TIMSS in Perspective:
Lessons Learned from IEA's Four Decades of International Mathematics Assessments

DRAFT
October 19, 2006

Ina V.S. Mullis and Michael O. Martin
Co-Directors, TIMSS & PIRLS International Study Center
Lynch School of Education, Boston College

Conducting International Assessments

International comparative studies of educational achievement had their origins in what is now known as the First International Mathematics Study (FIMS). FIMS was the first venture of the fledgling International Association for the Evaluation of Educational Achievement (IEA), the organization that pioneered international assessments of student achievement. Although FIMS was conducted principally between 1961 and 1965, its origins may be traced to a plan for a large-scale cross-national study of mathematics presented by the late Professor Benjamin S. Bloom in 1958 to colleagues in England and Germany (Husén, 1967).

The 1950s were a time of great educational development and expansion in many countries, with excellent descriptive studies of countries' educational systems conducted under the auspices of, for example, UNESCO and the OECD. However, the absence of any way of comparing the productivity or outputs of educational systems was keenly felt. Although the methodology for measuring student achievement through national surveys was established in many countries, it had not been implemented in a cross-national study. In FIMS, this methodology was applied for the first time in an international study of student achievement – mathematics in this case.

The overall aim of the FIMS researchers was to use psychometric techniques to compare outcomes of different education systems. An underlying idea was that the world could be considered as an educational laboratory, with each national education system an instance of a natural experiment in which national educational policies and practices are treated as inputs and student attitudes and achievement are treated as outputs.

Education was considered as part of a larger socio-political-philosophical system in which rapid change was taking place, and the most fruitful comparisons were considered to be those that took account of how education responded to changes in the society. One aim, therefore, was to study how mathematics teaching and learning had been influenced by such developments. The project addressed questions in three major areas: school organization, selection, and differentiation; the curriculum and methods of instruction; and the sociological, technological, and economic characteristics of families, schools, and societies.

There were several reasons for choosing mathematics as the subject of the first international study of student achievement. First, there was the importance of the subject itself. Countries were concerned with improving science and technology, and these are based in a fundamental way on the learning of mathematics. Second, there was the question of feasibility. Evidence from studies of the mathematics curriculum and the teaching of mathematics suggested that there was much in common between countries in terms of the curriculum and approach to teaching, and so there was a good chance that a mathematics test could be constructed that would be an acceptable match to countries' curricula. Furthermore, translation into the languages of instruction of participating countries was expected to be less problematic for mathematics than for other subjects because of the widespread acceptance of mathematical symbols and notation.

FIMS was conducted in 12 countries, with the results published in two volumes in 1967 (Husén, *op cit*). These volumes provide a wide-ranging summary of findings related to the structure and organization of the educational systems; the mathematics curriculum and how it is taught; the mathematics teaching force and how it is educated and trained;

and the attitudes, home background, and, most importantly, the mathematics achievement of students in the target populations. The target populations for the study were the grade in each country with most 13-year-old students and those students at the end of secondary education who had taken advanced courses in mathematics.

FIMS was important not only for the immense contribution it made in its own right to the comparative study of mathematics achievement, but also because of its pioneering role in establishing without doubt the feasibility of comparative study of student achievement across different countries with different languages and cultures. FIMS showed for the first time that international studies could produce data on educational contexts and student achievement (at least in mathematics) that could form the basis of policy research aimed at improving education around the world.

The Second International Mathematics Study

By the middle of the 1970s, IEA researchers had begun to think that the time had come for a second international study of mathematics, one that would build on the successes of the first study and incorporate all that had been learned about international studies in the intervening period. Planning for the second international mathematics study (SIMS) began in 1976, with data collection taking place in 1980-82. Like its predecessor, SIMS was a comparative study of student achievement in mathematics. However, in addition, SIMS placed great emphasis on the curriculum as a major organizing and explanatory factor, and also on the mathematics classroom – how teachers go about teaching mathematics to their students.

The three-level curriculum model introduced by SIMS has continued to underpin IEA studies in mathematics and science to this day. This model views the mathematics curriculum in a school system as having three aspects, each associated with a different level of the system. At the national or system level, the *intended curriculum* is comprised of the curricular goals and intentions that the country has for its students. These often are articulated through official documents such as national curriculum guides, course syllabi, and prescribed textbooks. The intended curriculum is what the ministry of education or similar responsible body expects to be taught in classrooms.

At the school or classroom level, the *implemented curriculum* comprises what is actually taught to students. Although, in an ideal world, there may be close alignment between the intended and implemented curriculum, in reality this may not be the case. The teacher in the classroom is the central agent in implementing the mathematics curriculum, and the choices that the teacher makes in terms of textbook or other instructional materials, emphasis on particular topics, and general teaching approach all have fundamental implications for the implemented curriculum. Teacher preparation can also have far-reaching implications, since teachers who have not received preparation in teaching the intended curriculum in mathematics are scarcely in a position to teach it effectively.

The third aspect of the curriculum model pertains to the student. The *attained curriculum* refers to the mathematics that the student has learned and the attitudes that the student has acquired as a result of being taught the curriculum in school. The attained curriculum may be considered the ultimate outcome of the educational process – how much of the mathematics intended at the national level and taught in the classroom has the student actually learned? The match between the intended, implemented, and attained curricula was an important focus of the second mathematics study.

SIMS was a very ambitious study, the results of which were summarized in a series of volumes published over several years. Analyses of the intended and implemented curricula were reported in Travers and Westbury (1989). The analysis of the intended curriculum provided a detailed picture of each participating country's coverage of the mathematics topics in the curriculum framework. Data on the implemented curriculum were collected from teachers through an innovative "Opportunity to Learn" questionnaire, which asked teachers to review the items in the SIMS mathematics test and indicate whether they had taught the mathematics needed to answer each item.

Student achievement in the 20 countries or educational systems that participated in SIMS was the focus of the second volume of results (Robitaille and Garden, 1989). Unlike in FIMS, where much effort was made to investigate causal relationships between the inputs of schooling and outputs in the form of student achievement, the focus of the SIMS achievement report was on describing a broad picture of mathematics education around the world. This included detailed information on student achievement in the major mathematics content areas addressed by SIMS – arithmetic, algebra, geometry and measurement, and descriptive statistics for the junior population (the grade with most 13-year-olds) and algebra, geometry, and elementary functions and calculus for the senior population (students at the end of secondary school having taken advanced preparation in mathematics). SIMS also reported extensively on student attitudes, preferences, and opinions, as well on teacher preparation and attitudes to mathematics.

TIMSS – The Third International Mathematics and Science Study

IEA's Third International Mathematics and Science Study (TIMSS) was the first to bring together mathematics and science in a single study. Building on the traditions of

FIMS and SIMS, as well as on IEA's two earlier studies of student achievement in science, TIMSS was the most ambitious IEA study to date. At the heart of TIMSS was a far-ranging assessment of student achievement in mathematics and science at three levels of the education system – grades three and four (the end of primary schooling), grades seven and eight (middle school or lower secondary) and twelfth grade (or final grade). The twelfth-grade assessment included tests of advanced mathematics and physics for students having taken advanced courses in these subjects, as well as tests of mathematics and science literacy for all students at this grade level.

TIMSS also included an in-depth analysis of mathematics and science curricula and an extensive investigation into home, school, and classroom contexts for learning (Schmidt, McKnight, Valverde, Houang, and Wiley, 1997). To provide fresh insights into how the curriculum is taught in mathematics classrooms, TIMSS incorporated video studies of instructional practices in Germany, Japan, and the United States (Stigler, Gonzales, Kawanaka, Knoll, and Serrano, 1999).

The launching of TIMSS at the beginning of the 1990s coincided with an upsurge of interest in international studies. There was a growing acceptance that effective mathematics and science education would be a crucial ingredient of economic development in the increasingly knowledge-based and technological world of the future. An additional impetus came from the recent breakup of the Soviet Union, which resulted in a host of newly-independent countries from Central and Eastern Europe anxious to participate in a study that promised to provide data to guide the revitalization of their education systems. Almost 50 countries participated in TIMSS in one way or another, making it the largest study of its kind at that time. The TIMSS data collection took place during the 1994-1995 school year.

TIMSS differed from earlier IEA studies in another important respect: it was well funded, principally by the U.S. National Center for Education Statistics and the National Science Foundation. With this support, TIMSS introduced a number of innovations that enhanced its validity and reliability as an international assessment of student achievement.

One of the essential underpinnings of TIMSS was an ambitious and extensive set of curriculum frameworks for mathematics and science (Robitaille, Schmidt, Raizen, McKnight, Britton, and Nicol, 1993). Designed to support both a comprehensive analysis of curriculum documents and textbooks and the development of a challenging and broad-based assessment of student achievement, the frameworks sought to improve the validity of the TIMSS assessments by specifying a wide range of mathematics and science content to be covered, employing a variety of cognitive skills, and recommending different types of item types and formats. The TIMSS achievement tests included not only multiple-choice items, which had commonly been used in international studies, but also constructed-response items that required students to provide a written response. TIMSS constructed-response items included both short-answer questions and questions requiring an extended response. There also was a performance assessment as a special component of the assessments at the fourth and eighth grades (Harmon, Smith, Martin, Kelly, Beaton, Mullis, Gonzalez, and Orpwood, 1997).

The TIMSS international reports of mathematics achievement in the middle school years (Beaton, Mullis, Martin, Gonzalez, Kelly, and Smith, 1996), in the primary school years (Mullis, Martin, Beaton, Gonzalez, Kelly, and Smith, 1997), and in the final year of secondary school (Mullis, Martin, Beaton, Gonzalez, Kelly, and Smith, 1998) describe student achievement in mathematics in the participating countries, and present a wealth of data about the social and educational context for learning in each country. The technical and operational aspects of TIMSS, and details of the methodology employed, are fully described in a series of technical reports (Martin and Kelly, 1996; 1997; 1998). TIMSS results were widely disseminated across participating countries, and the impact on educational systems has been considerable (see, for example, Robitaille, Beaton, and Plomp, 2000).

The drive towards valid and reliable measurement of student achievement in mathematics and science inevitably resulted in an assessment with far too many items to be administered in the time available to test any one student (90 minutes). This difficulty was readily solved by adopting a matrix-sampling booklet design, whereby each student took just a part of the assessment in the 90 minutes available. In the matrix-sampling approach, the assessment item pool is distributed across a number of student booklets, with each student responding to just one booklet. Although this booklet design solved the problem of administering the assessment, it created another problem: how to report student achievement in a range of mathematics and science content areas when each student has been administered just a few items in each area?

In a major step forward for international assessments, TIMSS addressed this issue by implementing a complex psychometric scaling technique that was developed by the U.S. National Assessment for Educational Progress (NAEP) for use in its national surveys of student achievement. This approach combines Item Response Theory (IRT) scaling with conditioning and multiple imputation methods to provides each student with estimated scores on the entire assessment even though the student has responded to just a part of it. With TIMSS, for example, this technique was used to assign each student scores in both mathematics and science overall, as well as in each of the TIMSS mathematics and science content areas, e.g., number, algebra, geometry, data, physics, chemistry, biology, and earth science, regardless of the booklet taken by the student. The student scores may be used to describe student achievement and also for analyses of the relationship between achievement and the many contextual variables that are measured as part of TIMSS.

The techniques described above that TIMSS pioneered in the field of international student assessment may be used in any study that is faced with the challenge of administering a student assessment that is much too large to be taken by a single student in the testing time available. The ability to combine a data-collection design that allows the assessment to be administered piecemeal with an analytic methodology that enables a student's scores on the entire assessment to be estimated from just the pieces of the assessment that the student took provides enormous flexibility to assessment designers. It means that subject matter specialists have great freedom to design wide-ranging assessments that do full justice to their assessment goals, secure in the knowledge that the resulting data can be reported in a meaningful way.

The TIMSS funding afforded unprecedented opportunities for development and analysis and reporting, but also brought with it the responsibility of ensuring that all aspects of TIMSS were conducted to the highest standards, and that this high quality was documented at every step of the way (see, for example, Martin and Mullis, 1996). IEA's studies had always applied state-of-the-art techniques to the development and implementation of its studies, but TIMSS was the first study with the resources to do this in a way that ensured high-quality comparable data from the participating countries. For educational data to be useful in an international arena, it is not enough for it to be of high quality from a national perspective, it also must be highly comparable from country to country.

More than any previous international study of student achievement, TIMSS placed the issue of data quality and data comparability at center stage. TIMSS identified a range of crucial areas for data comparability and went to great lengths to establish high standards, to report steps taken to meet these standards, and to attach consequences to failure to meet the standards. Among the questions that TIMSS focused on in striving for comparative validity were the following:

- Is curriculum coverage comparable?
- Are target populations comparable?
- Was sampling of populations conducted correctly?
- Are instrument translations comparable?
- Were the achievement tests administered appropriately?
- Was scoring of constructed-response items conducted correctly?
- Are the resulting data comparable?

This insistence on "comparative validity" has become a hallmark of TIMSS, and is one of the reasons that international comparative data on student achievement have become accepted as reliable instruments of educational policy analysis.

TIMSS – The Trends in International Mathematics and Science Study

The TIMSS results revealed a wide range of mathematics and science achievement both within and across countries. For several countries also, TIMSS revealed differences in relative performance across the grades. For example, in the United States, students performed just above the international average at the fourth grade but below the international average at the eighth grade. At the twelfth grade, average mathematics performance of U.S. students was among the lowest of the participating countries. This seems to suggest that U.S. students do reasonably well in the primary grades, but that the longer they stay in schools, the more they fall behind their peers in other countries.

Results such as these prompted the decision to administer TIMSS at the eighth grade again in 1999, four years after the original TIMSS administration in 1995. Since the cohort of students that was in fourth grade in 1995 would be in the eighth grade in 1999, TIMSS 1999 (or TIMSS-Repeat as it was known at the time) would allow countries to see if the relative performance of their eighth-grade students in 1999 reflected their performance when they were in the fourth grade in 1995, or whether the eighth-grade students of 1999 more resembled their eighth-grade compatriots in 1995.

In addition to the analysis of the fourth-to-eighth grade effect, TIMSS 1999 for the first time in an international study provided a solid measurement of trends in student achievement from two points in time – eighth-grade students in 1995 compared to eighth-grade students in 1999. This was possible because about one-third of the items from TIMSS 1995 had been kept secure for use in future studies, and these were combined with items newly-developed for 1999 to form the TIMSS 1999 assessment. Applying the

TIMSS scaling methodology to the data from 1995 and 1999 and capitalizing on the items common to both data sets, TIMSS was able to measure eighth-grade mathematics achievement in both 1995 and 1999 on the same scale. This provided countries that participated in both assessments with a direct measure of how much average eighth-grade mathematics performance had improved or declined over the four-year period (Mullis, Martin, Gonzalez, Gregory, Garden, O'Connor, Chrostowski, and Smith, 2000).

With the success of TIMSS 1995 and TIMSS 1999, IEA committed to TIMSS as a major component of its core cycle of studies. Now renamed the *Trends in International Mathematics and Science Study*, TIMSS is dedicated to providing sound measurement of trends in student achievement in mathematics and science at fourth and eighth grades every four years. TIMSS 2003 successfully reported achievement at both grades, extending the TIMSS trend line from 1995 through 1999 to 2003 for eighth-grade students and from 1995 to 2003 for students in fourth grade (Mullis, Martin, Gonzalez, and Chrostowski, 2004). TIMSS 2007, currently engaged in data collection, will further extend both fourth- and eighth-grade trend lines to include 2007 (Mullis, Martin, Ruddock, O'Sullivan, Arora, and Erberber, 2005).

First administered in 16 countries as part of the TIMSS 1995 twelfth-grade assessment, TIMSS Advanced assesses school-leaving students with special preparation in advanced mathematics and physics. With data collection in 2008, TIMSS Advanced will permit countries that participated in 1995 to determine whether the achievement of students having taken advanced coursework has changed over time (Garden, Lie, Robitaille, Angell, Martin, Mullis, Foy, and Arora, 2006).

International Comparisons in Mathematics Achievement

Since TIMSS 2003 represents the culmination of experience in assessing mathematics internationally, this section of the paper highlights results from TIMSS 2003 to provide a perspective about what has been learned student's mathematics achievement. For a full set of findings and a description of procedures, please see the *TIMSS 2003 International Mathematics Report* (Mullis, Martin, Gonzalez, and Chrostowski, 2004)

Variation Between and Within Countries in Mathematics Achievement

Exhibit 1 presents the TIMSS 2003 results for the 45 countries at the eighth grade (first page) and 25 countries at the fourth grade (second page) that participated fully in TIMSS 2003 meeting all guidelines for high-quality sampling and data collection. Countries are shown in decreasing order of average scale score, together with an indication of whether the country average is significantly higher or lower than the average of the TIMSS mathematics achievement scale which was set at 500 when the scale was developed in 1995. This represents a reanalysis of the presentation in the *TIMSS 2003 International Mathematics Report* which compared achievement to the average of the participating countries (467 at the eighth grade and 495 at the fourth grade). Using an approach dependent on participating countries has caused the international average to shift with each assessment, whereas the TIMSS mathematics achievement scale has a mean which is stable over time.

Insert Exhibit 1 about here

At the eighth grade, 13 countries performed above average on the TIMSS mathematics achievement scale, 7 countries performed about the same as the TIMSS scale average including the United States (average scale score 504), and 25 countries performed below average. The disproportionate number of countries performing below the scale average reflects the number of TIMSS 2003 countries facing economic and educational challenges. For example, countries with high values on the Human Development Index (HDI) provided by the United Nations enjoy long life expectancy, high levels of school enrollments and adult literacy, and a good standard of living as measured by per capita GDP. Only 12 of the TIMSS 2003 countries have an HDI greater than 0.9 (on a scale 0 to 1.0), including Australia, Belgium (Flemish), England, Israel, Italy, Japan, New Zealand, Norway, the Netherlands, Scotland, Sweden, and the United States. At the fourth grade, 13 countries performed above average including the United States (average scale score 518), 3 countries performed about the same as the TIMSS scale average, and 9 countries performed below average.

The TIMSS 2003 results make it clear that there is a substantial range in mathematics achievement across and within countries. Even though performance generally differed very little between one country and the next higher- or lower-performing country, with so many countries at the eighth grade from diverse cultural and economic backgrounds, the achievement range was particularly large – from an average of 605 for Singapore to 264 for South Africa. At the fourth grade, the range in achievement was from 594 in Singapore to 339 in Tunisia.

Exhibit 1 also shows a graphical representation of the distribution of student achievement within each country. The dark boxes at the midpoints of the distributions show the 95 percent confidence intervals around the average achievement in each country. Achievement for each country is shown for the 25th and 75th percentiles as well as for the 5th and 95th percentiles. Each percentile point indicates the percentage of students performing below and above that point on the scale. (For example, 25 percent of the eight-grade students in each country performed below the 25th percentile for that country, and 75 performed above the 25th percentile.)

In most countries, the range of performance for the middle group of students (50% of students between the 25th and 75th percentiles) was about 100 to 130 scale points—not too dissimilar from the difference between performance in Singapore, on average, and the average of the TIMSS mathematics scale. The range of performance between the 5th and 95th percentiles (90% of the students) was approximately 270 to 300 points in most countries.

As well as showing the wide spread of student achievement within each country, the percentiles also provide a perspective on the size of the differences among countries. For example, average achievement for Singaporean students at the eighth grade exceeded performance at the 95th percentile in the lower-performing countries such as Botswana, Saudi Arabia, Ghana, and South Africa. Similarly, at the fourth grade, average performance in Singapore exceeded performance at the 95th percentile in Iran, the Philippines, Morocco, and Tunisia. This means that only the most proficient students in the lower-performing countries approached the level of achievement of Singaporean students of average proficiency.

The TIMSS International Benchmarks of Achievement

Since many countries are working toward the goal of all students achieving at high levels in mathematics, it is interesting to look at mathematics achievement internationally from the perspective of how well countries are doing in having more students learn more mathematics. Considering achievement at four benchmarks along the

TIMSS mathematics scale provides a way to consider this question, and the success of the top-performing Asian countries in reaching the benchmarks demonstrates that mathematics proficiency for all students within a country is possible.

Because the TIMSS achievement scale summarizes the performance of each country's students across a large number (about 300 at eighth grade and 250 at fourth grade) of test questions, it is possible to examine achievement at various points along the scale in terms of the questions answered correctly by students performing at those scale levels (known as scale anchoring). Since the TIMSS test questions were designed to measure a wide range of student knowledge and proficiency in mathematics, analyzing student achievement at the four benchmarks provides a rich description of the mathematics that students know and can do at each benchmark. In analyzing performance across the test questions in relation to scores on the achievement scale at the eighth grade, it was determined that performance ranged from using relatively complex algebraic and geometric concepts and relationships at the advanced benchmark to having only some basic mathematical knowledge primarily in the area of number at the low benchmark.

While the full descriptions based on an ambitious scale-anchoring exercise conducted by international experts can be found in the *TIMSS 2003 International Mathematics Report*, achievement at the eighth-grade benchmarks can be summarized as follows:

Advanced (625): Students can organize information, make generalizations, solve non-routine problems, and draw and justify conclusions from data.

High (550): Students can apply their understanding and knowledge in a wide variety of relatively complex situations.

Intermediate (475): Students can apply basic mathematical knowledge in straightforward situations.

Low (400): Students have some basic mathematical knowledge.

At the fourth grade, students achieving at or above the advanced benchmark showed the ability to solve a variety of problems (appropriate to fourth-grade curriculum). Fourth-grade students achieving at the low benchmark demonstrated an understanding of whole numbers, the properties of basic geometric shapes, and how to read simple bar graphs. Performance at the fourth-grade benchmarks is summarized below.

Advanced (625): Students can apply their understanding and knowledge in a wide variety of relatively complex situations.

High (550): Students can apply their knowledge and understanding to solve problems.

Intermediate (475): Students can apply basic mathematical knowledge in straightforward situations.

Low (400): Students have some basic mathematical knowledge.

Exhibit 2 shows the percentage of students in each participating country that reached each international benchmark. Both the eighth- and fourth-grade results are presented in decreasing order by percentage reaching the advanced benchmark. The most striking finding is that such large percentages of students in the five top-performing Asian countries achieved at or above the advanced benchmark, especially at the eighth grade.

Insert Exhibit 2 about here

Most remarkably, 44 percent of the Singaporean eighth-grade students scored at or above the advanced benchmark. Not only was their average performance the highest at 605, but a substantial percentage scored at or above 625, demonstrating proficiency, for example, in solving non-routine problems in algebra and geometry. Four other Asian countries also had large percentages of eighth-grade students achieving at or above the advanced level. The Republic of Korea and Chinese Taipei both had more than one-third of their students performing at the advanced benchmark, 38 and 35 percent, respectively, followed by Hong Kong SAR with 31 percent. About one-fourth (24%) of the Japanese eighth-grade students achieved at or above the advanced benchmark.

Performance by students in the five top-performing Asian countries far exceeded that of the other participants. Besides the Asian countries, only Hungary and the Netherlands reached double digits, 11 and 10 percent, respectively, and 30 countries had 5 percent of their students or less reach the advanced benchmark. The United States had 7 percent of its eighth-grade students reach the advanced benchmark, which was average for the participating countries.

Although Exhibit 2 is organized to draw particular attention to the percentage of high-achieving students in each country, it conveys information about the distribution of middle and low performers also. Since students reaching the low benchmark may be considered to have reached some level of minimal competency in mathematics, it can be seen that the top-performing Asian countries also do an excellent job in educating all of their students. Ninety-nine percent of the Singaporean eighth-graders reached the low benchmark as did 96 to 98 percent of the students in the other four top-performing Asian Countries.

Trends in Mathematics Achievement

Exhibit 3 shows, in alphabetical order, the countries that have comparable data from two or three TIMSS assessments. At the eighth grade, 34 countries had data from two assessments, 17 with three-cycle trends. At the fourth grade, 15 countries had comparable data from 1995 and 2003.

Insert Exhibit 3 about here

At the eighth grade, only five countries showed a pattern of significant improvement over the 8-year period from 1995 to 2003. One of these countries was the United States, which improved about 9 scale-score points from 1995 to 1999 and another 3 points from 1999 to 2003 for a significant 12-point increase over the 8 years. Interestingly, two of the high-performing Asian countries showed improvement over the 8-year period—Hong Kong SAR and the Republic of Korea. Hong Kong SAR had a 17-point increase and the Republic of Korea had an 8-point increase. The remaining two countries with improvement were two Baltic countries, including Latvia (LSS) with a 17-point increase for their Latvian-speaking schools, and Lithuania with the dramatic improvement of 30 scale-score points, 20 of which were from 1999 to 2003. The improvements in Latvia (LSS) and Lithuania, as well as the strong performance of Estonia in its first TIMSS appearance, indicate successful educational reforms in that region of the world.

Unfortunately, more countries declined significantly in mathematics achievement at the eighth grade than improved. Countries showing a decrease in TIMSS 2003 from 1995, 1999, or both, included Belgium (Flemish), Bulgaria, Cyprus, Iran, Japan, Macedonia, Norway, the Russian Federation, the Slovak Republic, Sweden, and Tunisia.

At the fourth grade, many countries had significant increases in average achievement between 1995 and 2003. Participants showing improved performance included Cyprus, England, Hong Kong SAR, Latvia (LSS), New Zealand, and Slovenia. In contrast, the Netherlands and Norway showed significant declines. For the United States, achievement at the fourth grade remained essentially the same—518 in both 1995 and 2003.

Contexts for Learning Mathematics

At the early FIMS meetings in 1959, it was assumed that a cooperative approach to cross-national research would lead to better understanding of the more significant and effective educational procedures found anywhere in the world. In particular, the goal was to identify differences in education that were related to higher student achievement.

Even though a great deal of educational research is preoccupied with variations in instructional materials and strategies, the FIMS researchers observed that classrooms can look remarkably similar all over the world and concentrated on larger issues where countries tended to have the largest differences. In some sense, issues about the economic and cultural support for education remain the crux of the matter today as do the fundamentals of the what, who, and how of teaching—curriculum, teacher preparation, and effective instruction.

During the past 40 years of conducting international assessments, much has been learned about the conditions that can support and facilitate teaching and learning. This section of the paper describes some of the most important and extensively studied factors related to improving mathematics achievement. It must be emphasized, however, that there is little evidence of isolated factors consistently leading to excellence in mathematics achievement. Rather, it appears that high-achieving countries have systemic coherence across curriculum, teacher preparation, instruction, and assessment.

Economic Development

Since the national support for education is often related to level of economic development, Exhibit 4 shows the relationship between average eighth-grade mathematics achievement and gross national income per capita for the countries participating in TIMSS 2003. It is noteworthy that the wealthiest countries (those with GNI/capita of $20,000 or more) are to be found in the top half of the achievement distribution, while many of the poorer countries are in the lower half. The pattern is by no means clear-cut, however. For example, Korea and Chinese Taipei, two of the highest performing countries, have modest national income per capita. Also, many of the former Eastern Bloc countries have relatively low per capita income but relatively high average mathematics performance, perhaps the result of a traditional emphasis on mathematics education.

Insert Exhibit 4 about here

The Organization of Schools

The international studies of mathematics achievement have demonstrated that the school structure itself is one of the most important variables accounting for differences in student achievement. That is, what types of schools exist in a country, how do children progress through the schools, and how many of them are to be found at each stage? Is education universal for all students through the secondary level?

Although many of the world's nations still have major challenges in implementing universal education, considerable progress has been in this area since

FIMS. Of the twelve original participating countries, nine were European and the other three were Australia, Japan, and the United States—all high income, technologically developed nations, that realized the importance of raising the upper age of compulsory education and providing equality of educational opportunity for all. Still, among this group of countries the percentage of 16-year-olds remaining in school ranged from less than 25% in England to about 86% in the United States (Husén, 1967). It can be noted that the FIMS choice of the grade with the most 13-year-olds as a target population was based on the fact that students in some of the FIMS countries began leaving school after age thirteen. One of the most significant developments in educational systems around the world in the years since the first international mathematics assessment has been the large increase in the number of students completing upper secondary education. Unfortunately, considerable variation still exists among countries in completion rates.

Besides the universality of education, there remains the issue of differentiation or specialization, usually depending essentially on whether students are bound for university. Each country in the world has a set of policies and practices which influences the allocation and selection of students to each type of secondary school program, and which also determines the proportion of students who will complete each program, transfer to another program, or drop out of school.

This issue of completion rates as well as the percentages of students having taken advanced mathematics and physics was studied most recently in TIMSS 1995 as part of the assessment of the school-leaving population. To learn how much of the school-leaving cohort was still in school and represented by the TIMSS sample, a TIMSS Coverage Index (TCI) was computed for each country. The TCI was computed by forming a ratio of the size of student population covered by the TIMSS sample, as estimated from the sample itself, to the size of school-leaving age cohort, which was derived from official population census figures supplied by each country (Mullis, Martin, Beaton, Gonzalez, Kelly, and Smith, 1998).

One component of the school-leaving age cohort not covered by the TIMSS sample was the percentage of students no longer attending school. This latter percentage varied greatly from more than 50 percent in South Africa to about 35 percent in the United States to about 12 to 15 percent in Norway and France. The percentages of students having taken advanced mathematics ranged from 2 percent in the Russian Federation to 75 percent in Slovenia. For physics, the percentages ranged from 3 percent in the Russian Federation to 44 percent in Austria and Slovenia. One important aspect of TIMSS Advanced 2008 will be estimating any changes in the population of students taking advanced courses in mathematics and physics.

Curriculum

Beginning with SIMS (Travers and Westbury, 1989) and studied extensively in TIMSS 1995 (Schmidt, et al., 1997), much has been learned about the impact of the curriculum on mathematics achievement. Since the mathematics curricula across countries is discussed in particular elsewhere in this volume, the discussion here is confined to an overview of the most recent findings in TIMSS 2003.

In TIMSS 2003, most countries reported having the same intended curriculum for all students with no grouping of students. At the eighth grade, 38 participants reported this approach while nine countries reported having one curriculum for all students, but at different difficulty levels for groups of students with different ability levels. Only four countries—Belgium (Flemish), the Netherlands, the Russian Federation, and Singapore—reported having different curricula for different groups of students according to their ability level. At the fourth grade, all participants reported having just one curriculum for all students, and in most cases with no grouping by ability level.

Although the relationship between inclusion in the intended curriculum and student achievement was not perfect in TIMSS 2003, it was notable that several of the higher-performing countries reported high levels of emphasis on the mathematics topics in the intended curricula. For example, five of the six top-performing countries at the eighth grade (Chinese Taipei being the exception) included 80 percent or more the topics for all or almost all of their students. For most countries participating in TIMSS 2003 at the eighth grade, a substantial amount (70 percent for all students and another 6 percent for able students) of the mathematics topics assessed were in the intended curriculum. In only six countries were less than half of the topics included, and as might be anticipated, these countries had relatively low achievement (Botswana, Indonesia, Lebanon, Morocco, the Philippines, and Tunisia). Similarly, at the fourth grade, higher-performing countries had generally greater levels of coverage (e.g., Singapore and Japan about 70%) and lower-performing countries lesser levels (e.g., Tunisia and Morocco about 20%).

In the recent past, it has become common for countries' intended curricula to be updated regularly. At the time of the TIMSS 2003 testing, the official eighth-grade mathematics curriculum in about three-fourths of the countries was undergoing revision and in 27 countries had been in place for five years or less. At the fourth grade, about half the countries were undergoing curriculum revision and 20 had a curriculum that had been place for five years or less.

Having a rigorous curriculum is fundamental to giving students the opportunity to be exposed to increasingly challenging mathematical concepts and ideas, however, it is up to the teacher to actually provide classroom instruction in the topics. Given that the curriculum may be becoming somewhat of a moving target, it is interesting to note that according to their teachers at both grades, 72 to 73 percent of the students, on average, had been taught the TIMSS 2003 mathematics topics. At eighth grade, there was substantial agreement between the topics in the intended curriculum and the topics taught, but this was much less so at the fourth grade.

Teacher Experience and Preparation

Since the teacher can be a powerful force in determining the quality and quantity of educational achievement, each international study of mathematics achievement beginning with FIMS has sought to relate teacher variables to the test results, including the teacher’s level of training, amount and quality of teaching experience, and status as a professional worker. While there is little evidence in the data of a direct relationship between teacher training and student achievement, the studies have collected considerable information about policies and practices across countries.

For example, TIMSS 2003 collected information about five requirements for being certified or licensed to teach mathematics at the fourth and eighth grades, including obtaining a university degree, supervised practical experience (practicum), passing an examination, completion of a probationary period, and completion of an induction program. Interestingly, most countries required only two or three of these. Taking the requirements one by one for the 47 participating countries at the eighth grade, 33 of the countries required a university degree (or equivalent), 33 required some type of practicum, 28 required passing an examination, and 23 a probationary period. At the fourth grade, 19 out of 26 countries required some type of practicum, and 18 required at two or more of the following—passing an examination, a university degree, or completion of a probationary period. Only 11 countries at eighth grade and eight at fourth grade required completion of an induction program.

Teachers' reports about their preparation to teach mathematics revealed that approximately three-fourths of eighth-grade students and two-thirds of fourth-grade students were taught mathematics by teachers having at least a university degree or equivalent. At the eighth grade, on average, the majority of students had teachers who had studied mathematics (70%) or mathematics education (54%) or both (since teachers often reported that their study was focused in more than one area). As might be anticipated, the situation was different at the fourth grade where teachers typically studied primary or elementary education (for 80% of the students, on average).

Despite a certain lack of uniformity in their preparation and training, and achievement from students indicating the contrary, the teachers at both grades reported that they felt ready to teach nearly all the TIMSS 2003 topics. At the eighth-grade, at least 90 percent of the students had teachers who felt ready to teach all of the topics, with the exception of three of the data topics (sources of error, data collection methods, and simple probability). At the fourth grade, the results dipped below 90 percent for only two geometry topics (relationships between two- and three-dimensional shapes, and translation, reflection, and rotation).

Instruction

Since the types of learning experiences provided for students can vary from country to country, the question of whether different experiences lead to different achievement levels has been a significant area of inquiry in each international study of mathematics achievement. One obvious area of inquiry has been the amount of mathematics instruction given to students, since research about time on task and common sense suggest this as a prerequisite to learning. The TIMSS results, however, are not as conclusive as might be anticipated. For example, in TIMSS 2003, countries that reported providing the most instruction (more than 150 hours per year) at the eighth grade had relatively low achievement (the Philippines, Indonesia, and Chile). It is apparent that time alone is not enough, and that the time provided must be used effectively.

Interestingly, the instructional strategies used in mathematics classrooms around the world are remarkably similar. TIMSS consistently has found the textbook to be the foundation of instruction. In 2003, nearly two-thirds of the fourth-and eighth-grade students had teachers who reported using a textbook as the basis of their lessons. The three predominant activities at both grades, accounting for about 60 percent of class time, were teacher lecture, teacher-guided student practice, and students working on problems on their own.

Using technology to facilitate mathematics teaching and learning has emerged as topic of considerable study beginning with TIMSS 1995. Most recently, TIMSS 2003 found that calculator use at the eighth grade varied dramatically from country to country. The countries permitting calculator use for nearly all students (98% or more) included both high performing Hong Kong SAR and low performing Morocco. At the fourth grade, 57 percent of the students, on average were not permitted to use calculators. The countries permitting widespread calculator usage (90% of the students or more) all performed at about the international average—Australia, Cyprus, England, New Zealand, and Scotland. However, even in those countries, teachers reported asking relatively small percentages of students to do any calculator activities in half the lessons or more.

Access to computers remains a challenge in many countries. Teachers reported that, on average, internationally computers were not available for 68 percent of the eighth-grade students and 58 percent of the fourth-grade students. Even in countries with relatively high availability, using computers as often as in half the lessons was extremely rare (typically for 5% of the students or less) at either grade for any purpose. The greatest use at fourth grade was by the Netherlands and Singapore where 31 and 14 percent of the students, respectively, used the computer for practicing skills and procedures.

Home Support for Learning

IEA’s ongoing international assessments have shown that in almost every country students from homes with extensive educational resources have higher achievement in mathematics than those from less advantaged backgrounds. Most recently, TIMSS has focused on just a few central variables: level of parental education, students’ educational aspirations, speaking the language of the test at home, having a range of study aids in the home, and computer use at home and at school.

The different educational approaches, structures, and organizations across the TIMSS countries make comparisons of education levels difficult, and this is exacerbated by high levels of ‘do not know’ and missing responses in some countries. Nonetheless, TIMSS 2003 data make it clear that higher levels of parents’ education are associated with higher eighth-grade student achievement in mathematics in almost all countries. Further, students with university educated parents are particularly likely to expect to attend university themselves.

At both eighth and fourth grades, TIMSS 2003 found that students from homes where the language of the test is always or almost always spoken had higher average achievement than those who spoke it less frequently. Even though many countries tested in more than one language in order to cover their whole student population, on average, about 21 percent of students were from homes where the language of the test was spoken only sometimes, or never. Countries where the majority of students were from homes where the language of the test was spoken infrequently generally had relatively low mathematics achievement, including Botswana, Ghana, Indonesia, Lebanon, Morocco, the Philippines, and South Africa.

TIMSS 2003 (and previous studies) also found a clear-cut relationship, on average, between number of books in the home and mathematics achievement at both grades. In addition to literacy resources, having study aids such as a computer or a study desk or table at home was associated with higher student achievement. Mathematics achievement was positively related to computer usage, particularly at eighth grade, with average achievement highest among students reporting using computers at home and at school. At both grades, the data indicated that students were somewhat more likely to use a computer at home than at school. It should be emphasized, however, that at both grades the percentages of students reporting that they did not use a computer at all varied dramatically across countries. For example, at the eighth grade, the percentages varied from 0 percent in Hong Kong SAR and Chinese Taipei to about two-thirds in Botswana and Iran.

Attitudes and Values

The FIMS researchers noted that even though attitudes and values are directly related to educational practices and outcomes, this is a difficult area to measure and the inferences about cause and effect are more speculative. Nevertheless, they developed and administered six different scales about teaching and learning mathematics, the role of mathematics in society, and attitudes toward the environment to secure data on attitudes and values for descriptive purposes and later study.

More than four decades later, TIMSS 2003 reported results for two attitudinal scales—one about students' self-confidence in learning mathematics and one on the value students place on mathematics. At both grades, students had higher mathematics achievement, on average, if they were more self-confident in learning mathematics and placed higher value on the subject. It is clear that students' motivation to learn mathematics can be affected by whether they find the subject enjoyable, place value on the subject, and think it is important for success in school and for future career aspirations. Nevertheless, some countries with high average mathematics achievement (Korea, Japan, and the Netherlands) had less than 20 percent of their students placing a high value on learning mathematics. Some research has indicated that the students following a demanding curriculum may have higher achievement, but little enthusiasm for the subject matter (Shen, 2002)

Safe and Orderly Schools

Unfortunately, one of the emerging factors for study in TIMSS is school safety. TIMSS 2003 asked both teachers and students to characterize their perceptions of safety in their schools. At both grades, there was a positive relationship between teachers' reports of school safety and mathematics achievement. About three-fourths of the students were in schools the teachers characterized as safe, however, one-fifth were in schools where teachers were only in partial agreement, and about 4 to 6 percent were in schools deemed not to be safe. The students were asked about having things stolen, being hit or hurt, being made to do things, being made fun of, and being left out of activities. Fifteen percent of the eighth-grade students and 23 percent of the fourth-grade students answered "yes" to all five statements, and these students had lower average mathematics achievement than their counterparts in better school environments.

Summary

The FIMS pioneers of international mathematics assessment were experts in sampling, test and questionnaire construction, and statistical data analysis that sought to expand their field of expertise across countries to gain further insight into effective educational practices. By conducting an international assessment involving 12 economically developed countries, they sparked the possibilities for the future. Keeping in mind the model of the intended, implemented, and achieved curriculum provided by SIMS, as well as the rigor of comparative validity and innovative methodological advances introduced by TIMSS, many lessons have been learned over the past decades about how to conduct international assessments of high quality.

Much also has been learned about the educational process, and about what matters in teaching and learning mathematics. Education is arduous process. Different countries use different approaches, but to be effective always requires enormous effort. Success requires developing a rigorous and progressive curriculum, and providing all students an equal opportunity to learn it. Success also depends on economic resources and a strong-willed society to ensure that students are ready to learn and teachers are well prepared to provide instruction, as well as having the necessary facilities and materials.

Today, TIMSS 2007 data collection has begun at the fourth and eighth grades in the Southern Hemisphere, with more than 60 countries expected to take part by next June 2007. TIMSS Advanced for students the final year of secondary school (12^{th} grade in most countries) will occur in 2008, and planning is underway for TIMSS 2011. With the explosion of technological discoveries to facilitate improved measurement techniques as well as international communications, and the increasing number of diverse countries striving to become full participating members of our global enterprise, it seems that international assessments will continue to grow in both the quality of the data provided and the quantity of the participating countries.

References

Beaton, Albert E., Mullis, Ina V.S., Martin, Michael O., Gonzalez, Eugenio J., Kelly, Dana L., and Smith, Teresa A. 1996. *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Boston College, Chestnut Hill, MA.

Garden, Robert A., Lie, Svein, Robitaille, David F., Angell, Carl, Martin, Michael O., Mullis, Ina V.S., Foy, Pierre, and Arora, Alka 2006. *TIMSS Advanced Assessment Frameworks*. Boston College, Chestnut Hill, MA.

Harmon, Maryellen, Smith, Teresa A., Martin, Michael O., Kelly, Dana L., Beaton, Albert E., Mullis, Ina V.S., Gonzalez, Eugenio J., and Orpwood, Graham. 1997. *Performance Assessment in IEA's Third International Mathematics and Science Study*. Boston College, Chestnut Hill, MA.

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

Martin, Michael O. and Kelly, Dana L. (Eds) 1996. *Third International Mathematics and Science Study Technical Report Vol. I: Design and Development*. Boston College, Chestnut Hill, MA.

Martin, Michael O. and Kelly, Dana L. (Eds) 1997. *Third International Mathematics and Science Study Technical Report Vol. II: Implementation and Analysis, Primary and Middle School Years*. Boston College, Chestnut Hill, MA.

Martin, Michael O. and Kelly, Dana L. (Eds) 1998. *Third International Mathematics and Science Study Technical Report Vol. III: Implementation and Analysis, Final Year of Secondary School*. Boston College, Chestnut Hill, MA.

Martin, Michael O. and Mullis, Ina V.S. (Eds) 1996. *Third International Mathematics and Science Study: Quality Assurance in Data Collection*. Boston College, Chestnut Hill, MA.

Mullis, Ina V.S., Martin, Michael O., Beaton, Albert E., Gonzalez, Eugenio J., Kelly, Dana L., and Smith, Teresa A. 1997. *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study*. Boston College, Chestnut Hill, MA.

Mullis, Ina V.S., Martin, Michael O., Beaton, Albert E., Gonzalez, Eugenio J., Kelly, Dana L., and Smith, Teresa A. 1998. *Mathematics Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Study*. Boston College, Chestnut Hill, MA.

Mullis, Ina V.S., Martin, Michael O., Gonzalez, Eugenio J., Gregory, Kelvin D., Garden, Robert A., O'Connor, Kathleen M., Chrostowski, Steven J., and Smith, Teresa A. 2000. *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Boston College, Chestnut Hill, MA.

Mullis, Ina V.S., Martin, Michael O., Beaton, Gonzalez, Eugenio J., and Chrostowski, Steven J. 2004. *TIMSS 2003 International Mathematics Report: Findings form IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston College, Chestnut Hill, MA.

Mullis, Ina V.S., Martin, Michael O., Ruddock, Graham J., O'Sullivan, Christine Y., Arora, Alka, Erberber, Ebru 2005. *TIMSS 2007 Assessment Frameworks*. Boston College, Chestnut Hill, MA.

Robitaille, David F. and Garden, Robert A. (Eds) 1989. *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics*, Pergamon Press, Oxford.

Robitaille, David F., Schmidt, William H., Raizen, Senta, McKnight, Curtis, Britton, Edward, and Nicol, Cynthia. 1993. *Curriculum Frameworks for Mathematics and Science*. Pacific Educational Press, Vancouver.

Robitaille, David F., Beaton, Albert E., and Plomp, T. (Eds) 2000. *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science*, Pacific Educational Press, Vancouver.

Schmidt, William H., McKnight, Curtis C., Valverde, Gilbert A., Houang, Richard T., and Wiley, David E. 1997. *Many Visions, Many Aims Volume 1: A Cross-national Investigation of Curricular Intentions in School Mathematics*, Kluwer Academic Publishers, Dordrecht.

Shen, Ce. 2002. "Revisiting the Relationship Between Students' Achievement and their Self-perceptions: a cross-national analysis based on TIMSS 1999 Data." *Assessment in Education* 9 (2) 161-184

Stigler, James W, Gonzales, Patrick, Kawanaka, Takako, Knoll, Steffen, and Serrano, Ana. 1999. *The TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States. NCES 1999-074.* U.S. Government Printing Office, Washington, DC

Travers, Kenneth J. and Westbury, Ian. (Eds) 1989. *The IEA Study of Mathematics I: Analysis of Mathematics Curricula*, Pergamon Press, Oxford.