DOCUMENT RESUME

ED 423 262 TM 029 060

| | |
|---|---|
| AUTHOR | Stevenson, Harold W. |
| TITLE | A TIMSS Primer. Lessons and Implications for U.S. Education. |
| INSTITUTION | Thomas B. Fordham Foundation, Washington, DC. |
| PUB DATE | 1998-07-00 |
| NOTE | 40p.; Foreword by Chester E. Finn, Jr. |
| AVAILABLE FROM | Thomas B. Fordham Foundation, 1015 18th Street, N.W., Suite 300, Washington, DC 20036. |
| PUB TYPE | Collected Works - Serials (022) -- Reports - Evaluative (142) |
| JOURNAL CIT | Fordham Report; v2 n7 Jul 1998 |
| EDRS PRICE | MF01/PC02 Plus Postage. |
| DESCRIPTORS | *Academic Achievement; Achievement Tests; Case Studies; Curriculum; Elementary Secondary Education; Foreign Countries; International Education; *International Studies; Low Achievement; *Mathematics Achievement; Mathematics Education; *Performance Factors; *Science Education; Standards; Surveys; Tables (Data); Teacher Role; Teachers |
| IDENTIFIERS | *Science Achievement; *Third International Mathematics and Science Study |

ABSTRACT

Results are now available from the Third International Mathematics and Science Study (TIMSS) with its 5 main components, 41 cooperating countries, over 500,000 participants, and coverage of the full spectrum of mathematics and science from grades 4 through 12. American educators, parents, and policy makers have found the results to be both startling and disturbing, especially because of the decline in relative standing of U.S. students as they progress from elementary school through high school. This report describes how the TIMSS was conducted and discusses some lessons learned about the bases of these differences. The TIMSS included five main components: (1) curriculum analyses; (2) achievement tests; (3) questionnaire surveys of students, teachers, and administrators; (4) case studies of subjects in the United States, Germany, and Japan; the working environment and training of teachers; methods for dealing with differences in ability; and the role of school in adolescents' lives; and (5) a video study of classroom lessons in the United States, Germany, and Japan. The reports by members of the TIMSS staff express extreme caution in coming to firm answers concerning the poor performance of U.S. students. Nevertheless, it is possible to make some comments about American students' performance. Possible explanations begin with the fragmented, and nonsequential curricula in the United States, and the school's emphasis on developing rules that are automatically applied to problems rather than understanding the basis for the rules. Other problems are the lack of clear and tough standards, the mind-set that academic success is mostly determined by family background rather than by hard work, the demands placed on teachers, and their relatively low status within American culture. Demographic factors play a role, as does the associated phenomenon of placing some students in less challenging curricula. (Contains 6 tables, 4 figures, and 12 references.) (SLD)

THOMAS B. FORDHAM FOUNDATION
OUTSIDE THE BOX

# A TIMSS Primer





**by Harold W. Stevenson**

**JULY 1998**

*Fordham Report*

Vol. 2, No. 7 July 1998

# A TIMSS Primer

*Lessons and Implications for U.S. Education*

by
Harold W. Stevenson

THOMAS B. FORDHAM FOUNDATION
OUTSIDE THE BOX

# Table of Contents



4

# *Tables and Figures*

# *Foreword*

The results from TIMSS, the Third International Mathematics and Science Study, have bombarded America over the past several years. Some of the news was cheerful: our fourth graders scored among the best in the world in math and science. But our eighth grade scores were mediocre, and our twelfth grade scores were downright miserable. The longer our kids remain in school, it seemed, the worse they do, at least in math and science, at least in relation to most of the rest of the planet.

Then came the backlash. From education "experts" and pundits came word that we need not be upset by the TIMSS results. One main line of attack tried to invalidate the tests and comparisons on "technical" grounds. There must be something wrong with the tests or how they were administered or how their results were analyzed.

The other major critique—actually more like a dose of Prozac—said the country is doing fine so how could the schools have a problem? "Low Scores are No Disgrace" soothed one. "Stupid Students, Smart Economy?" asked another.

Perplexed? We sought clarification from the U.S. scholar who knows the most about international comparisons of K-12 education, Professor Harold Stevenson of the University of Michigan. Please tell us about TIMSS, we pleaded.

And he has responded with his usual mastery and clarity. In the pages that follow, you will read in plain English a definitive description of TIMSS and what its findings mean for the United States.

Stevenson begins by overviewing the study, detailing how participants were selected, and countering critics' allegations about bogus methods. He then describes its various components, from case studies and video catalogues to achievement tests and questionnaires. Finally, he digs into the findings and explains them in the context of cultural and school system differences. The result is a user-friendly guide through the complicated world of TIMSS. Hopefully, it will whet your appetite for the many more interesting TIMSS findings yet to come.

Harold Stevenson is very likely already known to you. Professor of Psychology at the University of Michigan, he is without question the foremost U.S. authority on international comparisons of K-12 education, not just comparisons of results but also of attitudes and values, of education systems and practices, of parenting and teaching. Of his many writings in this field, perhaps the best known is *The Learning Gap*, co–authored with James Stigler in 1992, which brilliantly explicates the differences between U.S. and Asian elementary schools.

As the director of the TIMSS case study project (which is described in this report), Dr. Stevenson developed its methodology and is responsible for producing the findings. He is author, with Roberta Nerison-Low, of the forthcoming comparative study, *It All Adds Up*. Readers wishing to contact him directly may write him at the Center for Human Growth & Development, The University of Michigan, 300 North Ingalls, Room 1000SW, Ann Arbor, MI 48109-0406 or e-mail hstevens@umich.edu.

The Thomas B. Fordham Foundation is a private foundation that supports research, publications, and action projects in elementary/secondary education reform at the national level and in the Dayton area. Further information can be obtained from our web site (http://www.edexcellence.net) or by writing us at 1015 18th Street, N.W., Suite 300, Washington, D.C. 20036. (We can also be e-mailed through our web site.) This report is available in full on the Foundation's web site and hard copies can be obtained by calling 1-888-TBF-7474 (single copies are free).

Chester E. Finn, Jr., President
Thomas B. Fordham Foundation
Washington, D.C.
July 1998

# Executive Summary

Results are now available from the Third International Mathematics and Science Study (TIMSS) with its five main components, 41 cooperating countries, over 500,000 participants, and coverage of the full spectrum of mathematics and science from fourth to twelfth grade. There has been widespread interest, both in the rankings of the participating countries and in possible explanations of the widely disparate levels of performance. American educators, parents, policy makers, and others interested in education have found the results to be both startling and disturbing, especially because of the decline in the relative standing of the U.S. students as they progressed from elementary school through high school.

***Why should U.S. students receive such low scores? What kinds of schooling lead to such marked differences between the performance of U.S. students and students in East Asian and some European countries?***

Why should U.S. students receive such low scores? What kinds of schooling lead to such marked differences between the performance of U.S. students and students in East Asian and some European countries? The purpose of this report is to attempt to answer these questions by describing how TIMSS was conducted and by discussing some of the lessons that have been learned about the bases of these differences.

## The Study Itself

TIMSS included five main components:

1) ***Curriculum Analyses***. Investigations of the academic standards of the various nations and the actual classroom curricula.
2) ***Achievement Tests***. Examinations that included multiple-choice and open-ended questions.

3) ***Questionnaires***. Surveys of students, teachers, and administrators regarding background characteristics, study habits, professional training, school culture, etc.
4) ***Case Studies***. In-depth analyses of four subjects in the U.S., Germany, and Japan: the implementation of national standards, the working environment and training of teachers, methods for dealing with differences in ability, and the role of school in adolescents' lives.
5) ***Video Study***. Recording and analysis of classroom lessons in the U.S., Germany, and Japan .

# Implications of the Study

The reports published by the TIMSS staff hesitate to draw any firm conclusions from the study. However, analysis of its five components, and especially the case studies and video study, leads to a few possible explanations for poor U.S. performance:

- U.S. schools' fragmented, non-sequential curricula.
- The emphasis on developing rules that are automatically applied to problems rather than an understanding of the basis for such rules.
- The lack of clear, tough academic standards.
- The mind-set that academic success is mostly determined by family background rather than by hard work.
- Overwhelming demands placed on teachers without adequate professional development or time.
- The low status awarded teachers within our culture.
- Demographic factors, such as inequitable school funding, and the associated phenomenon of tracking some students into less challenging curricula.

# *Overview of the TIMSS Study*

TIMSS, like its two predecessors of the 1960s and 1980s, was sponsored by the International Association for the Evaluation of Educational Achievement (IEA). It is by far the most ambitious and complex of the three studies and is, according to its organizers, "the largest, most comprehensive, and most rigorous international comparison of education ever undertaken."

My purpose in the following pages is to describe what was done in this ambitious study, what was found, and the implications it holds for American education.

## Supporters and Critics

Any effort with strong implications for sensitive topics such as education and social policy is bound to have ardent supporters and passionate critics. TIMSS has been no exception. For example, Alexandra Beatty, in her introduction to a report on a symposium at the National Research Council, wrote: "TIMSS has yielded an unprecedented body of data with which to explore both targeted questions about mathematics and science achievement and large questions about the structure and curricular goals of education systems in different nations." Supporters have pointed to the innovative methods employed in the study and the care that went into all aspects of its preparation.

***TIMSS is "the largest, most comprehensive, and most rigorous international comparison of education ever undertaken."***

No one suggests that the study is without fault. Not all of the nations were able to follow the recommendations for selecting participants. Similarly, it was not possible to ensure that all of the problems in the tests nor all of the items in the questionnaires were equally relevant to all participants. Even so, vigorous efforts were made to obtain the approval of representatives from the various countries before items were included. Despite its problems, the study has been widely commended for the depth and scope of its findings.

Critics have been less interested in offering specific criticisms than in rejecting the study as a meaningful contribution to education policy. Howard Gardner of the Harvard Graduate School of Education, for example, simply dismissed the measures of academic achievement on which the study was based: "These tests," he wrote in the *New York Times*, "don't measure whether students can think scientifically or mathematically, they just measure a kind of lowest common denominator of facts and skills. So getting students to do well on them doesn't mean much in the real world." Gerald Bracey, a perennial critic of all studies involving comparisons between the United States and other countries, was more harsh:

"[Pascal] Forgione (U.S. Commissioner of Education Statistics) has called TIMSS 'rough around the edges.' I say rotten to the core. The official TIMSS story is an exercise in political rhetoric and comes very close to being a hoax perpetrated on the whole world." Bracey then attempts to bolster his views by discussing data such as the ages of students in Cyprus and the years of physics studied by students in Norway, but fails to provide any meaningful discussion of why older American students or American students studying physics fall short in their performance.

While acknowledging that there are both those who praise TIMSS and those who regard it as an unfair indictment of American education, it is not productive to engage in further discussion of these controversial views. The vast majority of those who have read parts of the study consider it to be the best study that could have been done, both methodologically and substantively.

Even though numerous reports have been written, there is no single source of information about all of the components of TIMSS. A general overview should be of help in interpreting the various findings. Other than indicating the constraints that some of the methodological factors may pose for unambiguous interpretations of the results, further discussion will be devoted to an overview of the information that was available in the late spring of 1998.

## Designing the Study

A moment's reflection quickly suggests the enormity of the task involved in carrying out a study of the magnitude of TIMSS. Developing cross-culturally relevant and interesting items for the tests and questionnaires, selecting schools and gaining the cooperation of school authorities, analyzing the results from thousands of participants, and writing comprehensive reports of the results are extremely demanding challenges. The demands were made even greater in TIMSS by the fact that countries participated by choice. Because each country was required to pay for the collection of its own data, there was no basis for requiring participation in all components of the study by those nations that were willing to cooperate in any particular component. As a result, some countries, such as China, chose not to participate because of the expense. Others, such as Germany and Singapore, chose only to participate in parts of the study.

## Selecting the Participants: By Age or Grade?

One of the first tasks in organizing any research study is to decide who will participate. The organizers of TIMSS immediately faced the question of whether the participants within each country would be chosen on the basis of their chronological age or their number of years in school. Because countries have different requirements for the age at which children enter elementary school and for the time they graduate or leave secondary school, problems emerge with the adoption of either index. In the end, the decision was made to include three groups of students: those who were midway through elementary school, midway through secondary school, and at the end of upper-secondary school. Specifically, this included the grades containing the most nine-year-olds (termed Population 1), the most thirteen-year-olds (Population 2), and, regardless of age, those who were completing their secondary education (Population 3). This meant that Population 1 included both third and fourth graders in some countries and second and third graders in others, depending on which grades contained the greatest percentages of nine-year-olds. For Population 2, the participants could be in grades 7 and 8 or in grades 6 and 7. Most of the students in Population 3 were in grade 12, but it was possible in some countries for students enrolled in grades 9 to 13 to be included in Population 3, depending on the grade after which students left or graduated from high school.

***It was inevitable that the ages of students in each group differed among the various countries.***

It was inevitable, therefore, that the ages of students in each group differed among the various countries. To the degree that acquiring information about mathematics and science is believed to be dependent on chronological age, this could be considered an important drawback. If, however, the number of years of schooling is considered to be a more important index of academic knowledge, such differences in age can be assumed to have little impact on the results of achievement tests.

Explicit criteria for participating in TIMSS were developed in order to ensure that the samples of participants would be representative of each nation involved in the study. These criteria were an acceptance rate of 50 percent following the initial invitation, a participation rate of at least 75 percent after the recruitment of replacements, and the inclusion of samples representing at least 90 percent of the nation's eligible population. In addition, the participating classrooms within a school were to be selected randomly and participants were expected to be enrolled in the appropriate grades. It was not always possible, however, to obtain the cooperation of the schools selected or to meet all of the other sampling criteria.

Data for the eighth graders illustrate the types of problems that were encountered. Two nations faced such severe sampling problems that their data were withdrawn. There were so many questions about the data from sixteen other nations that the TIMSS International Study Center questioned the degree to which their results could be accepted with confidence. Readers of the various TIMSS reports, it should be noted, are informed in the accompanying tables about which nations experienced sampling problems.

# The Components of TIMSS

A brief description of the five main components of TIMSS illustrates the diversity of methods and topics that were included. These components included: analyses of the mathematics and science curricula of the participating countries; tests of mathematics and science knowledge; questionnaires for teachers, schools, and students; case studies of three participating countries; and a video study of classroom teaching in those three countries. Students in all participating countries were given tests of science and mathematics and questionnaires were completed by teachers, students, and school administrators. The case studies and video studies were conducted in only three countries: Germany, Japan, and the United States.

## Curriculum Analyses

Information about what is taught in the various nations is of interest in its own right for two reasons: first, without such information it was impossible to assess what students at the different grade levels in various countries are expected to know; second, this type of background information is necessary when attempting to construct culture-free tests. Armed with information about the content of the curriculum of their own countries, representatives from the participating countries were able to evaluate the relevance of the items for their country's students.

## Achievement Tests

A standard component of international comparative studies of student achievement is paper-and-pencil tests. In the case of TIMSS, these tests included multiple-choice and open-ended items that were administered to all participants during regular class periods. Each student was given a subset of questions from a larger pool; for example, the items for the various versions of the eighth grade test were selected from a pool of 102 mathematics items and 97 science items. In addition, subsets of randomly selected groups of students were given performance tests that required involvement with hands-on problems. Two additional types of special groups were formed for Population 3 by selecting students enrolled in advanced classes in mathematics and/or physics.

An international panel of subject-matter and assessment experts met to select items for use in a pilot study for Populations 1 and 2. New items were written and other items were selected from the tests used in the second IEA studies of mathematics and science. Items were retained for use in TIMSS if they were judged to be appropriate for more than 70 percent of the countries. The four types of tests developed for Population 3 included general and advanced tests of mathematics and science.

The tests for each population purposely contained comprehensive coverage of the topics that are generally included in mathematics and science curricula. For example, the mathematics test for Population 2 included fractions and number sense; geometry; algebra; data representation, analysis and probability; measurement; and proportionality. Five topics were included in the Population 2 science test: earth science; life science; physics; chemistry; and the combined topics of environmental issues and the nature of science.

## Questionnaires

Questionnaires were developed to elicit contextual information that would be useful in interpreting the results from the achievement tests. Items for students in Populations 1 and 2 asked about out-of-school activities, family demographics, attitudes toward mathematics and science, home language, use of calculators and computers, and the child's practices concerning studying and homework.

Questionnaires for teachers covered a broad range of topics, including the teacher's views about teaching mathematics and science, their background and professional training, responsibilities limiting their teaching practices, current teaching assignments, ways of handling certain kinds of material, coverage of the various aspects of mathematics, attitudes about homework, and strategies for teaching and managing other classroom activities.

The third questionnaire was designed for school authorities. This questionnaire contained items dealing with school administration, such as: who was responsible for the content of courses, how teachers were assigned to classes, discipline policies, the mathematics and science courses that were offered, tracking practices, and graduation requirements.

## Innovations

Two innovations in TIMSS departed markedly from the components of prior IEA comparative studies. During the very early planning phases of TIMSS, the need for more contextual information was pointed out. In the two previous IEA studies there had been no one-on-one interaction with parents, teachers, or students and no one had visited classrooms to observe teaching practices, or studied the customs or cultures of the participating countries. Consequently, there was little basis in the earlier studies for interpreting why students of different countries obtained different scores. Following these discussions, the decision was made to include case studies of selected policy issues and video studies of classroom lessons in TIMSS.

## Case Studies

The Case Studies Project was designed to focus on four topics of special concern to U.S. policy makers and to investigate how these topics were handled in the U.S., Japan, and Germany. The topics included the implementation of national standards, the working environment and training of teachers, methods for dealing with differences in ability, and the role of school in adolescents' lives. Each topic was studied through interviews with a broad range of students, parents, teachers, and education specialists. Supplementing the personal interactions were classroom observations of mathematics and science lessons.

Each topic was studied in three regions in each of the three countries and, when appropriate, at the fourth, eighth, and twelfth grades. Cities and schools were selected in consultation with advisors from each of the countries. Two goals were kept in mind: first, to select cities in the three countries that were as comparable to each other as possible in terms of population, industries, and socioeconomic and cultural status; and second, to obtain representative samples of respondents in each location. Because of the large commitments of time that were required of the case study participants, overlap of schools with those in the equally time-consuming main TIMSS study was avoided.

Interviews and conversations were held for more than 1300 hours. All were tape recorded, translated into English (in the case of German and Japanese), and entered in a computer program with key words necessary for easy retrieval of information. Supplementing the interviews and conversations were over 250 hours of observations of mathematics and science lessons in the three countries.

## Video Study

The Videotape Study of German, Japanese, and U.S. classrooms was conducted to gather more in-depth information about the classroom context in which learning takes place, the techniques of teaching, and the responses of students. An hour of regular classroom instruction was videotaped in nationally representative samples of mathematics classrooms in Population 2 that had been included in the main TIMSS study.

Data for the video study consisted of videotapes of representative samples of mathematics lessons in Germany (100), Japan (50), and the United States (81). Building on previous observational studies, a system was developed for combining the observations into a database that resulted from translating all materials into English, digitizing them, and transferring them to a CD-ROM.

The use of videotapes stored on CD-ROM eliminates the need for written narrative descriptions of the content and conduct of a lesson and provides illustrative examples of the classroom behavior of the teachers and their students. An effort was made to place the videotaped lesson within the context of everyday practices by instructing teachers to present a typical lesson. They were given a brief questionnaire to record their reactions to what was videotaped.

The methodological as well as substantive components of TIMSS obviously represent important advances over the prior comparative studies sponsored by IEA. And even more information will be available when additional analyses are made of the TIMSS data and TIMSS-R, a partial repetition of TIMSS that has been announced for 1999. Thirty of the TIMSS nations have agreed to participate in the study. Videotaping will also be extended to include the Netherlands, the Czech Republic, Korea, and Singapore, as well as the United States and Japan.

# Test Results

Discussions of the TIMSS results can proceed most effectively by reviewing each of the five components of the study separately and then discussing the resulting conclusions. The best place to start is by describing the findings from the achievement tests, and the first question to be asked is whether the U.S. is graduating students who are competitive with their peers in other industrialized nations. Data for the U.S. are used as the focus for comparisons.

Before describing the resulting scores, it is useful to know how to interpret the data provided in the various TIMSS reports. In an effort to make generalizations across subjects or grades possible, the "raw" scores obtained by the students were transformed into a new distribution with the ideal of having 500 as the international mean and 100 as the standard deviation. For several reasons, this ideal was hard to realize. Essentially, however, the scores can be interpreted roughly as percentiles if one knows, for example, that 16 percent of the scores in this new distribution lie below one standard deviation below the mean, 50 percent lie below the mean, and 84 percent lie below one standard deviation above the mean.

## Population 3 Results

The outcome of primary and secondary schooling in the participating countries is evident in the scores of students in Population 3. The overall mean scores of the U.S. students in Population 3 were below the international average and departed further from the average in mathematics than in science.

In addition to information about the mean for each country, the tables reporting the achievement test scores also indicate the standard error. This statistic gives an indication of how representative the mean is as an estimate of the mean of the population from which the sample was obtained. A commonly used index is found by taking the mean plus and minus two times the standard error. This encompasses 95 percent of the likely values for the population mean. The smaller the value of the standard error, the more reliable is the sample mean for representing the average score of the whole population of students from that country leaving or graduating from high school. The standard error for the U.S. is among the smallest obtained for any nation. In marked contrast, for example, is the Czech Republic, whose standard error is among the largest.

The standard error is also useful in determining whether the means for two samples differ from each other to a degree that cannot be attributed to chance. The larger the standard error of a mean, the more difficult it is to obtain statistically significant differences from other means.

On the basis of statistical tests, the TIMSS analyses divide the countries into three groups: those that receive scores significantly above the average for the U.S., those that do not differ from the average for the U.S., and those that are significantly below the

U.S. average. These three categories provide a concise indication of a nation's status in relation to other nations. More detailed information is presented in Tables 1 and 2 in terms of the mean score for each country, the standard error, and whether or not the country met the selection criteria established for inclusion in the study.

The data concerning inclusion are somewhat worrisome, for only five of the countries met the necessary criteria for inclusion in the Population 3 sample. It should be noted, too, that the East Asian nations did not participate in Population 3. Had they been included, the number of nations exceeding the U. S in their mathematics and science scores is likely to have been even greater.

**Table 1**
National Average Mathematics Performance
Compared with the U.S.
Population 3 (Twelfth Grade)

| Nation | Mean | Standard Error |
|---|---|---|
| Average score significantly higher than U.S. | | |
| Netherlands | 560 | 4.7 |
| *Sweden | 552 | 4.3 |
| Denmark | 547 | 3.3 |
| *Switzerland | 540 | 5.8 |
| Iceland | 534 | 2.0 |
| Norway | 528 | 4.1 |
| France | 523 | 5.1 |
| *New Zealand | 522 | 4.5 |
| Australia | 522 | 9.3 |
| Canada | 519 | 2.8 |
| Austria | 518 | 5.3 |
| Slovenia | 512 | 8.3 |
| Germany | 495 | 5.9 |
| *Hungary | 483 | 3.2 |
| Average score not significantly different from U.S. | | |
| Italy | 476 | 5.5 |
| Russian Federation | 471 | 6.2 |
| Lithuania | 469 | 6.1 |
| *Czech Republic | 466 | 12.3 |
| United States | 461 | 3.2 |
| Average scores significantly lower than U.S. | | |
| Cyprus | 446 | 2.5 |
| South Africa | 356 | 8.3 |

* = Nation meeting international guidelines.

It will be recalled that special tests were given to students enrolled in advanced mathematics and physics classes. Their scores placed the U.S. participants near the bottom of the 16 countries that administered the physics and advanced mathematics tests. The U.S. average score in mathematics was the lowest obtained by the 16 nations participating in the testing. For the group enrolled in physics classes, the average score of the U.S. students was the lowest, except for that of Austria, obtained by any of the participating countries.

It is clear that the U.S. ends up in the untenable position of producing students who, by the time they are ready to leave secondary school, are below average in both mathematics and science. This conclusion holds whether the whole range of students is considered or only those who have taken advanced-level courses in physics or mathematics.

**Table 2**
National Average Science Performance
Compared with the U.S.
Population 3 (Twelfth Grade)

| Nation | Mean | Standard Error |
|---|---|---|
| Average score significantly higher than U.S. | | |
| *Sweden | 559 | 4.4 |
| Netherlands | 558 | 5.3 |
| Iceland | 549 | 1.5 |
| Norway | 544 | 4.1 |
| Canada | 532 | 2.6 |
| *New Zealand | 529 | 5.2 |
| Australia | 527 | 9.8 |
| *Switzerland | 523 | 5.3 |
| Austria | 520 | 5.6 |
| Slovenia | 517 | 8.2 |
| Denmark | 509 | 3.6 |
| Average score not significantly different from U.S. | | |
| Germany | 497 | 5.1 |
| France | 487 | 5.1 |
| *Czech Republic | 487 | 8.8 |
| Russian Federation | 481 | 5.7 |
| United States | 480 | 3.3 |
| Italy | 475 | 5.3 |
| *Hungary | 471 | 3.0 |
| Lithuania | 461 | 5.7 |
| Average scores significantly lower than U.S. | | |
| Cyprus | 448 | 3.0 |
| South Africa | 349 | 10.5 |

* = Nation meeting international guidelines.

This is a bleak conclusion, but one that should not come as a great surprise. It replicates findings that had been obtained in the first IEA study of previous decades. Looking at students in the second IEA study who were studying mathematics in their final year of secondary school, the mean score for the Japanese students was over twice the mean for the U.S. students (31.4 versus 13.8 points). For students not studying mathematics, the scores were lower, but the differences between the averages were of similar magnitude (25.3 versus 8.3 points). In the first IEA study of science, Japanese students received the highest scores at the elementary and middle school levels but did not participate at the high school level. The highest average for the high school students who did participate was obtained by those from New Zealand, whose scores were over twice those received by the U.S. students.

***The U.S. ends up in the untenable position of producing students who, by the time they are ready to leave secondary school, are below average in both mathematics and science.***

The second IEA study yielded similar conclusions. In none of the analyses were the scores of the U.S. students at or above the international average. In fact, U.S. students' scores were generally among those at the bottom fourth of the countries in the six tests given. Even those students who were enrolled in calculus classes, often considered the best mathematics students in the U.S., were at or near the average levels of achievement attained by their counterparts in the fourteen other participating countries.

In view of the consistency of the results over 30 years, questions must be raised about the usefulness of TIMSS-R planned for 1999. Are improvements expected in the few years since the data for TIMSS were collected?

## Population 2 Results

The major emphasis in TIMSS was on Population 2. This part of the study included the largest number of cooperating countries and the most participating students. Only eighth grade classrooms were included in the video study and more emphasis was placed on eighth graders in the case studies than on any other group.

Five nations outperformed the U.S. in both mathematics and science (see Tables 3 & 4). Three were from East Asia (Singapore, Korea, and Japan) and two were from Europe (Hungary and the Czech Republic). The Netherlands, Austria, Slovenia, and Bulgaria also received significantly higher scores than the U.S., but because of sampling problems their scores are subject to question. The only nations that the U.S. students outperformed in both mathematics and science were Cyprus, Iran, Lithuania, and Portugal—hardly the countries whose educational systems would seem to be competitive with those of the U.S.

In mathematics, the average scores of students in twenty nations were statistically higher than those obtained by the U.S. eighth graders.

**Table 3**
National Average Mathematics
Performance Compared with the U.S.
Population 2 (Eighth Grade)

| Nation | Mean | Standard Error |
|---|---|---|
| Average score significantly higher than U.S. | | |
| *Singapore | 643 | 4.9 |
| *Korea | 607 | 2.4 |
| *Japan | 605 | 1.9 |
| *Hong Kong | 588 | 6.5 |
| Belgium-Flemish | 565 | 5.7 |
| *Czech Republic | 564 | 4.9 |
| *Slovak Republic | 547 | 3.3 |
| Switzerland | 545 | 2.8 |
| Netherlands | 541 | 6.7 |
| Slovenia | 541 | 3.1 |
| Bulgaria | 540 | 6.3 |
| Austria | 539 | 3.0 |
| *France | 538 | 2.9 |
| *Hungary | 537 | 3.2 |
| *Russian Federation | 535 | 5.3 |
| Australia | 530 | 4.0 |
| *Ireland | 527 | 5.1 |
| *Canada | 527 | 2.4 |
| Belgium-French | 526 | 3.4 |
| *Sweden | 519 | 3.0 |
| Average score not significantly different from U.S. | | |
| Thailand | 522 | 5.7 |
| Israel | 522 | 6.2 |
| Germany | 509 | 4.5 |
| *New Zealand | 508 | 4.5 |
| England | 506 | 2.6 |
| *Norway | 503 | 2.2 |
| Denmark | 502 | 2.8 |
| United States | 500 | 4.6 |
| Scotland | 498 | 5.5 |
| Latvia | 493 | 3.1 |
| *Spain | 487 | 2.0 |
| *Iceland | 487 | 4.5 |
| Greece | 484 | 3.1 |
| Romania | 482 | 4.0 |
| Average scores significantly lower than U.S. | | |
| Lithuania | 477 | 3.5 |
| *Cyprus | 474 | 1.9 |
| *Portugal | 454 | 2.5 |
| *Iran, Islamic Republic | 428 | 2.2 |
| Kuwait | 392 | 2.5 |
| Colombia | 385 | 3.4 |
| South Africa | 354 | 4.4 |

* = Nation meeting international guidelines. Note: Among the 22 nations that failed to meet the criteria for inclusion, 16 nations' departures were so great as to call into question the reliability of their data. To check the influence of these cases, the international mean was re-calculated including only the 25 nations that met the sampling criteria.

**Table 4**
National Average Science Performance
Compared with the U.S.
Population 2 (Eighth Grade)

| Nation | Mean | Standard Error |
|---|---|---|
| Average score significantly higher than U.S. | | |
| *Singapore | 607 | 5.5 |
| *Czech Republic | 574 | 4.3 |
| *Japan | 571 | 1.6 |
| *Korea | 565 | 1.9 |
| Bulgaria | 565 | 5.3 |
| Netherlands | 560 | 5.0 |
| Slovenia | 560 | 2.5 |
| Austria | 558 | 3.7 |
| *Hungary | 554 | 2.8 |
| Average score not significantly different from U.S. | | |
| England | 552 | 3.3 |
| Belgium-Flemish | 550 | 4.2 |
| Australia | 545 | 3.9 |
| *Slovak Republic | 544 | 3.2 |
| *Russian Federation | 538 | 4.0 |
| *Ireland | 538 | 4.5 |
| *Sweden | 535 | 3.0 |
| United States | 534 | 4.7 |
| Germany | 531 | 4.8 |
| *Canada | 531 | 2.6 |
| Norway | 527 | 1.9 |
| *New Zealand | 525 | 4.4 |
| Thailand | 525 | 3.7 |
| Israel | 524 | 5.7 |
| *Hong Kong | 522 | 4.7 |
| Switzerland | 522 | 2.5 |
| Scotland | 517 | 5.1 |
| Average scores significantly lower than U.S. | | |
| *Spain | 517 | 1.7 |
| *France | 498 | 2.5 |
| Greece | 497 | 2.2 |
| *Iceland | 494 | 4.0 |
| Romania | 486 | 4.7 |
| Latvia | 485 | 2.7 |
| *Portugal | 480 | 2.3 |
| Denmark | 478 | 3.1 |
| Lithuania | 476 | 3.4 |
| Belgium-French | 471 | 2.8 |
| *Iran, Islamic Republic | 470 | 2.4 |
| *Cyprus | 463 | 1.9 |
| *Kuwait | 430 | 3.7 |
| Colombia | 411 | 4.1 |
| South Africa | 326 | 6.6 |

* = Nation meeting international guidelines.

Students from seven nations performed less effectively than the U.S. students. U.S. students fared somewhat better in science. They obtained significantly higher scores than their peers in fifteen nations and were outperformed by students in only nine.

To check the influence of the 16 nations that did not meet the criteria for sampling, the international mean was re-calculated including only the 25 nations that met the sampling criteria. The U.S. mathematics score was still below the international average, but the score for science was no longer significantly different from the average of the 25 nations.

Another question is how the U.S. would compare with other countries if only the top students from the various countries were considered. Perhaps U.S. strength lies in a small percentage of top students, rather than in the performance of its average students. If a group of students representing the top 10 percent of the students from all nations were assembled, what percentage of U.S. students would be included? The answer appears in Figure 1. Only 5 percent of the U.S. students would be chosen in mathematics and 13 percent in science. How about the top 50 percent? Would the relative contributions be maintained? As is evident in Figure 1, the U.S. makes a notably lower contribution to the top 50 percent than do Japan or Singapore, two top-scoring countries.

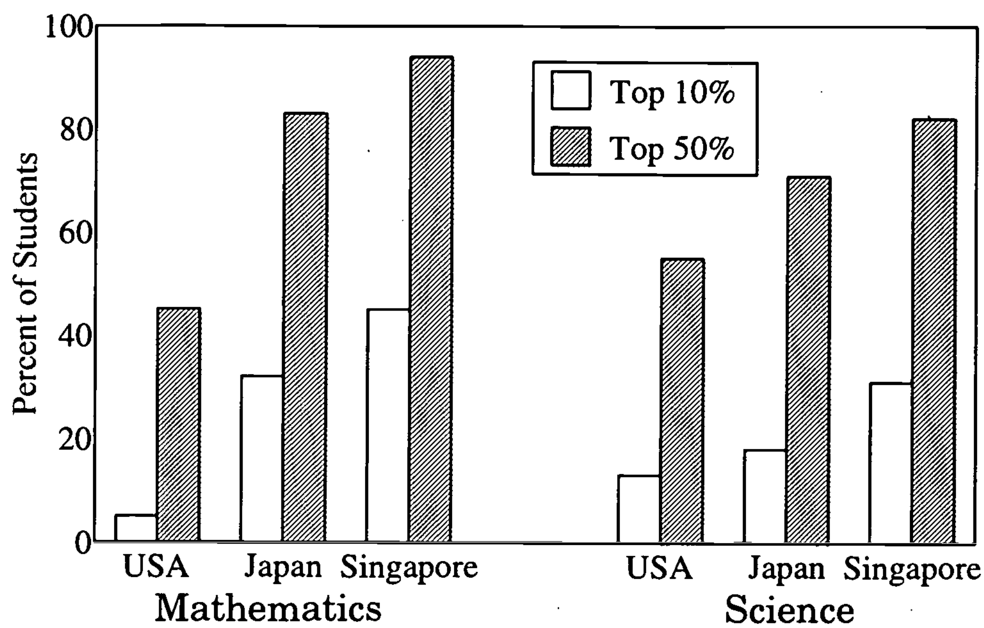


Figure 1. If the top 10% of 8th graders from all nations were assembled, what percentage of U.S. students would be included? What would the percentages be if the top 50% were considered?

An additional phase of the testing at middle school considered seventh and eighth graders. In the case of the U.S., 4,000 seventh graders and approximately 7,000 eighth graders took the TIMSS tests. A comparison of performance at these two grades yields an informative index of what is learned during the eighth grade. The smallest increments among the 25 participating countries in students' scores in mathematics from the seventh to the eighth grade were made by the U.S. and Belgium. The increments in the scores between seventh and eighth grade for science followed a similar pattern. The average increment for the U.S. was again one of the two smallest obtained by any of the countries.

***Once behind and with smaller increments of knowledge each year, it becomes increasingly difficult for U.S. students to catch up to their peers in other countries.***

If this trend is maintained at all grade levels, it is easy to see why U.S. students fall further behind their peers in other industrialized countries as their grade level increases. Once behind and with smaller increments of knowledge each year, it becomes increasingly difficult for them to catch up to their peers in other countries.

## Population 1 Results

The most effective performance by U.S. students occurred at the fourth grade (see Tables 5 & 6). U.S. fourth-graders scored above the international mean in both mathematics and science. Furthermore, the U.S. average in science was surpassed only by that of Korea and was higher than those of 19 of the 26 participating nations. Conclusions were unchanged when the international mean was computed only for nations that met all of the criteria for participation in TIMSS. The scores were not due to greater strength in one area of mathematics or science than in other areas. In mathematics, for example, U.S. students were above the average in five of the six areas included in the test. In science, U.S. students were above the international average in all four areas of science included in the test.

**Table 5**
National Average Mathematics
Performance Compared with the U.S.
Population 1 (Fourth Grade)

| Nation | Mean | Standard Error |
|---|---|---|
| Average score significantly higher than U.S. | | |
| *Singapore | 625 | 5.3 |
| *Korea | 611 | 2.1 |
| *Japan | 597 | 2.1 |
| *Hong Kong | 587 | 4.3 |
| Netherlands | 577 | 3.4 |
| *Czech Republic | 567 | 3.3 |
| Austria | 559 | 3.1 |
| Average score not significantly different from U.S. | | |
| Slovenia | 552 | 3.2 |
| *Ireland | 550 | 3.4 |
| Hungary | 548 | 3.7 |
| Australia | 546 | 3.1 |
| *United States | 545 | 3.0 |
| *Canada | 532 | 3.3 |
| Israel | 531 | 3.5 |
| Average scores significantly lower than U.S. | | |
| Latvia | 525 | 4.8 |
| Scotland | 520 | 3.9 |
| England | 513 | 3.2 |
| *Cyprus | 502 | 3.1 |
| *Norway | 502 | 3.0 |
| *New Zealand | 499 | 4.3 |
| *Greece | 492 | 4.4 |
| Thailand | 490 | 4.7 |
| *Portugal | 475 | 3.5 |
| *Iceland | 474 | 2.7 |
| *Iran, Islamic Republic | 429 | 4.0 |
| *Kuwait | 400 | 2.8 |

* = Nation meeting international guidelines.

**Table 6**
National Average Science Performance
Compared with the U.S.
Population 1 (Fourth Grade)

| Nation | Mean | Standard Error |
|---|---|---|
| Average score significantly higher than U.S. | | |
| *Korea | 597 | 1.9 |
| Average score not significantly different from U.S. | | |
| *Japan | 574 | 1.8 |
| *United States | 565 | 3.1 |
| Austria | 565 | 3.3 |
| Australia | 562 | 2.9 |
| Netherlands | 557 | 3.1 |
| *Czech Republic | 557 | 3.1 |
| Average scores significantly lower than U.S. | | |
| England | 551 | 3.3 |
| *Canada | 549 | 3.0 |
| *Singapore | 547 | 5.0 |
| Slovenia | 546 | 3.3 |
| *Ireland | 539 | 3.3 |
| Scotland | 536 | 4.2 |
| *Hong Kong | 533 | 3.7 |
| Hungary | 532 | 3.4 |
| *New Zealand | 531 | 4.9 |
| *Norway | 530 | 3.6 |
| Latvia | 512 | 4.9 |
| Israel | 505 | 3.6 |
| *Iceland | 505 | 3.3 |
| *Greece | 497 | 4.1 |
| *Portugal | 480 | 4.0 |
| *Cyprus | 475 | 3.3 |
| Thailand | 473 | 4.9 |
| *Iran, Islamic Republic | 416 | 3.9 |
| Kuwait | 401 | 3.1 |

*= Nation meeting international guidelines.

As was the case with Population 2, attention was also given to the top performers in each nation. Again, the percentage of fourth graders that would be included in the top ten percent of all fourth graders participating in TIMSS was determined. The results appear in Figure 2. The pattern evident in this figure is very similar to that appearing in Figure 1. There was little difference in the contribution of U.S. to the top 10 percent of students in science, but there was a marked difference in the contribution of each country to the top 10 percent of students in mathematics.

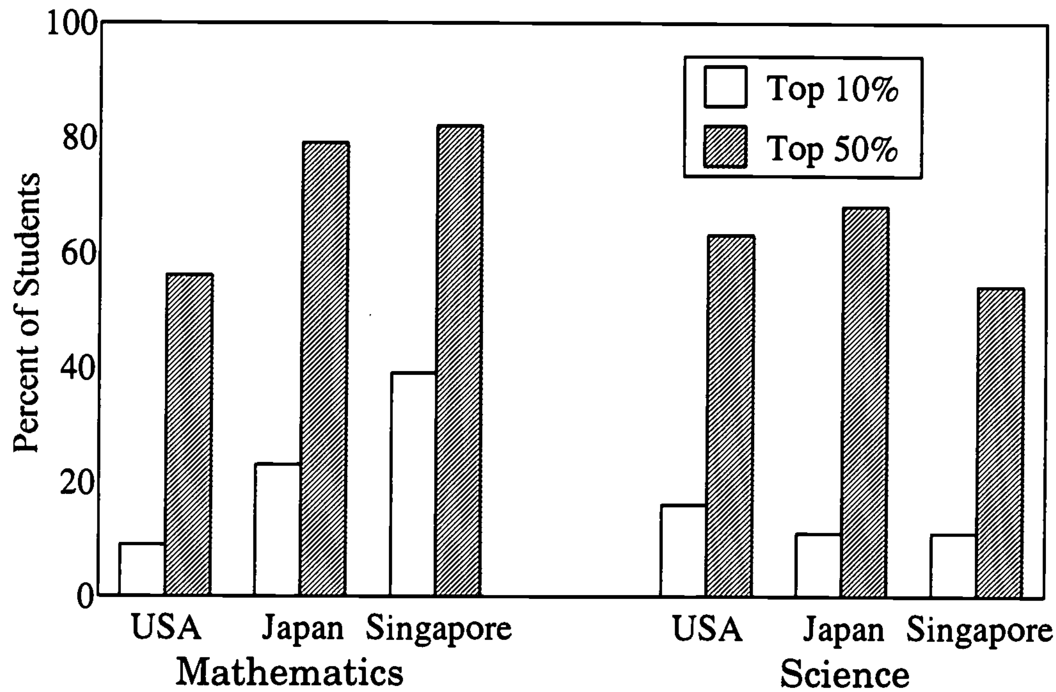


Figure 2. If the top 10% of 4th graders from all nations were assembled, what percentage of U.S. students would be included? What would the percentages be if the top 50% were considered?

There are no obvious explanations as to why U.S. fourth graders should appear to be so much stronger in mathematics and science than their older U.S. counterparts. Some writers have suggested that the fourth graders have benefited from improved teaching practices resulting from the adoption of the standards published by the National Council of Teachers of Mathematics. This is possible, but there is little concrete evidence showing a relation between adoption of the standards and students' performance. The fourth grade findings remain among the most tantalizing of the issues arising from the TIMSS data. Will these students maintain their high standing in successive years? Or will their performance be hindered by whatever factors resulted in the lower performance of U.S. students in Populations 2 and 3?

## Conclusions

Is it not time to accept the fact that U.S. students, except perhaps in the lower elementary grades, experience more serious difficulties in learning mathematics and science than do their peers in many other industrialized countries? Or, stated another way, is there any convincing evidence that students from typical American middle or high schools are as effective or more effective in mathematics and science than their peers in other industrialized countries? There might be reason to answer these questions less confidently if TIMSS were the only study that had been conducted. But this is not the case. A series of both large- and small-scale studies has yielded the same conclusion: U.S. schools are in need of attention and improvement. No one of these studies is perfect, but the accumulation of carefully conducted studies, all yielding similar conclusions and covering several decades, compels the reviewer to reach this conclusion.

***Large- and small-scale studies have yielded the same conclusion: U.S. schools are in need of attention and improvement.***

Arguing that schools are better now than they were a decade or two ago begs the central question: Are U.S. schools competitive with the schools found in other advanced industrialized nations, such as those of East Asia and Central Europe? TIMSS may make its most useful contribution to U.S. education by demonstrating the dramatic differences that exist among schools throughout the world in their ability to impart information and skills to their pupils. For the U.S., the major contribution is to point out that, despite a high financial investment in education, U.S. schools are clearly not among the world's most successful.

# The Context of Achievement

## Case Study Project

A primary goal of international comparative studies of achievement in mathematics and science is to evaluate the levels of achievement of students in various countries. An equally important goal is to attempt to understand and explain the bases of whatever differences emerge. The first impulse is to rely on questionnaires as a means of obtaining relevant information. Indeed, questionnaires are an obvious choice when it is necessary to collect large amounts of data on an array of topics at the least expense. How could information from all the participants in TIMSS have been obtained if the organizers had not relied on questionnaires? Case studies, which involve observations, long conversations, and interviews, are much more time-consuming, require more highly trained researchers, and are necessarily more expensive than questionnaire studies. Nevertheless, through the use of relaxed interactions and observations in everyday settings, case studies offer the possibility of gaining a depth of understanding that is difficult to reach with more impersonal questionnaires, especially when the studies involve different cultures and languages.

> ***Case studies offer a depth of understanding that is difficult to reach with more impersonal questionnaires, especially when the studies involve different cultures and languages.***

Because of the magnitude of TIMSS and the limited amount of time available to conduct it, it was impossible in the case studies to cover all facets of education or to include all of the participating nations. As a result, the project was limited to the four topics and three countries mentioned earlier. As far as we know, this is the largest, most complex cross-cultural project using the case study method that has ever been conducted in the social sciences or education.

In order to prepare for the project, a first step was to become familiar with the current literature in English, Japanese, and German related to the four topics on which the case studies were focused. With this background of information, it was possible to decide what was missing and what should be emphasized in the interviews and conversations.

The one-on-one interactions between researchers and participants in the case studies provided access to information that would be difficult, if not impossible, to discover through other means. Spending days rather than hours in the schools provides the variety of experiences through which a better understanding of the context for learning can be obtained. Conducting interviews and monitoring lessons in the participant's language and being able to probe and question the participants about their answers help to reduce the problems of translation and to clarify the content and meaning of the participant's responses. Having multiple researchers interact with participants reduces the likelihood of bias that might occur if the interviews were conducted only by a single investigator.

***Teaching in the United States is conducted in an individualistic, isolated fashion.***

Although the case-study researchers had lived in the countries where they conducted the research and most had completed their doctoral dissertations in that country, they continued to uncover new information and re-interpretations of what was presumably common knowledge. A few brief illustrations indicate the kinds of information that emerged.

Descriptions of interactions among teachers in the three countries offer a good example. Teaching in the United States is conducted in an individualistic, isolated fashion. After completing undergraduate work and spending a term practice teaching, the new teacher is placed in complete charge of a classroom. From that time on, U.S. teachers engage in few discussions with other teachers about the content of lessons or methods of teaching. In contrast, becoming a teacher in Japan is to engage in extensive interactions with other teachers throughout the teacher's career. Rather than relying primarily on university classes or practice teaching, Japanese teachers are expected to learn from each other. In Germany, the acquisition of teaching skill is dependent upon a two-year apprenticeship with skilled teachers, but once that is completed, teachers spend little time learning from each other. Reading the teachers' own descriptions of their desire for improving their professional training provides a vividness to the current problems that is missing in the reports of outside evaluations.

A second example of differences among countries deals with the how students in the three countries prepare for their end-of-school or college entrance examinations. The most prolonged preparation occurs in Japan, where students spend most of their senior year in high school studying for the examinations and attending special classes offered at school and at private academies (*juku*). In the United States, preparation for the college entrance examinations is more casual. Students, knowing that other factors than their scores on the examination are also important in determining admission to college, allocate little time to preparing for the examinations. Because the German exit examinations are directly tied to a small number of their high school courses, German students do not find it necessary to spend as much time studying as Japanese students do, but they do find it necessary to prepare more thoroughly than American students.

A third example deals with students' motivation for studying. The students indicated that their enthusiasm about studying depended on their perception of the relevance of their courses for their future careers, the quality of teaching, and the respect they had for their teachers. Their motivation was also influenced by their interactions with peers, dating, and part-time work. As students progressed through successive years in school, parents tended to become less and less directly involved in their children's education. Providing a supportive environment, access to after-school help, and books, among other study aids, were the main expressions of parental interest in all three countries. Personally helping their children became less likely as the difficulty of the curriculum increased and as the opportunities for interactions among members of the family became less frequent. As a consequence, adolescents in all three societies became increasingly dependent on their peers.

***Deductive reasoning occurred in 21 percent of the German lessons and in 62 percent of the Japanese lessons. It was never found in the U.S. lessons.***

These are glimpses of what the researchers heard about the attitudes, beliefs, and practices of parents, students, and teachers, all of which enter importantly into the effectiveness in educating students. Much larger amounts of information are contained in the case study reports, but even the case study reports provide only partial coverage of what is available from the researchers' reports of their interactions and observations.

## Videotape Study

Most American parents have spent little or no time visiting and observing their children's classrooms. Even U.S. teachers, after their practice teaching assignments, typically fail to observe each other's lessons, and few have had the opportunity to observe classrooms in other countries. This lack of experience, coupled with the remarkable differences among countries in teaching methods and procedures for classroom management, help to account for the high interest that has been shown in the TIMSS videotape study.

The study had several purposes: to describe what happens in eighth grade classrooms in the three countries and to provide quantitative indices of the teaching practices, to compare actual teaching practices with those recommended in current reform documents, and to evaluate the utility of applying videotape methods in future studies of instructional practices.

Although the initial goal of videotaping half of the classrooms participating in TIMSS proved to be unattainable, videotapes were made of a total of 231 lessons—a huge repository of videotapes. The permanent record of teaching practices and classroom activities contained in the videotapes makes many kinds of analysis easier to conduct than is the case with traditional narrative records. Teachers can readily observe their own strengths and weaknesses and those of other teachers, professionals can rate the effectiveness of different teachers and of different practices, and curriculum experts can evaluate the academic level at which lessons are taught. Applying these techniques produced many provocative findings, some of which are the following:

Post-secondary mathematics teachers were asked to view the tapes and attempt to determine the grade level of the topics contained in the videotape samples. The average grade level assigned to the U.S. lessons was seventh grade; to German lessons, mid-eighth grade; and to Japanese lessons, the beginning of ninth grade. These data suggest that Japanese students may excel in mathematics partly because their curriculum contains more advanced coverage of mathematics than is the case in Germany and the United States.

The viewing group was also asked to determine the quality of the lessons by judging the percentage of lessons that required the students to engage in deductive reasoning. This occurred in 21 percent of the German lessons and in 62 percent of the Japanese lessons. It was never found in the U.S. lessons.

A third type of judgment made in viewing the tapes was whether the teacher merely stated the principle by which a type of problem could be solved or attempted to help the children develop an understanding of the basis for the solution. It is evident in Figure 3 that vastly more topics in the U.S. than in the German or Japanese tapes contained concepts whose application was simply stated rather than developed logically. Additional evidence of the cognitive basis of mathematics appeared when judgments were made of the frequency with which lessons relied on the development of understanding rather than acquisition of routine skills. Figure 4 depicts the teachers' efforts to guide the students to an understanding of the concepts. This occurred nearly three times as often among Japanese as among U.S. and German lessons.

The overall effect of the video study is to describe a set of conditions that characterize the mathematics lessons of two groups of students: those who have displayed remarkably high levels of

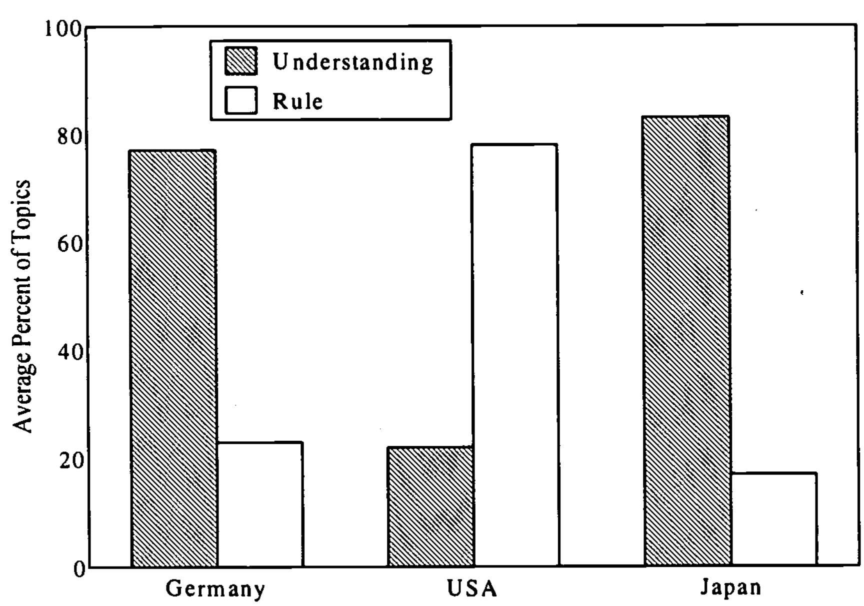


Figure 3. Percentage of topics for which teachers attempted to develop understanding versus following a rule.

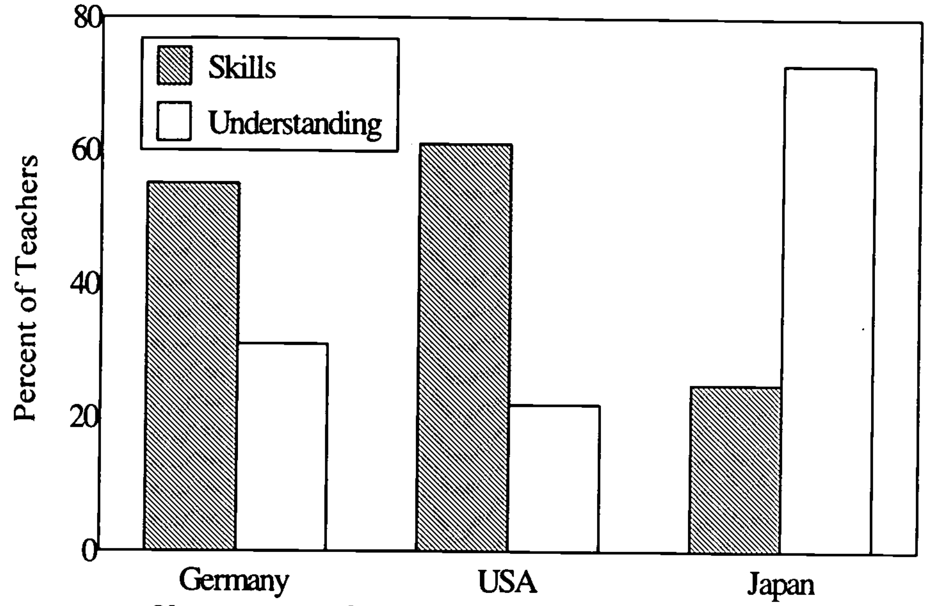

Figure 4. Percentage of lessons emphasizing understanding or a routine skill.

achievement in mathematics—the Japanese—and those who have not—the Germans and Americans.

These findings raise a number of questions: Do the differences between the successful and unsuccessful countries also appear in analyses of successful and less successful students within a country? Do other nations whose students are successful in mathematics share the characteristics that describe the Japanese lessons? Can interventions that attempt to modify teaching practices result in improvement in students' performance? These questions cannot be answered from what we know now about the relation of teaching practices and academic achievement.

The video study proved to be very productive, but whether it also describes what happens at other grade levels and for other subjects is a question for future research. Moreover, because the analyses were based on what occurred during a single lesson in mathematics, there is no information about the sequence of lessons involved in the development of a topic over successive days, or about situations where teachers and students are unaware they are being videotaped. However, the video study is an initial exploratory study, inaugurating a new research tool that represents the first effort to include observations of a nationally representative sample of classrooms. Expecting more from the research group is unrealistic within the constraints under which the project was conducted.

## Curriculum Analysis

A third innovation was the analysis of the curricula represented in the textbooks and teachers' guides of countries involved in TIMSS. These analyses served two purposes: the first was simply to develop a catalogue providing details concerning the topics covered in various countries; the second was to provide information about topics that were appropriate to include in the mathematics and science tests to be developed for use in TIMSS.

The sources of information used in the studies of the curricula were textbooks and teachers' manuals. In each country, teams of practicing mathematicians, scientists, educators, and specialists in assessment and implementation reviewed materials using a common framework. Each team member was given specific portions of the materials to review. After assigning scores to the material, they met to discuss their individual evaluations and to reach a consensus in their ratings.

Attempting to characterize the curricula in countries supporting such heterogeneous grouping of students as Germany and the U.S. is extremely difficult. In the U.S., mathematics and science education varies widely among states, districts, and even schools within districts. In Germany, a board of ministers of education from the various states meets to decide on recommendations for the academic curricula. Whether the recommendations are enacted into a law depends on the legislative body of each state. In marked contrast, the presence of national guidelines in Japan greatly reduces the need to select among curricula, for all schools and all textbooks in the country must comply with the curricular guidelines devised by the education ministry.

U.S. textbooks were found to lack focus and integration in the topics covered, in their difficulty, and in the expectations implied for student performance. Fewer topics were covered in the Japanese textbooks and the courses were integrated in the sense that they did not divide a subject such as mathematics into separate courses. For example, Mathematics I, the first course in Japanese high school mathematics, covers algebra, trigonometry, geometry, and statistics, rather than separate courses in each of these topics.

Some of the conclusions reached in the analyses of curricula were not especially novel. "The heart of the story," conclude the authors of one of the reports of the curricular analyses, "appears simple, almost self-evident. Classroom practices really do differ considerably among countries." Such a conclusion is of little use to policy makers. The analyses were of value for those who constructed the TIMSS tests, especially when the content of the curricula of various countries was being discussed. Other groups for which the detailed analyses would be useful are persons who wish to compare the content of their nation's curricula with those of other nations. In general, however, the curriculum analyses provide far more detail than the ordinary reader would find useful.

The arguments made by those who conducted the curriculum analyses emphasized the "splintered" nature of the U.S. curricula in science and mathematics. What is the reaction that can be made to the suggestion that "no intellectually coherent vision" guides the mathematics and science curricula in the United States? The alternative would appear to be either a national curriculum, which can bring coherence and integration to what occurs throughout the country, or national guidelines promoted by professional organizations. The first alternative is inimical to the American belief in the control of education by states, and the second lacks any power to secure the adoption of the curricula throughout the country.

# Conclusions

## Conclusions Concerning Methods

The methods used in TIMSS represented a great advance in terms of their comprehensiveness and diversity compared to the previous IEA studies of academic achievement. The inclusion of both science and mathematics within the same populations of students made comparisons of these two subjects possible. The case study and video study are useful complements to the quantitative data resulting from the tests and questionnaires in the main TIMSS study. Unless there is an understanding of the context in which learning occurs—what teachers teach, how they teach, whether the students are engaged, how parents participate, relations between school and home—the quantitative data remain indices of status that are difficult to interpret. On the other hand, doing nothing to evaluate the frequency or degree with which topics are mentioned or solutions are suggested leaves open the possibility for misrepresentations based on the reports of a few especially impressive conversations or observations. In short, the possible contributions of qualitative and quantitative methods are enhanced as information is supplied through each approach.

***"There are no educational characteristics that are present in every high-performing TIMSS country."***

## Policy Conclusions

One may question whether it was necessary or even desirable to attempt a study with over 500,000 participants. Carrying out the study and reporting the data proved to be extremely demanding in terms of time and funds, with the result that the reports are slow in being published. Data from the questionnaires have only been scratched and reports from the case studies and video study are only approaching the publication stage.

The authors of the various reports by members of the TIMSS staff express extreme caution in coming to any firm answers concerning the poor performance of the U.S. students and the seeming deterioration of their performance as students enter higher grades. For example, they have written: "No single factor or easily identifiable set of factors is clearly responsible for high achievement. Furthermore, every characteristic of a high performing country does not necessarily 'cause' its high achievement." No one would disagree, especially since some of the high-performing countries did not participate in all phases of TIMSS. Without their data it is impossible to check the consistency of many of the findings.

If one looks for definitive answers or interpretations of the performance of U.S. students in the various reports of the TIMSS main study, the search is bound to be frustrating. The writers indicate that analyses are incomplete, and that, "if anything, TIMSS suggests that there may be multiple recipes for excellence and that different combinations of factors may contribute to high achievement in different countries. There are no educational characteristics that are present in every high-performing TIMSS country." Even while agreeing that final answers must await further analyses of TIMSS data and the collection of additional information, it is possible to make some comments about American students' performance with reasonable levels of confidence. Despite the fact that the results from the video study and the case studies may not be definitive, they do provide strong hints about the kinds of variables that are likely to be associated with U.S. students' levels of performance.

***Adopting higher standards would not only have a positive effect on students' need to strive hard to improve their performance, but also on the publication of textbooks that represent more demanding curricula.***

## Possible Explanations for Poor U.S. Performance

Suggestions begin with the curriculum. The widespread adoption of what is sometimes called the "spiral curriculum" means that American teachers tend to spend little time on any topic because they assume that the topic will be covered again at later grades. Teachers also feel free to omit some topics completely, assuming that these topics, too, will be covered later. As a result, students in different classes at the same grade level cover widely varying topics, and in order to accommodate the interests of all types of teachers, the curricula for U.S. schools is, as it is termed in TIMSS, "a mile wide and an inch deep." In contrast, the curricula in many other countries are linear, comprehensive, and cumulative. If early steps are omitted or are weakly represented, later progress is impeded. The cumulative deficits, enhanced when steps are missing, may help to account for why U.S. students are behind their age mates in so many other countries, and why U.S. eighth graders had a lower standing among the cooperating nations than did U.S. fourth graders.

The lower standing of U.S. students may be due to their greater likelihood of acquiring rules that are automatically applied to problems rather than an understanding of the basis for such rules. This situation appears to be more likely when education standards are not high and students are expected simply to solve problems, rather than to understand the basis of their solution. Adopting higher standards would not only have a positive effect on students' need to strive hard to improve their performance, but also on the publication of textbooks that represent more demanding curricula.

The three societies held different interpretations of the feasibility of expecting all students to learn the curriculum. Mentioned only in the case studies is the Japanese emphasis on the role of effort and the belief that all students can learn the curriculum—attitudes that are in line with long-held

Confucian beliefs about the malleability of human beings. In contrast, the more biologically oriented German view holds that the primary influence is derived from inherited characteristics. The position of Americans is less clear. While not denying the importance of innate factors, the most frequent explanation of differences in academic ability offered by the American respondents was in terms of experiences resulting from the degree of family stability and support for education.

***The three societies held different interpretations of the feasibility of expecting all students to learn the curriculum.***

Another possible explanation of the American students' performance lies in the demands that are made of American teachers. Teachers talked about their heavy teaching loads, insufficient time for lesson preparation, concern about the adequacy of their professional training, their need to assume functions of child-rearing formerly performed by parents, families' lack of involvement in their children's education, and the need to adapt to ever-changing curricula. Attempting to respond to these demands has resulted in the high level of fatigue reported by American teachers. They had little to say about the usefulness of extending the length of the school day or school year, of allowing parents to choose the school their child will attend, or of establishing charter schools. They focused, instead, on the importance of improving the qualifications and working environments of those who are ultimately responsible for students' education: the teachers.

Attracting and retaining good teachers also depends on the status accorded them by the society in which they live. The U.S. public does not appear to be willing to support a professional status for teachers equivalent to that of professionals in other fields, such as law, engineering, and medicine. This is evident in their compensation, prestige in society, and in such a simple activity as being interrupted by others in the flow of their lessons, something that was rare in Germany and inconceivable in Japan.

Demographic factors are also obviously involved in students' academic achievement. Children attending poorly supported schools in impoverished or inner-city schools do not perform as well as those in affluent areas where funds are readily available to provide technology, laboratory, and library facilities or other types of equipment and supplies needed for lessons in various subjects. As long as the financial support of education depends strongly on real estate taxes, inequities are bound to continue in the quality of education provided students in different locations. Moreover, American students in different tracks enroll in different mathematics and science curricula. For example, the mathematics taught to vocational school students is different from that provided for those preparing to enter college. This is not the case in Japan, for example. Calculus is required of all high school students, regardless of their track, but the version taught to vocational school students is less rigorous in its proofs than is the calculus taught to students in the academic high schools.

There is no doubt that American schools could be improved if some of these alternatives were given appropriate attention and financial support. What is needed now is not the continued affirmation of the poor performance by American students or rationalization of why this should be the case. What is needed are firm data that will assist in explaining to the American public and policy makers what they can do to improve our students' ability to understand and apply the contents of contemporary science and mathematics. The widespread concern expressed after each successive publication of TIMSS results is indicative of the interest that has been aroused in Americans about their schools. There is reason, too, for optimism about the ability of American students to achieve at higher levels. When seven Chicago-area high schools in upper-income areas took the TIMSS tests recently, their scores were within the range of the top-scoring countries. The question that remains is what happens in these schools in order for interest and concern to be translated into such high performance.

# References

Three summary reports cover the initial findings obtained from TIMSS:


U.S. Department of Education. *Pursuing Excellence: A Study of U.S. Fourth- Grade Mathematics and Science Achievement in International Context.* (Washington, D.C.: U.S. Government Printing Office, 1997).

U.S. Department of Education. *Pursuing Excellence: A Study of U.S. Eighth Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context.* (Washington, D. C. : U.S. Government Printing Office, 1996).

U.S. Department of Education. *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context.* (Washington, D.C.: U.S. Government Printing Office, 1998).


More detailed information about the results for mathematics and science achievement appears in six publications from the TIMSS International Study Center in the School of Education at Boston College, Chestnut Hill, MA. (Both these and the preceding reports are available on the Internet: http://www.csteep.bc.edu/timss.)


Beaton, Albert E. et al. *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* (Chestnut Hill, MA, Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, 1996).

Beaton, Albert E. et al. *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* (Chestnut Hill, MA, Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, 1996).

Martin, Michael O. et al. *Science Achievement in the Primary School Years: ; IEA's Third International Mathematics and Science Study.*(Chestnut Hill, MA, Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, 1997).

Mullis, Ina V.S. et al. *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study.* (Chestnut Hill, MA, Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, 1997).

Mullis, Ina V.S. et al. *Mathematics and Science Achievement in the Final Year of Secondary School; IEA's Third International Mathematics and Science Study.* (Chestnut Hill, MA, Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, 1997).

Other overview publications include:

Beatty, Alexandra (Ed.). *Learning from TIMSS: Results of the Third International Mathematics and Science Study. Summary of a Symposium.* (Washington, D. C. National Academy Press, 1997).

Schmidt, William et al. *A Summary of Facing the Consequences: Using TIMSS for a Closer look at United States Mathematics and Science Education.* (Dordrecht, Kluwer Academic Publishers, 1998).

Stevenson, Harold W. "A Study of Three Cultures: Germany, Japan, and the United States—An Overview of the TIMSS Case Study Project." *Phi Delta Kappan* ( March 1998): 524-529.

Stigler, James W. and James Hiebert, "Understanding and Improving Classroom Mathematics Instruction: An Overview of the TIMSS Video Study." *Phi Delta Kappan* (September 1997): 14-21.

THOMAS B.
FORDHAM
FOUNDATION
OUTSIDE THE BOX


**The Thomas B. Fordham Foundation**

1015 18th Street, N.W. • Suite 300 • Washington, D.C. 20036
Telephone: (202) 223-5452 • FAX: (202) 223-9226
http://www.edexcellence.net



To order publications: 1-888-TBF-7474 (single copies are free)

# NOTICE

## REPRODUCTION BASIS

☑ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

(9/92)