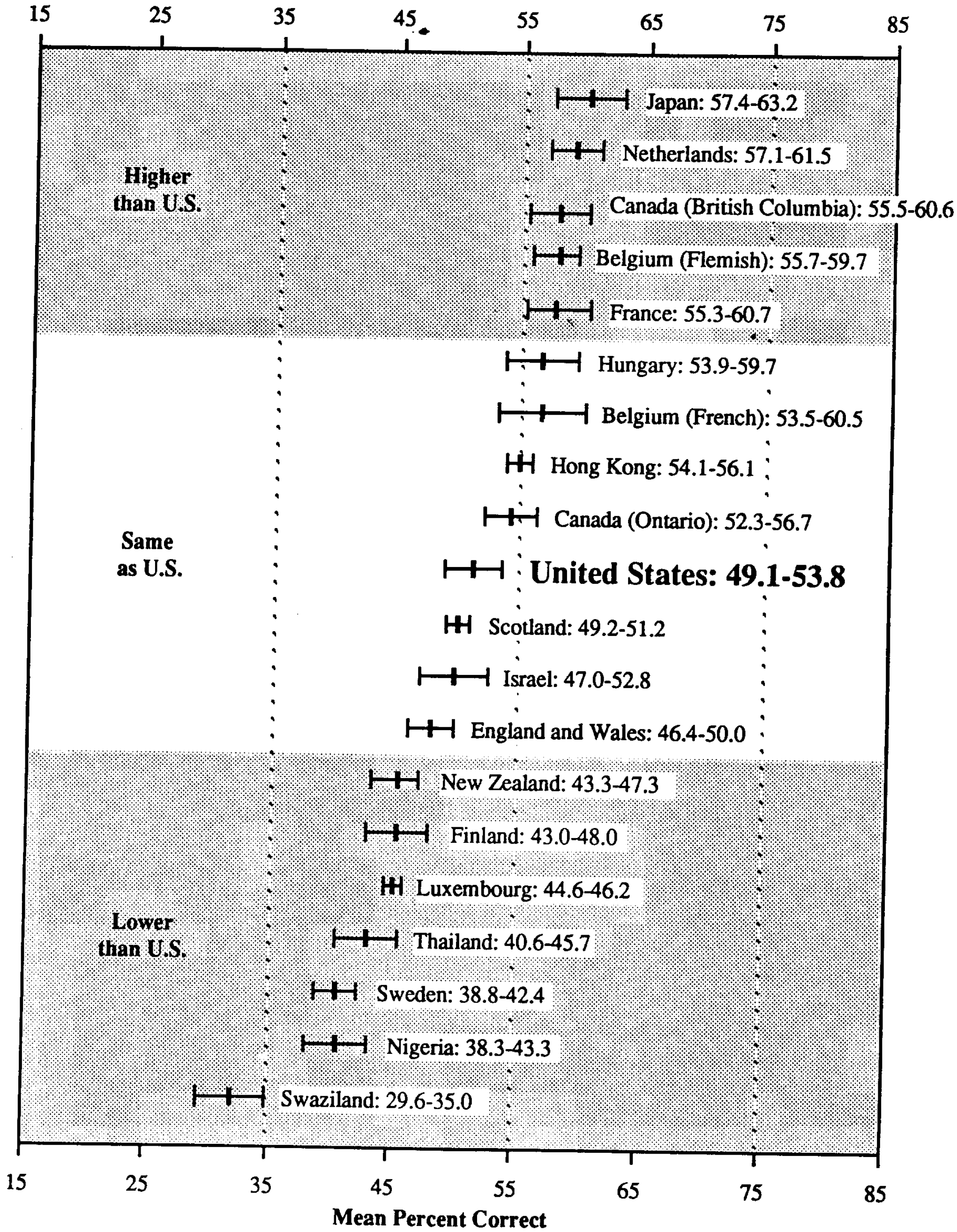


NOTE: On account of missing data confidence intervals could not be calculated for France or the Netherlands.

NOTE: Mean scores are denoted by the bold vertical ("❙"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 7 comparisons with the United States.

SOURCE: See Appendix B.

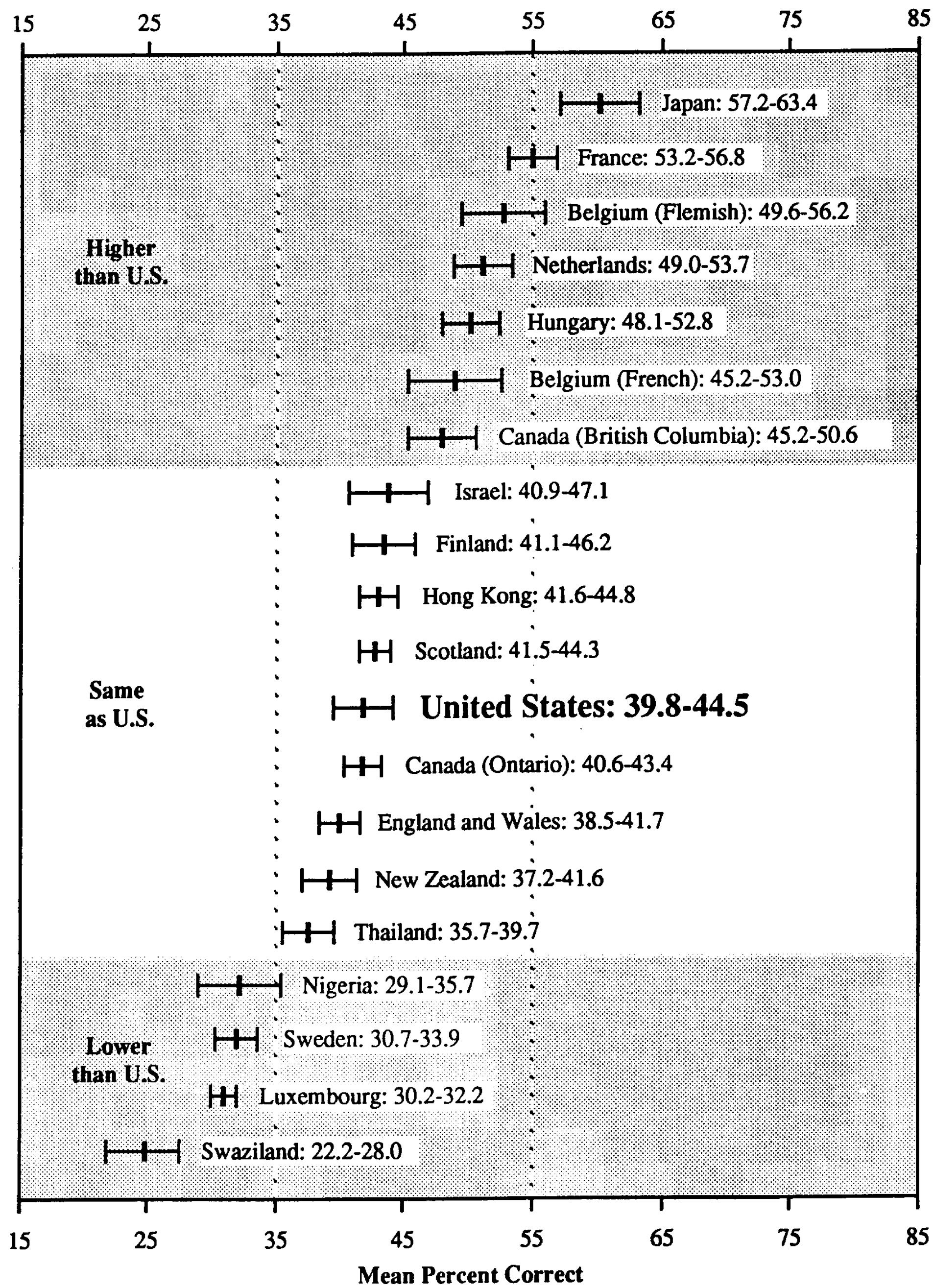
Figure E.4
Mean scores and confidence intervals for participating educational systems: Second International Mathematics Study, 13-year-olds (eighth grade) (46 core items—arithmetic)



NOTE: Mean scores are denoted by the bold vertical (" ❙ "). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 19 comparisons with the United States.

SOURCE: See Appendix B.

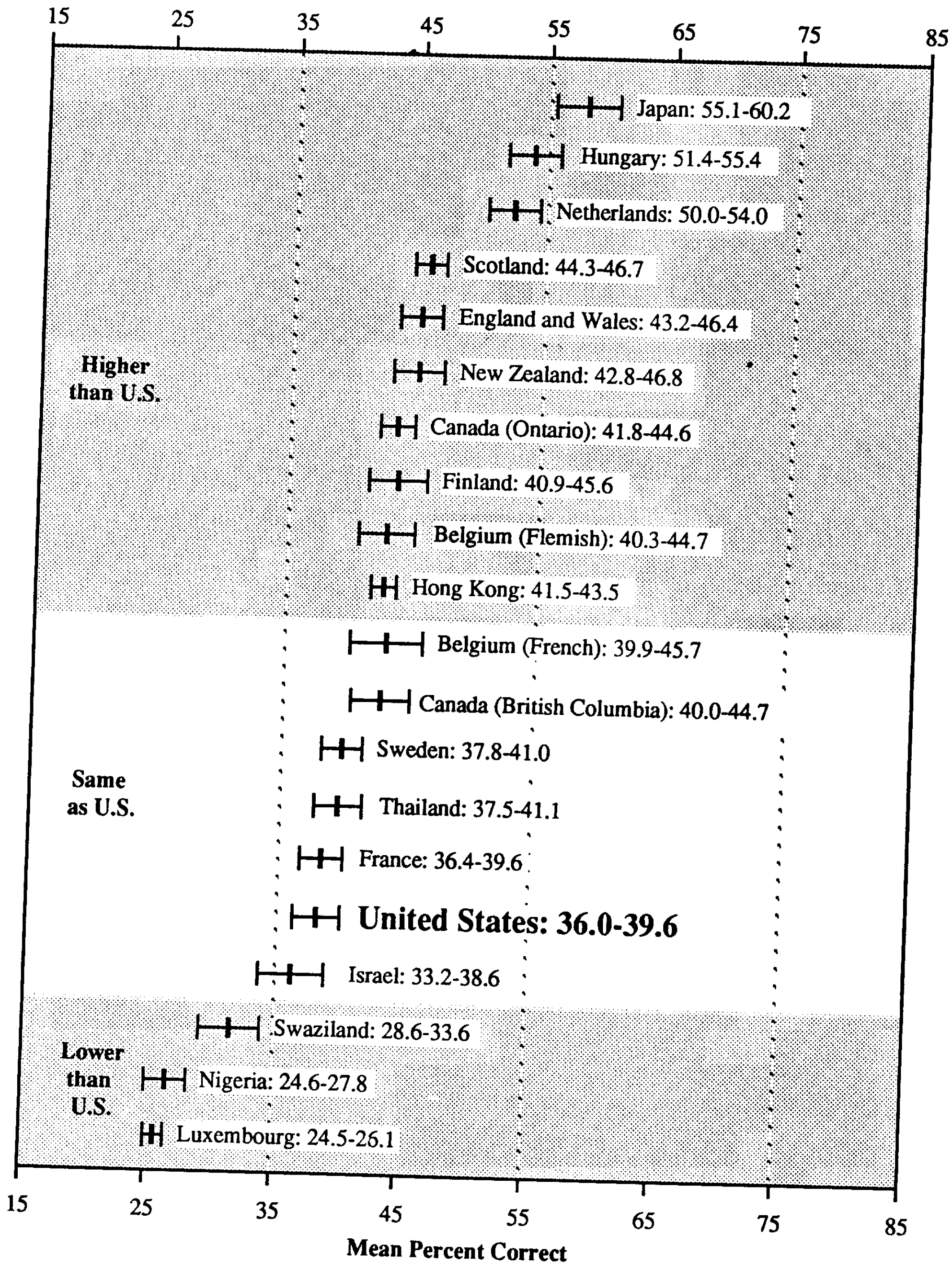
Figure E.5
Mean scores and confidence intervals for participating educational systems: Second International Mathematics Study, 13-year-olds (eighth grade) (30 core items—algebra)



NOTE: Mean scores are denoted by the bold vertical (" | "). The simple 95 percent confidence interval for each mean is denoted by ("|-|-|"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 19 comparisons with the United States.

SOURCE: See Appendix B.

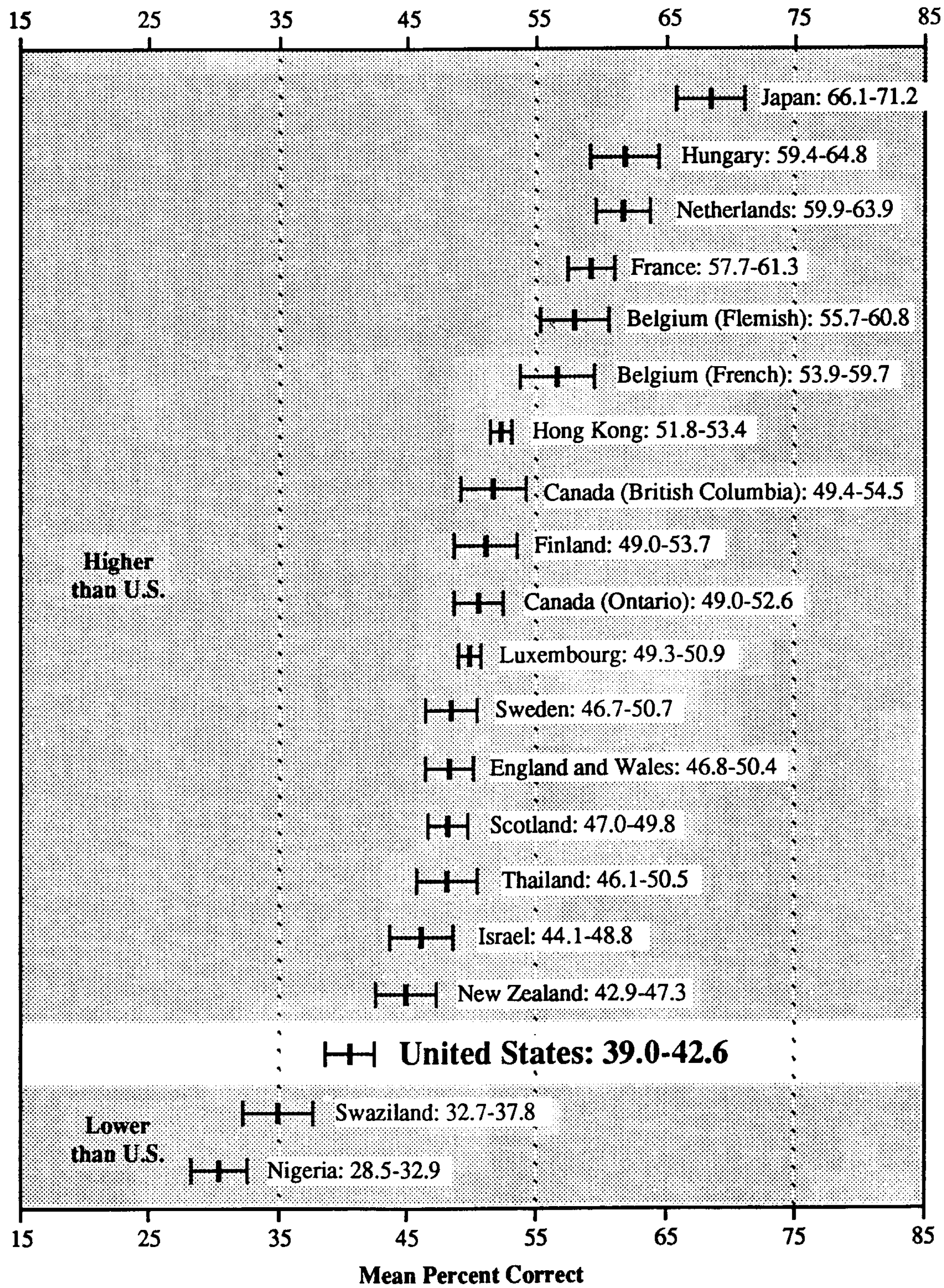
Figure E.6
Mean scores and confidence intervals for participating educational systems: Second International Mathematics Study, 13-year-olds (eighth grade) (39 items—geometry)



NOTE: Mean scores are denoted by the bold vertical ("❙"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 19 comparisons with the United States.

SOURCE: See Appendix B.

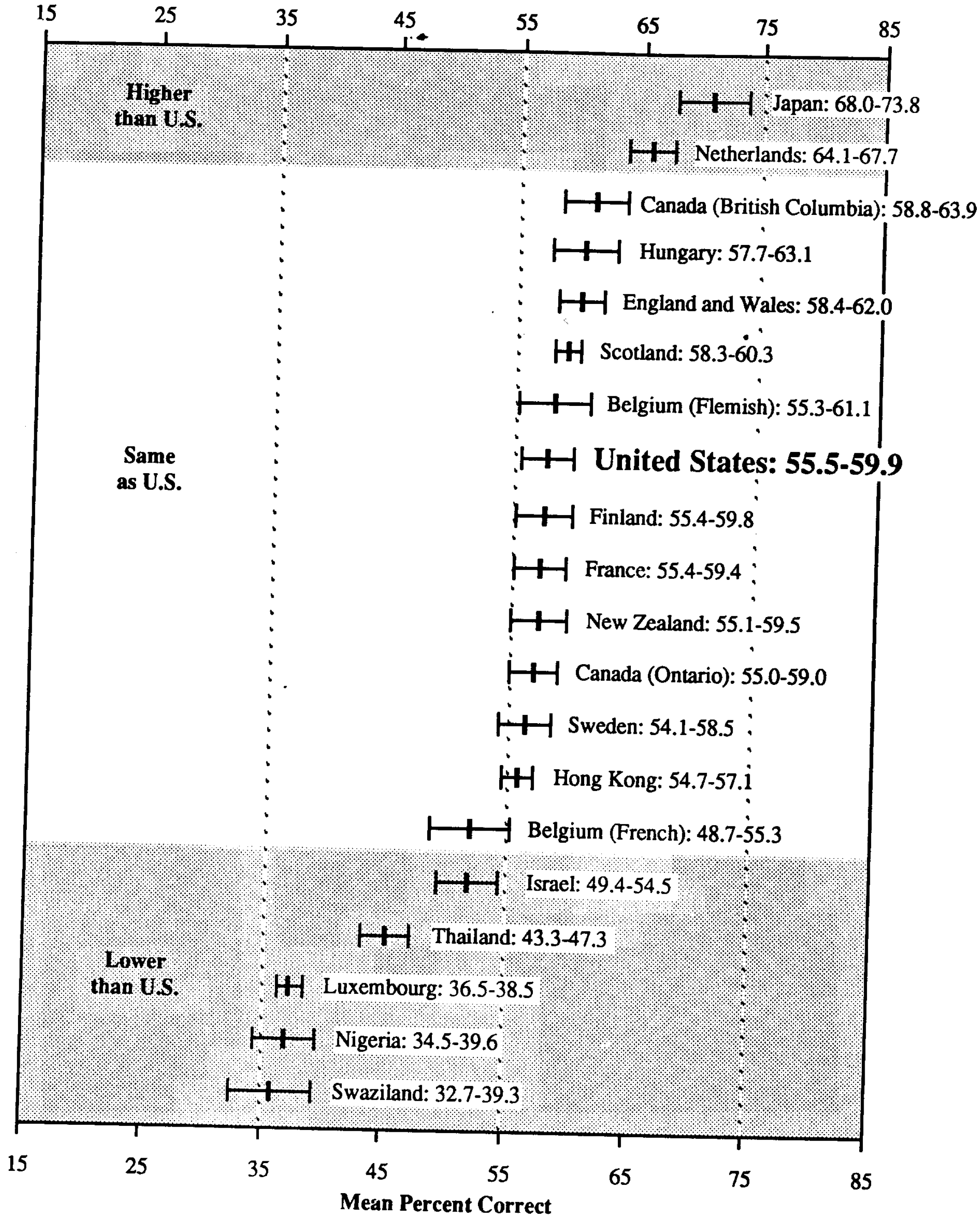
Figure E.7
Mean scores and confidence intervals for participating educational systems: Second International Mathematics Study, 13-year-olds (eighth grade) (24 core items—measurement)



NOTE: Mean scores are denoted by the bold vertical ("|"). The simple 95 percent confidence interval for each mean is denoted by ("|-|-|"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 19 comparisons with the United States.

SOURCE: See Appendix B.

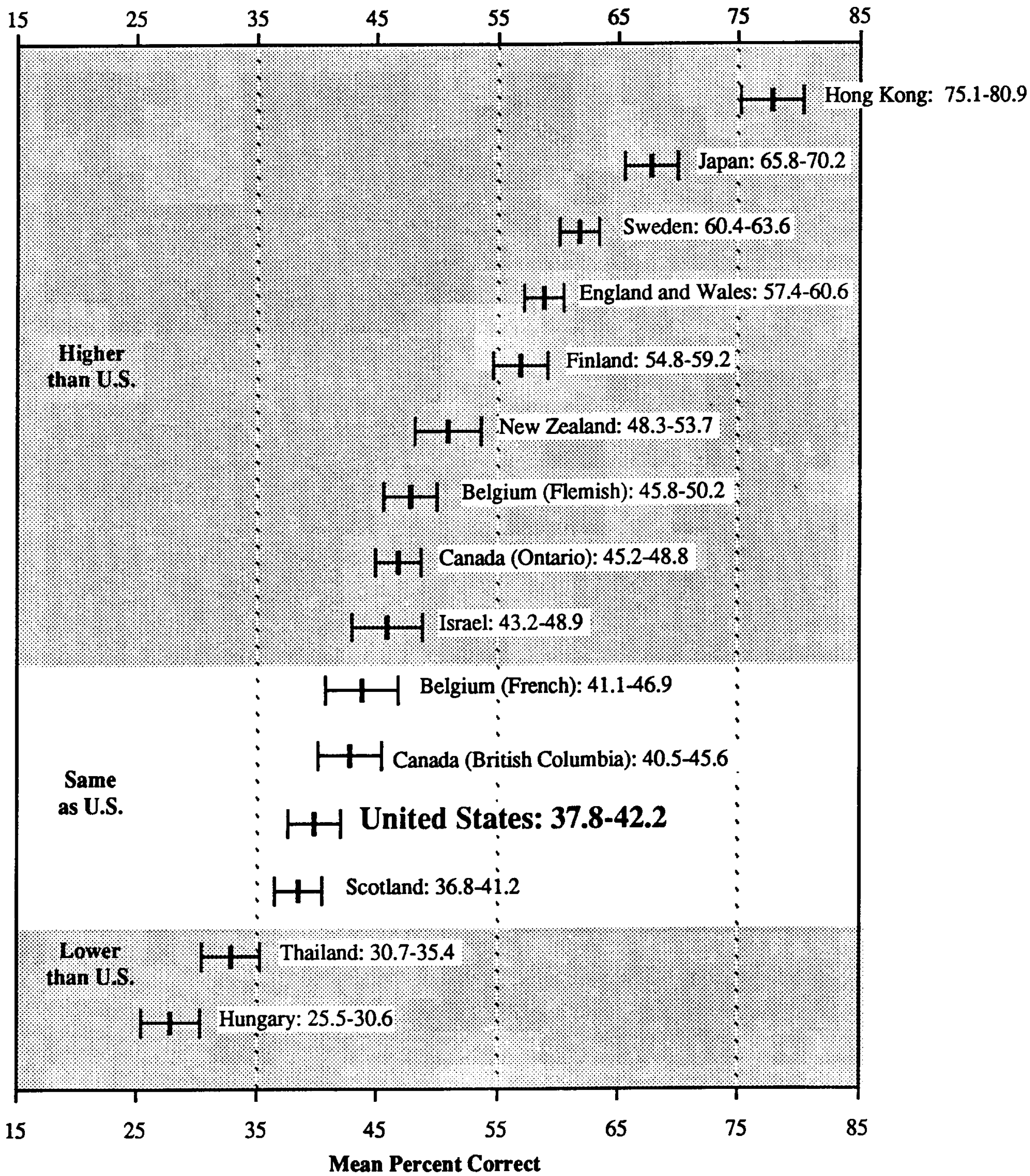
Figure E.8
Mean scores and confidence intervals for participating educational systems: Second International Mathematics Study, 13-year-olds (eighth grade) (18 core items—descriptive statistics)



NOTE: Mean scores are denoted by the bold vertical ("❙"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 19 comparisons with the United States.

SOURCE: See Appendix B.

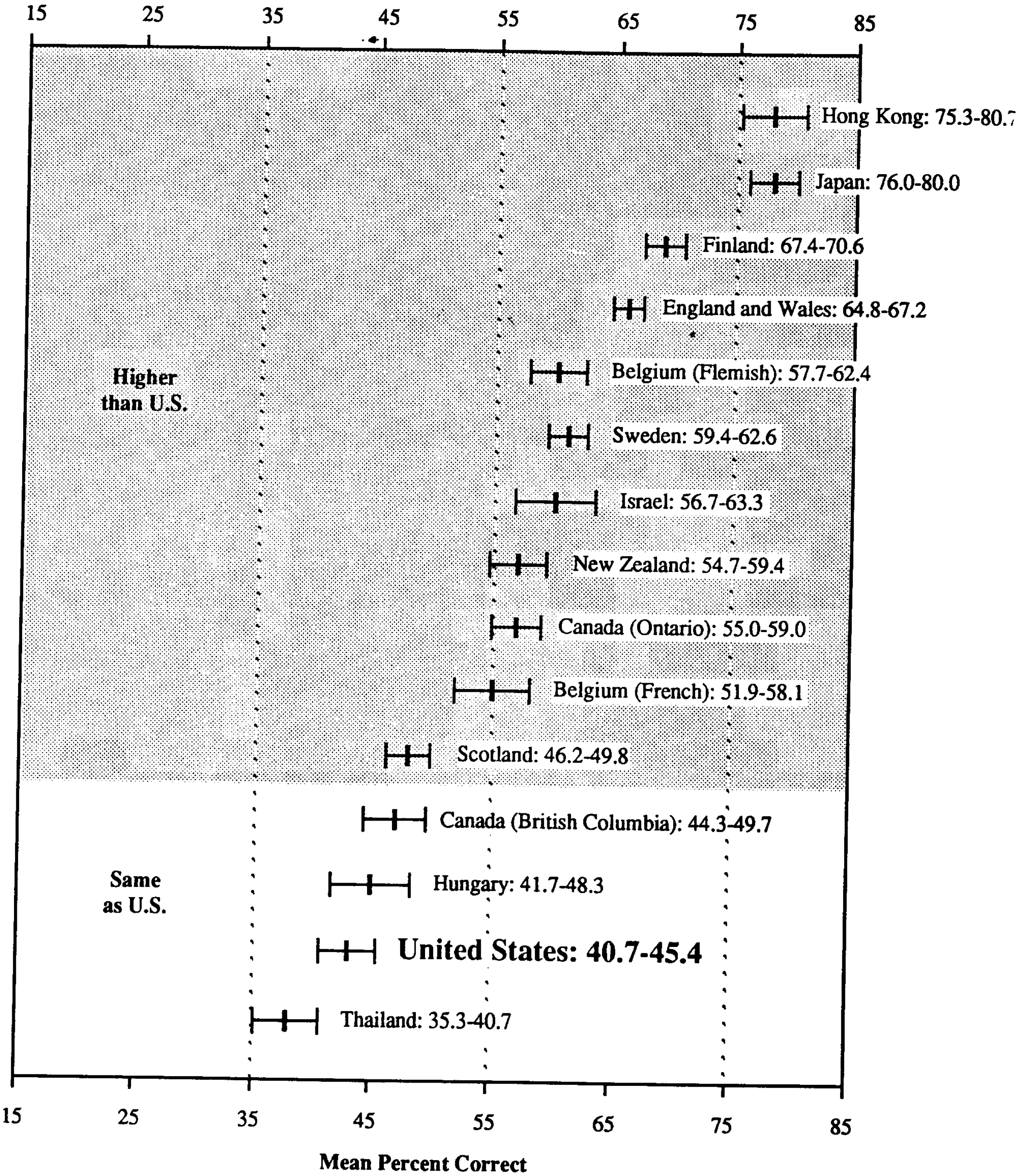
Figure E.9
Mean scores and confidence intervals for participating educational systems: Second International Mathematics Study, Last year of secondary school (17 items—number systems)



NOTE: Mean scores are denoted by the bold vertical (" ❙ "). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 14 comparisons with the United States.

SOURCE: See Appendix B.

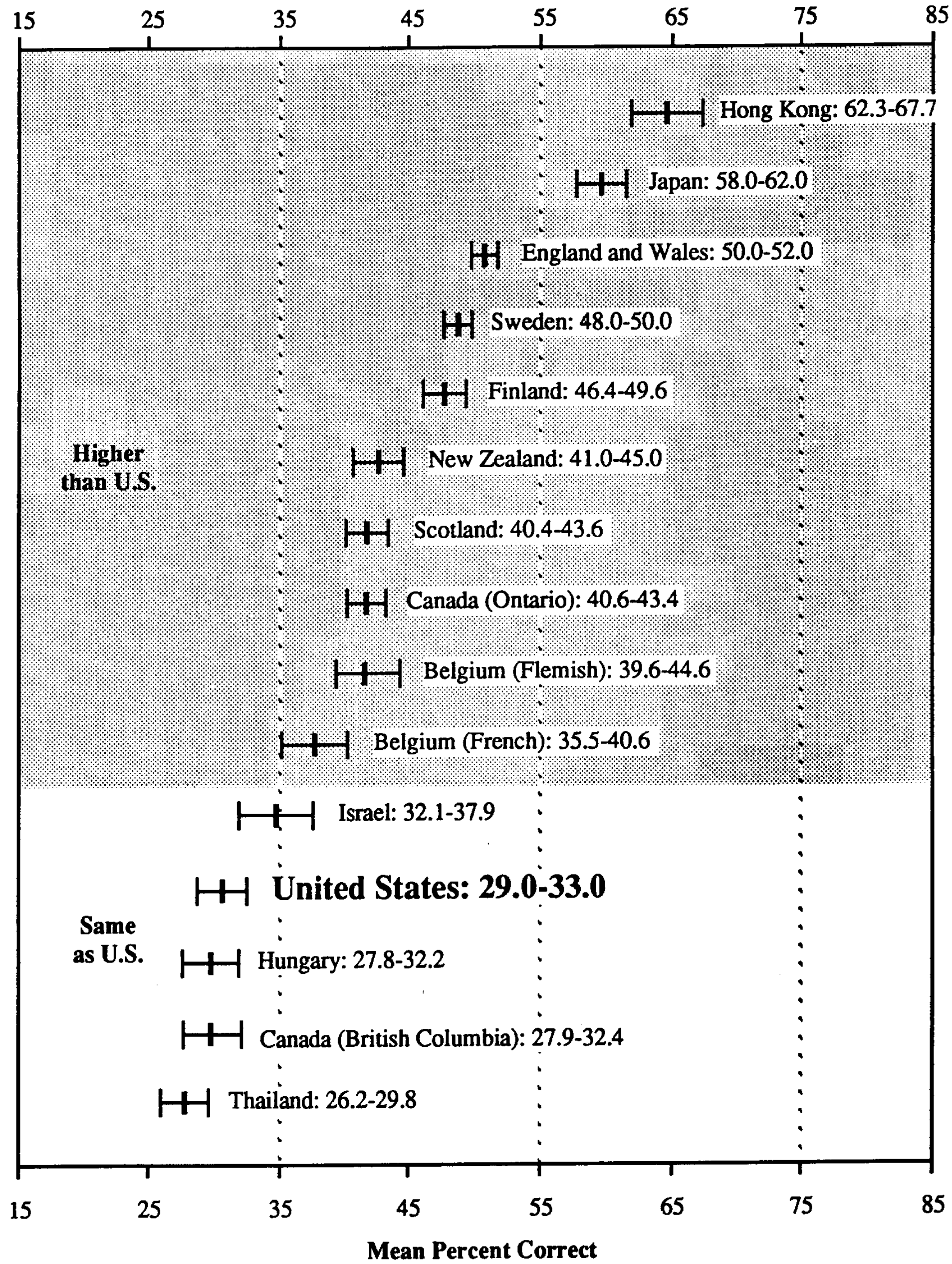
Figure E.10
Mean scores and confidence intervals for participating educational systems: Second International Mathematics Study, Last year of secondary school (26 items—algebra)



NOTE: Mean scores are denoted by the bold vertical ("❙"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 14 comparisons with the United States.

SOURCE: See Appendix B.

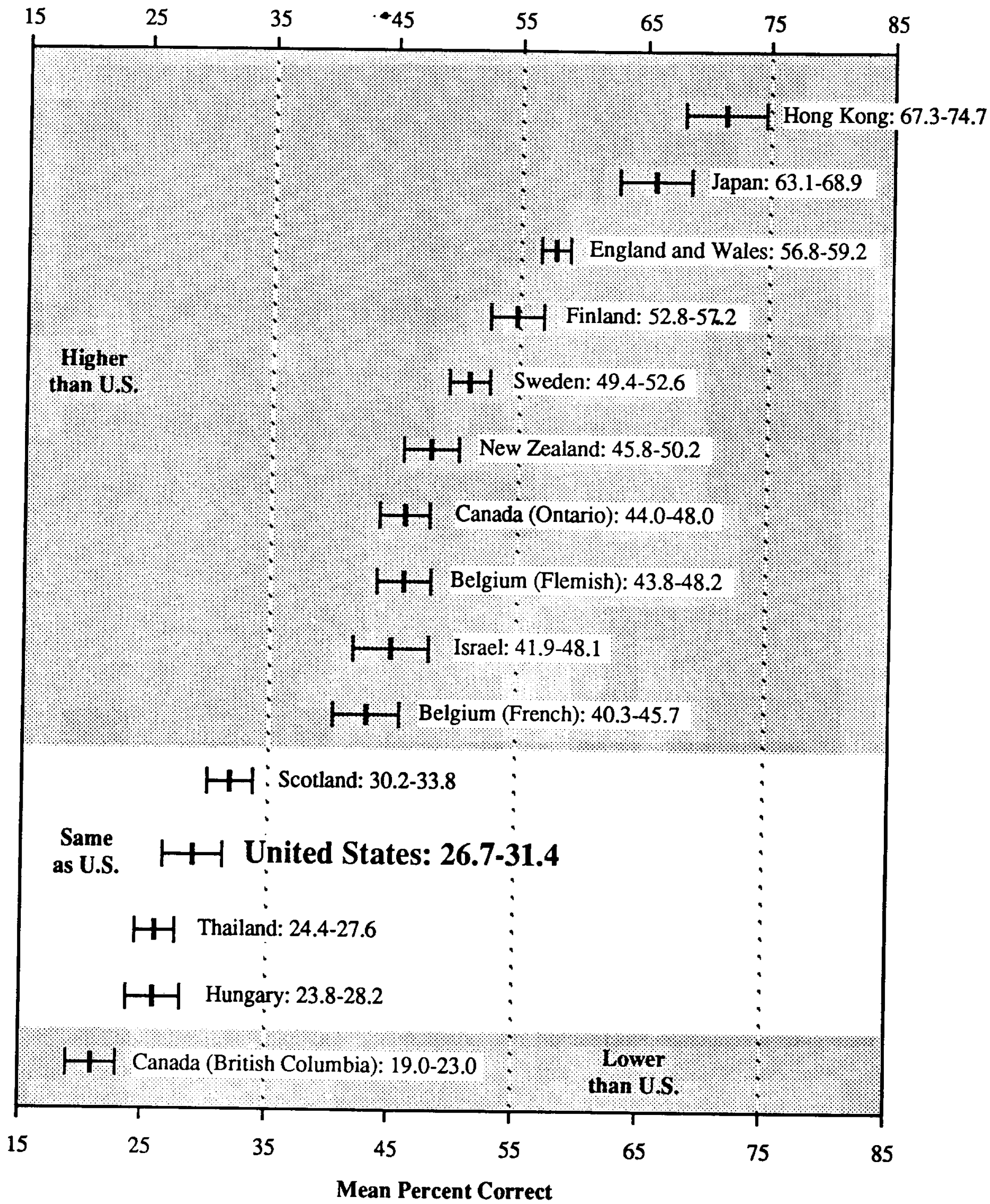
Figure E.11
Mean scores and confidence intervals for participating educational systems: Second International Mathematics Study, Last year of secondary school (26 items—geometry)



NOTE: Mean scores are denoted by the bold vertical ("❙"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 14 comparisons with the United States.

SOURCE: See Appendix B.

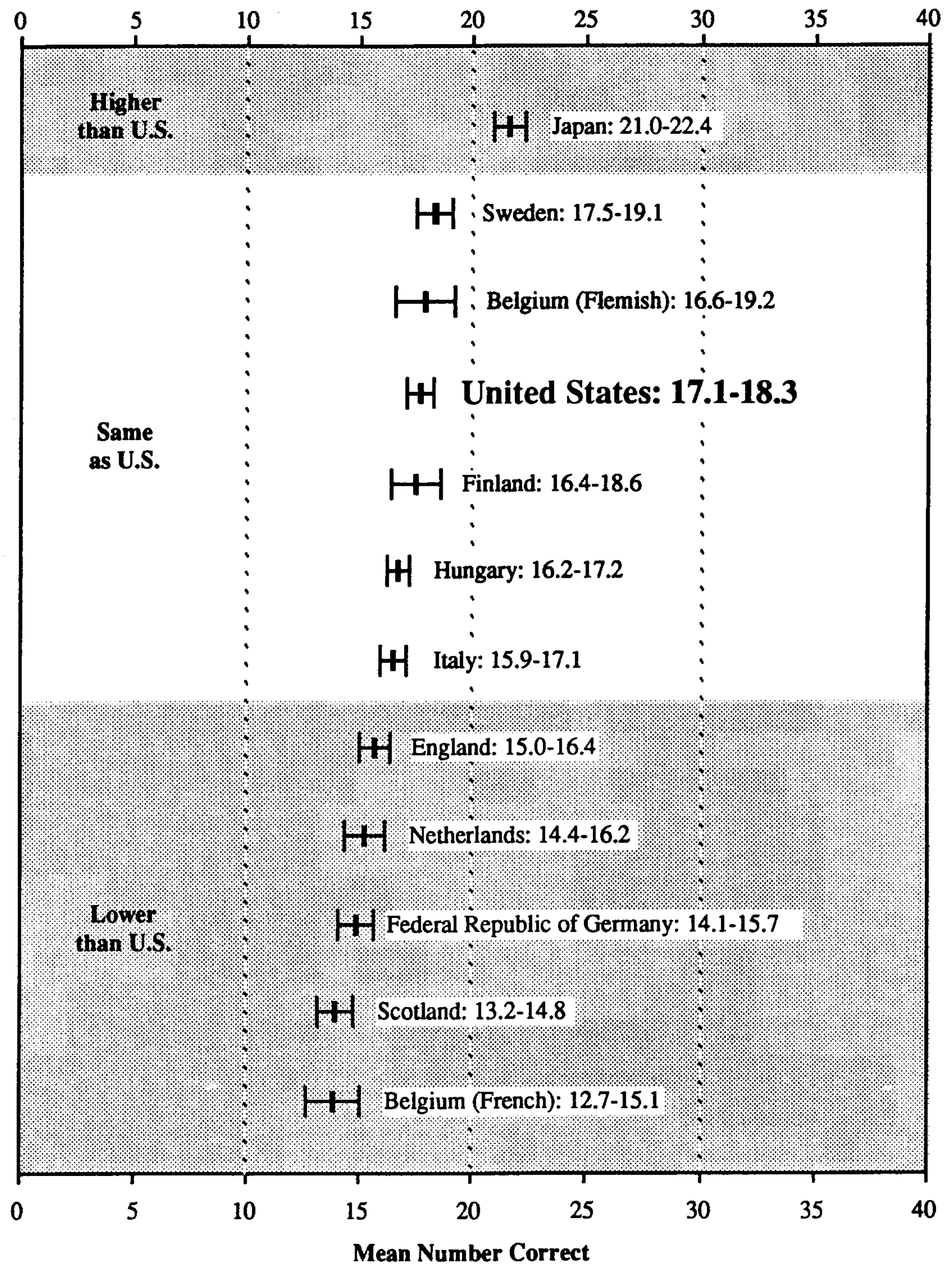
Figure E.12
Mean scores and confidence intervals for participating educational systems: Second International Mathematics Study, Last year of secondary school (46 items—elementary functions and calculus)



NOTE: Mean scores are denoted by the bold vertical ("|"). The simple 95 percent confidence interval for each mean is denoted by ("|-|-|"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 14 comparisons with the United States.

SOURCE: See Appendix B.

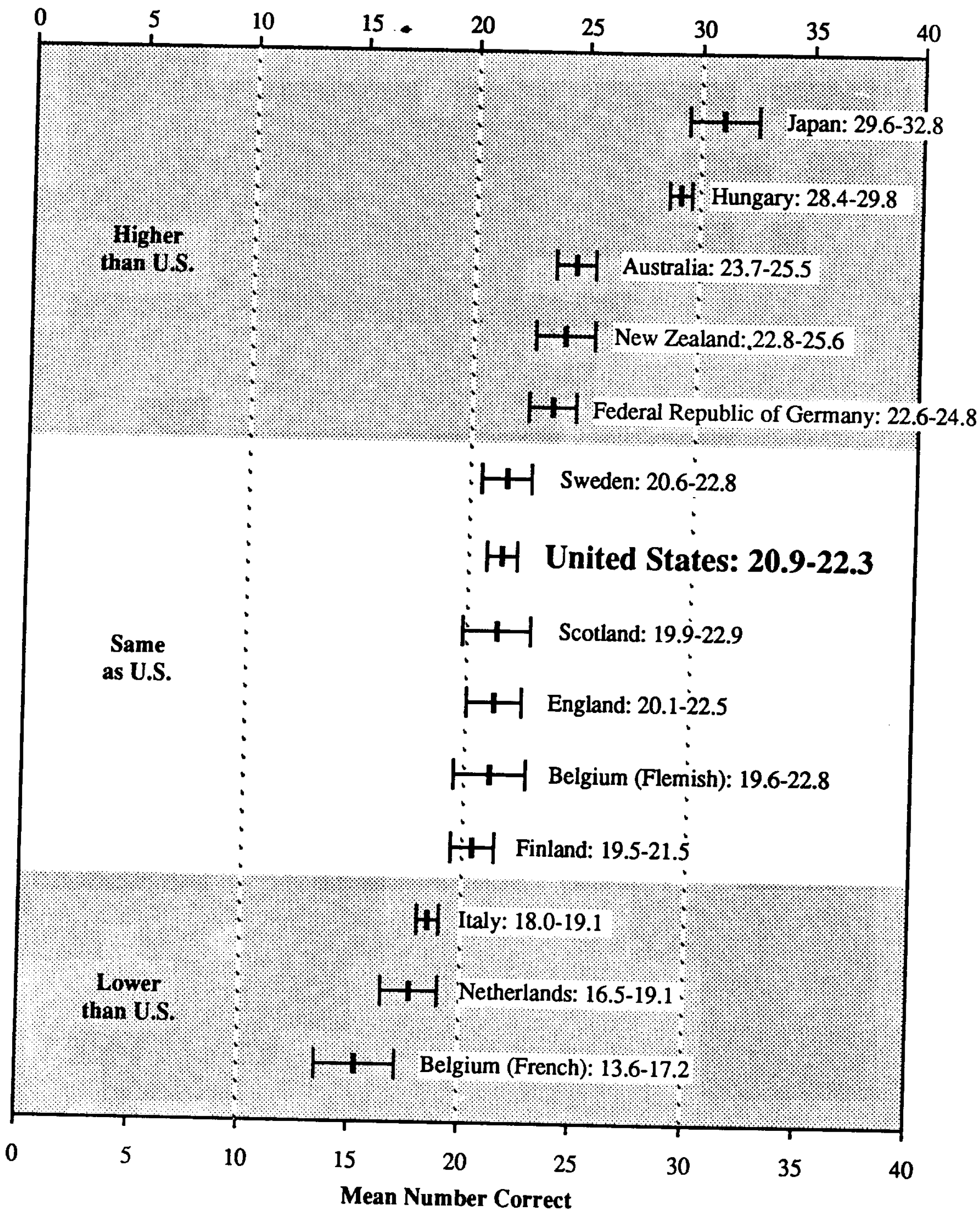
Figure E.13
Mean scores and confidence intervals for participating educational systems: First International Science Study, 10-year-olds (40 core items)



NOTE: Mean scores are denoted by the bold vertical ("|"). The simple 95 percent confidence interval for each mean is denoted by ("|-|-|"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 11 comparisons with the United States.

SOURCE: See Appendix B.

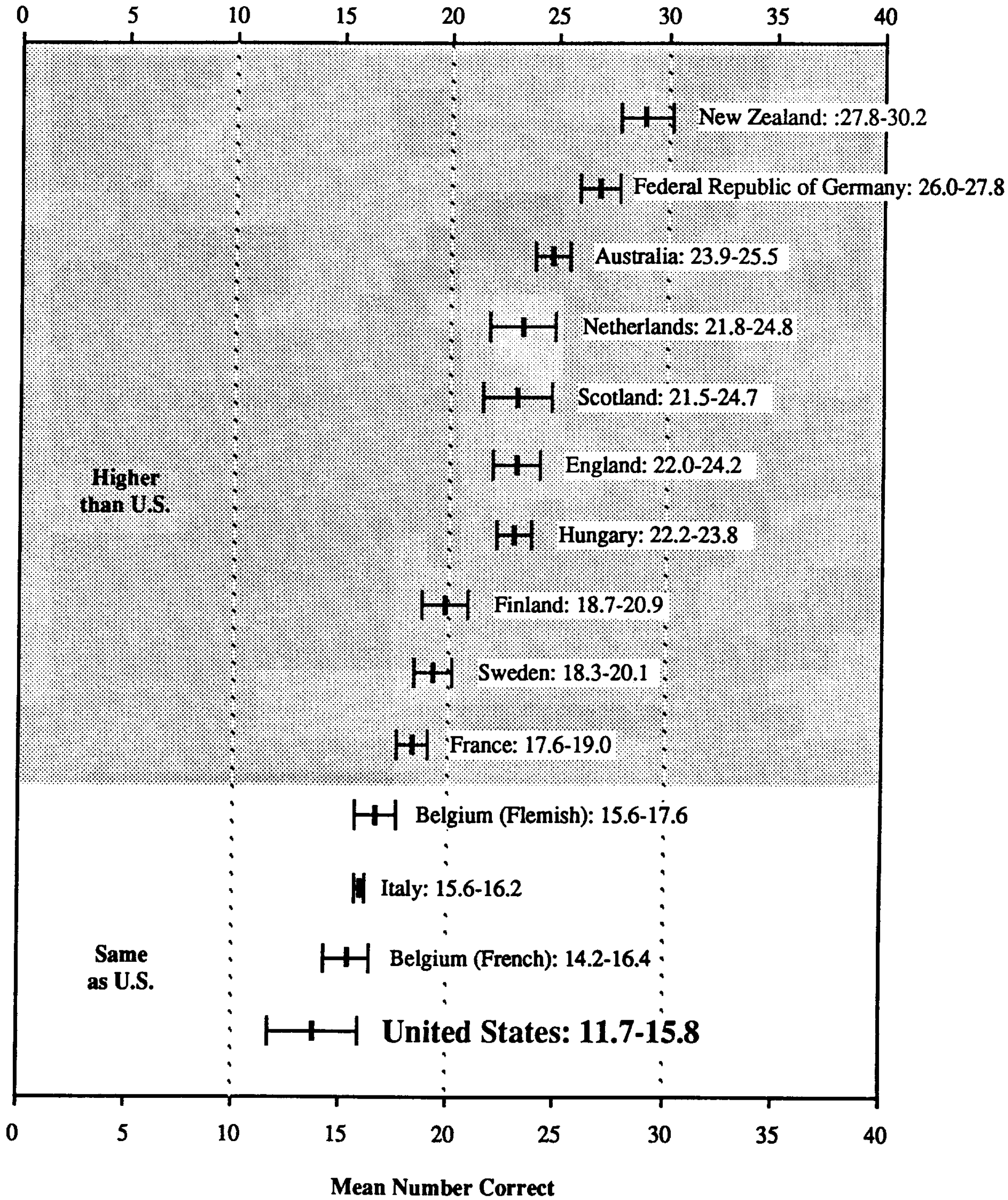
Figure E.14
Mean scores and confidence intervals for participating educational systems: First International Science Study, 14-year-olds (80 core items)



NOTE: Mean scores are denoted by the bold vertical ("❙"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 13 comparisons with the United States.

SOURCE: See Appendix B.

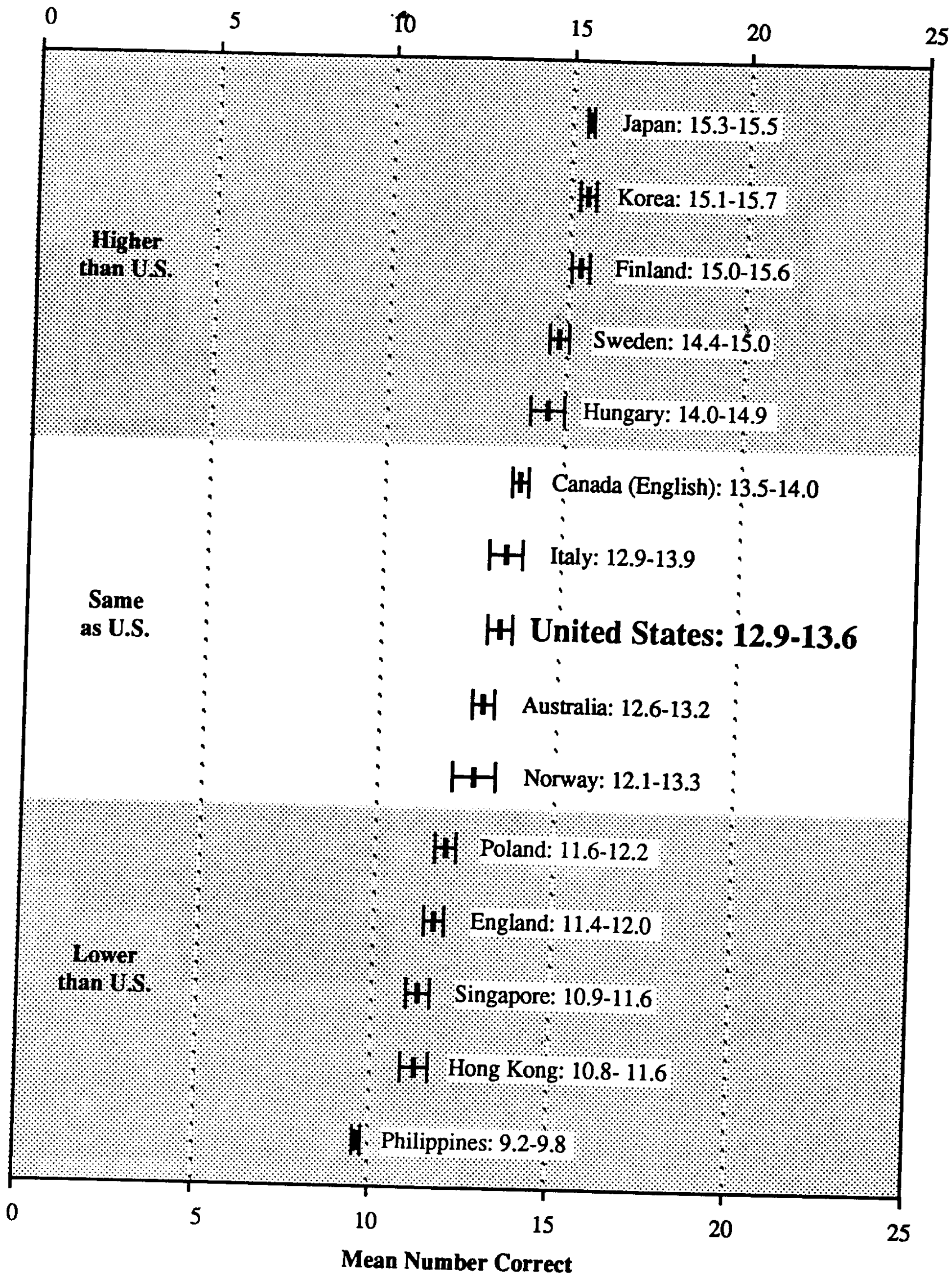
Figure E.15
Mean scores and confidence intervals for participating educational systems: First International Science Study, Last year of secondary school (60 core items)



NOTE: The mean score is denoted by the bold vertical ("**|**"). The simple confidence interval for the means is denoted by ("|-|-|"). Intervals are 95% confidence intervals. Statistical significance is based on comparison to the United States using Bonferroni adjusted t-test for 13 comparisons with the United States.

SOURCE: See Appendix B.

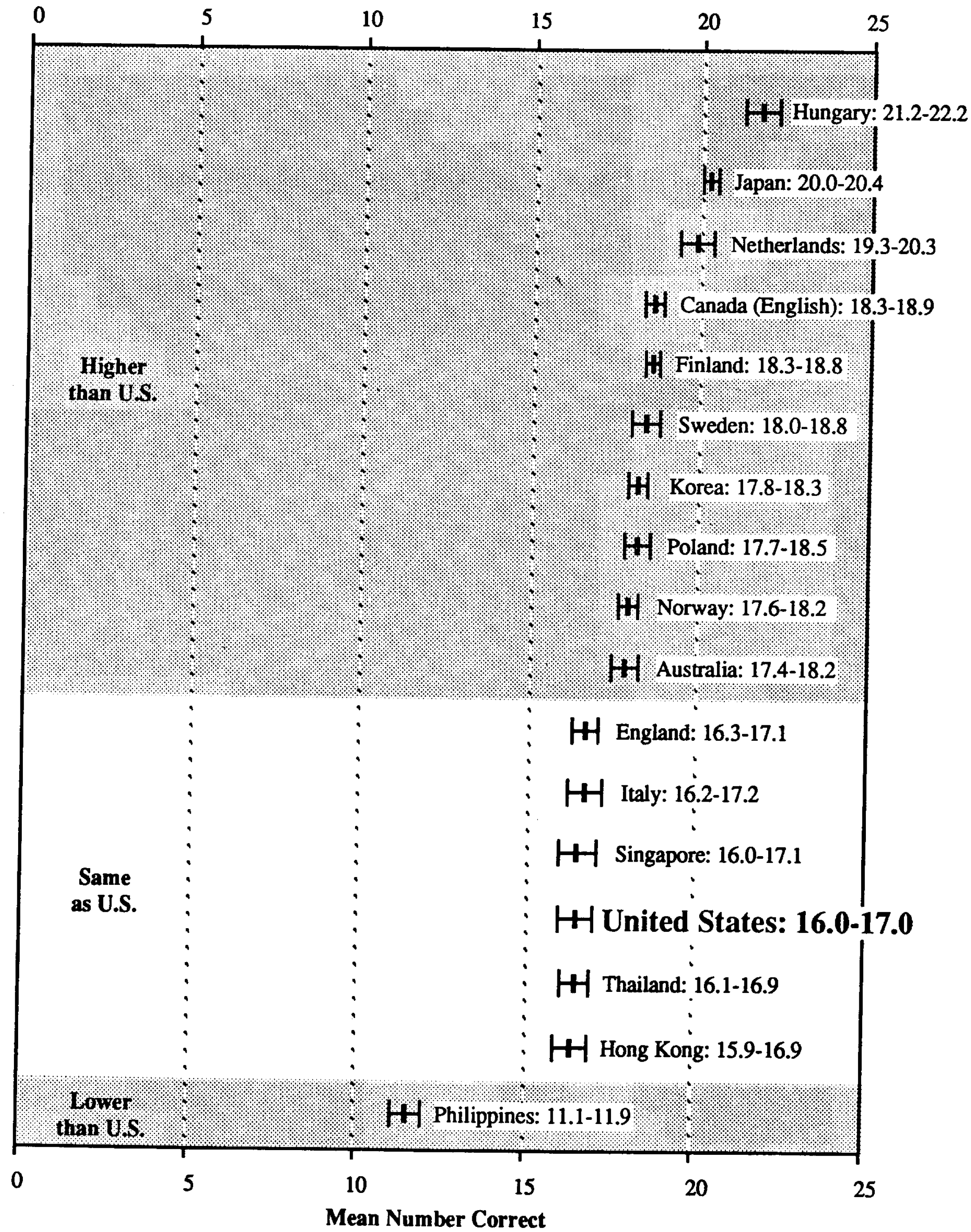
Figure E.16
Mean scores and confidence intervals for participating educational systems: Second International Science Study, 10-year-olds (24 core items)



NOTE: Mean scores are denoted by the bold vertical ("❚"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 14 comparisons with the United States.

SOURCE: See Appendix B.

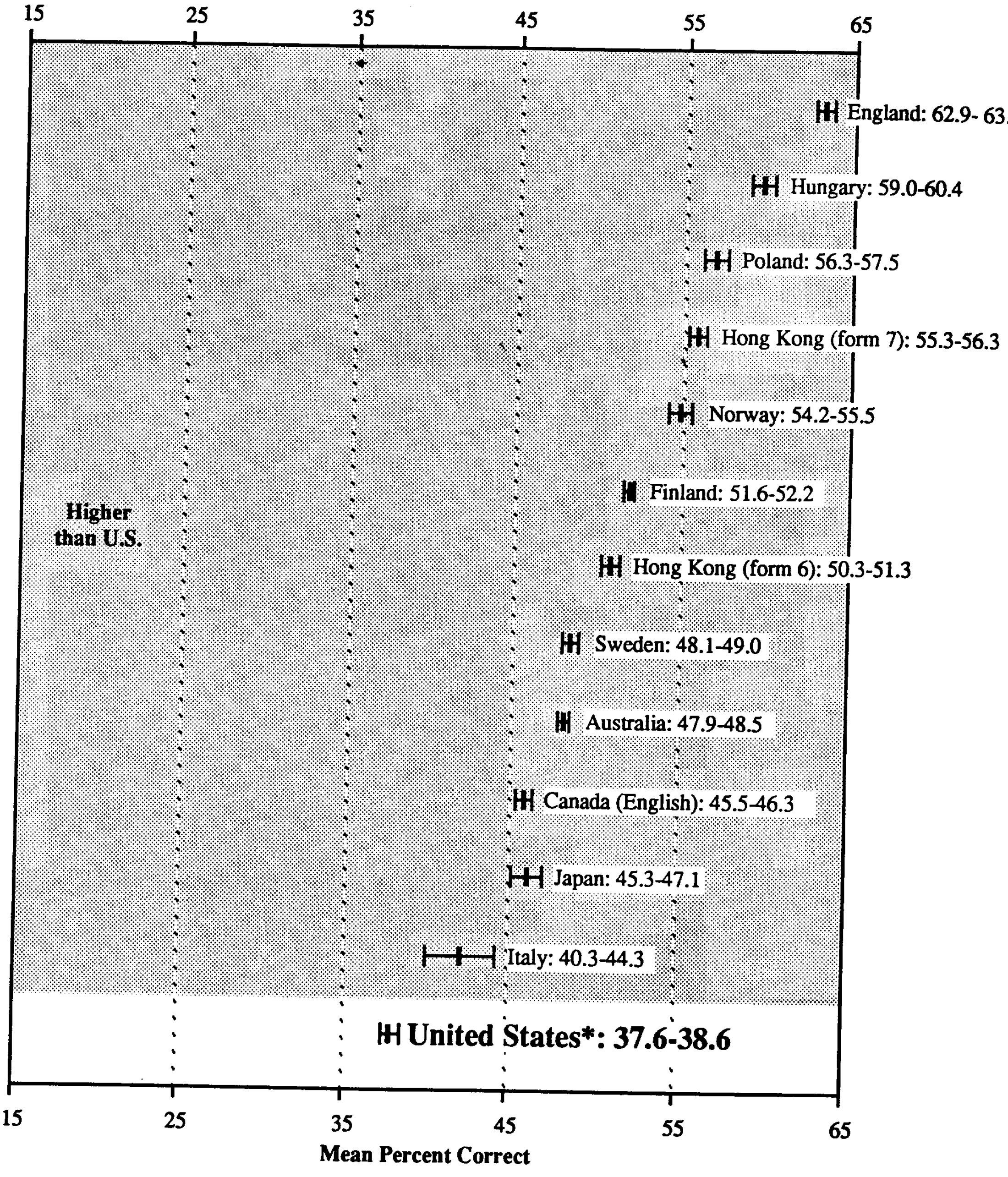
Figure E.17
Mean scores and confidence intervals for participating educational systems: Second International Science Study, 14-year-olds (30 core items)



NOTE: Mean scores are denoted by the bold vertical ("|"). The simple 95 percent confidence interval for each mean is denoted by ("|-|-|"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 16 comparisons with the United States.

SOURCE: See Appendix B.

Figure E.18
Mean scores and confidence intervals for participating educational systems: Second International Science Study, Last year of secondary school (30 core items—biology)



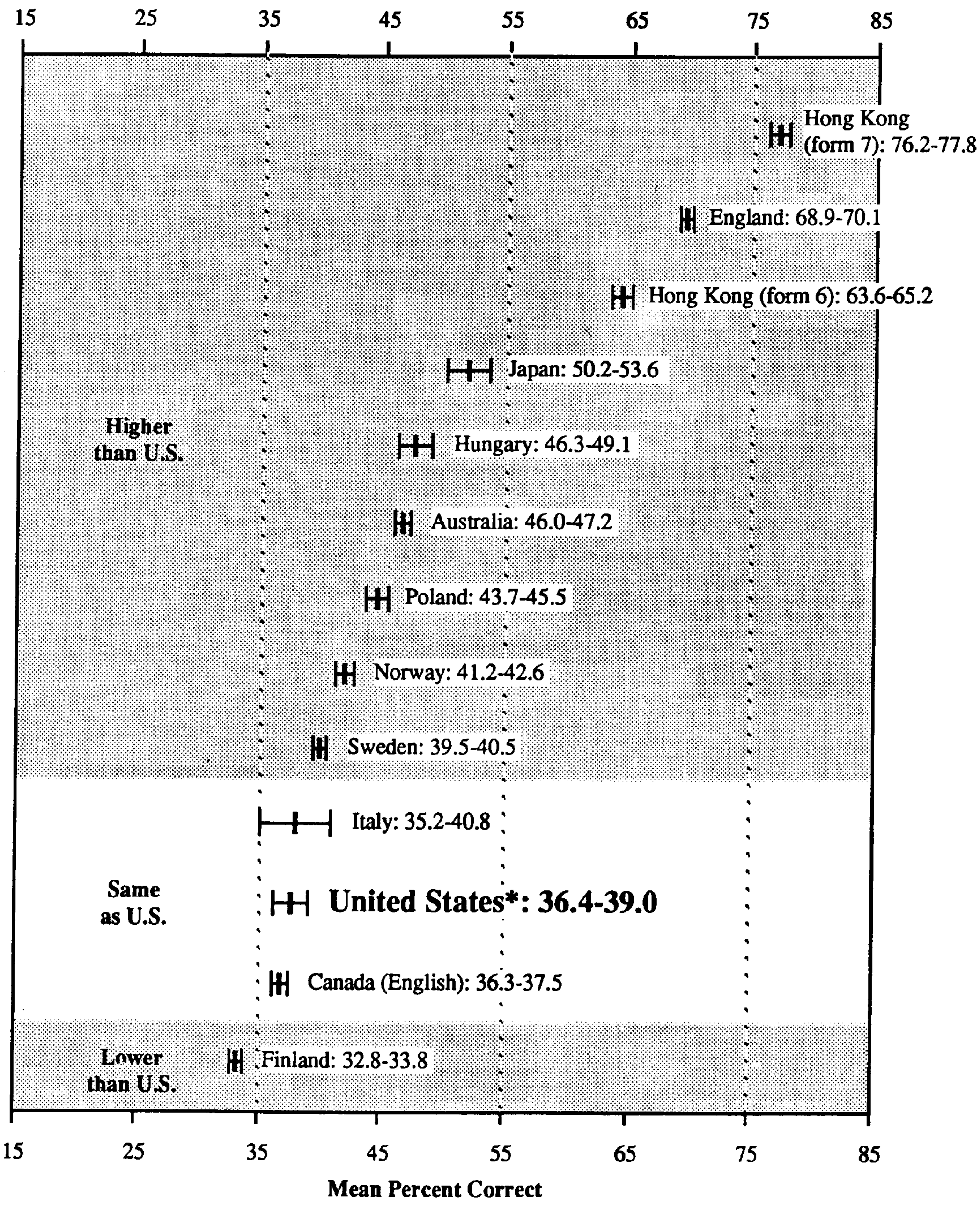
*United States test core 25 items; Australia 29 items; all other countries 30 items.

NOTE: On account of missing data confidence intervals could not be calculated for Singapore.

NOTE: Mean scores are denoted by the bold vertical ("❙"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 12 comparisons with the United States.

SOURCE: See Appendix B.

Figure E.19
Mean scores and confidence intervals for participating educational systems: Second International Science Study, Last year of secondary school (30 core items—chemistry)



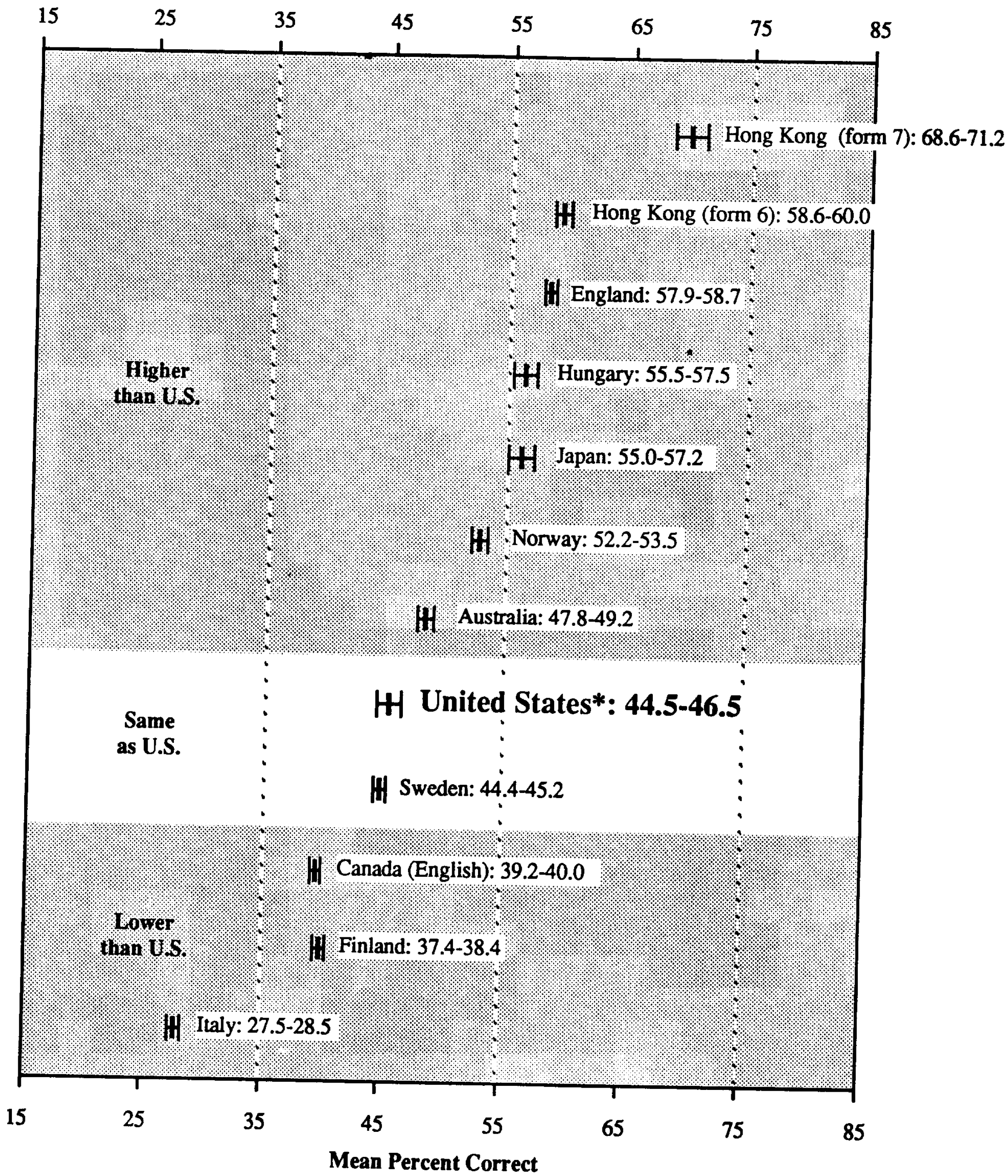
*United States test core 25 items; all other countries 30 items.

NOTE: On account of missing data confidence intervals could not be calculated for Singapore.

NOTE: Mean scores are denoted by the bold vertical ("|"). The simple 95 percent confidence interval for each mean is denoted by ("|-|-|"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 12 comparisons with the United States.

SOURCE: See Appendix B.

Figure E.20
Mean scores and confidence intervals for participating educational systems: Second International Science Study, Last year of secondary school (30 core items—physics)



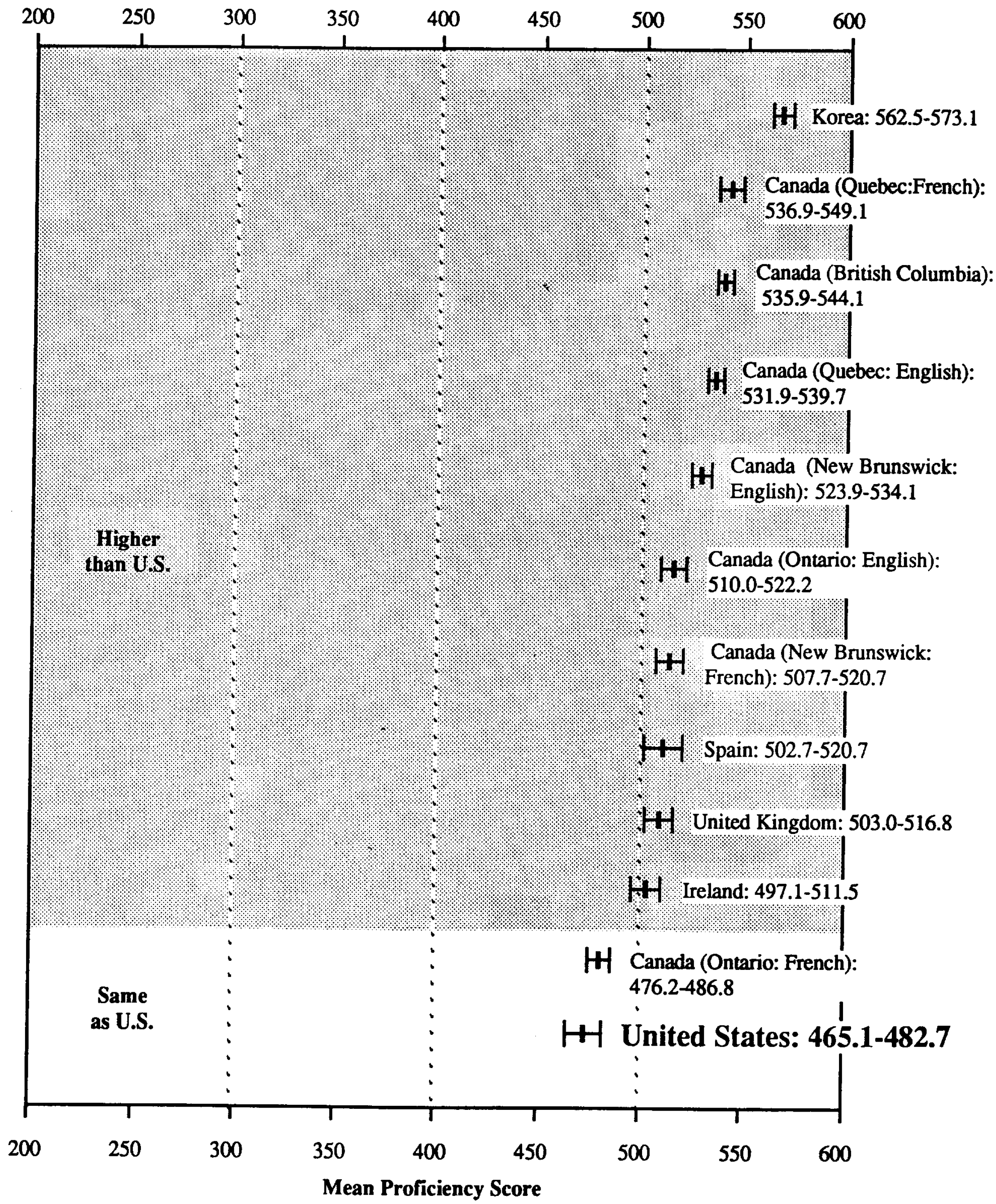
*United States test core 25 items; Canada (English) 26 items; all other countries 30 items.

NOTE: On account of missing data confidence intervals could not be calculated for Singapore and Poland.

NOTE: Mean scores are denoted by the bold vertical ("|"). The simple 95 percent confidence interval for each mean is denoted by ("|-|"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 12 comparisons with the United States.

SOURCE: See Appendix B.

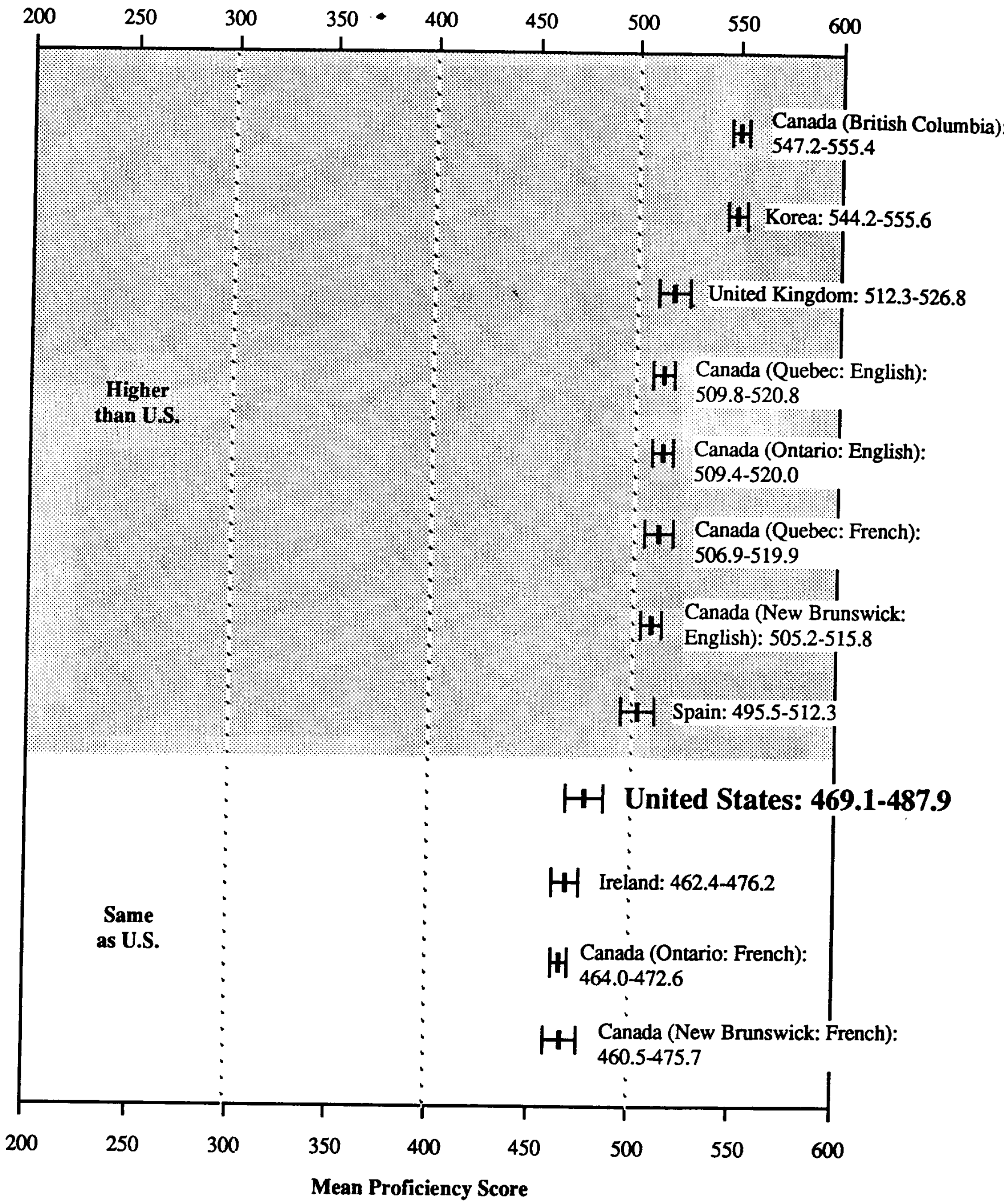
Figure E.21
Mean scores and confidence intervals for participating educational systems: International Assessment of Educational Progress, 13-year-olds (63 items—mathematics proficiency)



NOTE: Mean scores are denoted by the bold vertical ("❙"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 11 comparisons with the United States.

SOURCE: See Appendix B.

Figure E.22
Mean scores and confidence intervals for participating educational systems: International Assessment of Educational Progress, 13-year-olds (60 items—science proficiency)



NOTE: Mean scores are denoted by the bold vertical ("❙"). The simple 95 percent confidence interval for each mean is denoted by ("├┼┤"). Statistical significance of comparisons to the United States is based on a Bonferroni-adjusted t-test for 11 comparisons with the United States.

SOURCE: See Appendix B.

*U.S. Government Printing Office: 1992 — 311-087/62979