POLICY FORUM: SCIENCE EDUCATION

What Can We Really Learn from TIMSS?

William H. Schmidt and Curtis C. McKnight

There has recently been great concern about the quality of precollege education in science and mathematics in the United States and other countries. Evaluation of the state of education as well as policy options requires data on the practices and results of national education systems. The goal of the Third International Mathematics and Science Study (TIMSS) was to provide such data. It included an analysis of the education systems of over 40 countries, including textbooks and several achievement tests. The admittedly complex results of such a multifaceted study (*1*) have received considerable publicity and some criticism. Some have questioned whether TIMSS can even be used to inform policies related to science and mathematics education (*2*). Here we summarize some of the results (*3–8*), argue that many of the findings are robust, and discuss the implications for education policy.

Dimensions of TIMSS

The TIMSS study was wide-ranging, with many facets. Cross-national comparative achievement testing was done in each participating country in science and mathematics. Testing was done for the two adjacent grades containing the most 9-year-olds (hereafter referred to as third and fourth grades, as was the case in the United States), the two containing the most 13-year-olds (hereafter, seventh and eighth grades), and the final year of secondary school. The latter population included a sample of the entire population of students in the final year of secondary school who took science and mathematics general knowledge tests. It also included a specialized sample of those in their final year who had taken advanced courses in either mathematics or science (or both): that is, only physics for science and the appropriate advanced mathematics course (a mixture of precalculus and calculus in the United States).

Although this has received less attention, TIMSS also included analysis of official curriculum documents and textbooks as well as surveys of students, teachers, school officials, and national officials *(3)*. The goal was to develop a map of the structure of each national education system, both to inform study of that system and to guide sampling for achievement testing. A referee monitored sampling designs and implementations, and significant deviations were indicated in reported results.

The use of adjacent grades in the third/fourth- and seventh/eighth-grade populations allowed the estimation of differences between cross-section samples of grade pairs, which is a fair surrogate for gains that might have been measured by a true longitudinal design.

W. H. Schmidt is with the TIMSS U.S. National Research Center, Michigan State University, East Lansing, MI 48824, USA. E-mail: bschmidt@msu.edu. C. C. McKnight is in the Department of Mathematics, University of Oklahoma, Norman, OK 73019, USA. E-mail: cmcknight@ou.edu

Some Results of TIMSS

The TIMSS results show a decline in the relative standing of U.S. students from fourth to eighth grade in both mathematics and science, as compared to those in other countries. In science, U.S. third- and fourth-grade students were near the top of participating countries, whereas U.S. seventh- and eighth-grade students placed just above the cross-national average. In mathematics, the drop was from above the cross-national average to below it.

To relate achievement results more directly to curricular emphases, around 20 sub-areas were defined from the test items in each of mathematics and science for these two pairs of grades. Sub-areas where U.S. students did score somewhat better (for example, earth science, fractions, and decimals) corresponded to topics that received more extensive emphasis and coverage as revealed by the curriculum document and textbook analyses. The lower U.S. seventh- and eighth-grade scores for mathematics, compared to third- and fourth-grade scores, were true not only globally but also across most of the 20 sub-areas in mathematics. Use of paired grades also allowed single-grade "difference scores" to be computed. In the United States, in both science and mathematics, the seventh- to eighth-grade differences were consistently smaller than the third- to fourth-grade gains.

These achievement data seem to be clearly consistent with information gathered as part of the curriculum, textbook, and teaching data. U.S. curricula, as an aggregate, consistently covered more topics than did the curricula of virtually all other TIMSS countries. The U.S. mathematics and science textbooks included more topics and were the largest (literally) among those of the TIMSS countries. U.S. teachers covered more topics than those of most other TIMSS countries. They spent little time on most topics. For example, the only topic to receive more than 19 class periods of instruction in eighth grade was common fractions, a topic typically finished in earlier grades in many high-achieving TIMSS countries.

There was a close match between curricular differences and achievement differences. At fourth grade, only one nation (Korea) outperformed U.S. students on the total science score. When examined in detail, however, U.S. students' performance was weakest in all four areas related to physics and physical science. Similar weaknesses held at eighth grade. These results suggest that no serious foundation in physics was provided through the first 8 years, as compared to that provided by other TIMSS countries. Direct data on curricular emphases (for example, whether attention was to be focused on physical science topics) and textbook content (for example, pages devoted to physical science topics), as well as teacher reports of time spent covering physical science topics, support this suggestion (*7, 8*).

Similarly, in geometry the relative performance of U.S. students was much lower by the seventh and eighth grades than it was in the third and fourth grades (see the table). Analyses of documents and textbooks showed that U.S. treatment of geometry was comparatively weak after fourth grade [little focus on geometry in official curricula and few textbook pages devoted to it (and those only to simple aspects of geometry such as naming polygons)] (*6, 8*).

Other aspects of mathematics and science curricula were also consistent with the pattern of results from third to eighth grade. The U.S. curriculum appears not only to have been unfocused but highly repetitive, lacking coherence, and providing little rigorous intellectual challenge during the middle years, particularly when compared to those of other TIMSS countries (*4, 8*). Children in most TIMSS nations began the study of algebra, geometry, physics, and chemistry during fifth through eighth grades. In contrast, U.S. students continued studying elementary arithmetic and science as late as eighth grade.

In addition, out of almost 40 areas in mathematics and science combined and at fourth and eighth grades, the United States

Grade 4 Nation	% Correct	Grade 8 Nation	% Correct
Nations with Average Scores Significantly Higher Than the U.S.			
Hong Kong	74	Japan	80
Australia	74	Singapore	76
		Korea	75
		Hong Kong	73
		Czech Republic	66
		France	66
		Bulgaria	65
		Belgium-Flemish	64
		Russian Federation	63
		Slovak Republic	63
		Thailand	62
		Slovenia	60
		Hungary	60
		Switzerland	60
		Netherlands	59
		Belgium-French	58
		Canada	58
		Australia	57
		Israel	57
		Austria	57
		Latvia	57
		International*	56
		New Zealand	54
		England	54
		Denmark	54
Nations with Average Scores Not Significantly Different From the U.S.			
England	74	Lithuania	53
Scotland	72	Romania	52
Japan	72	Scotland	52
Singapore	72	Iceland	51
Korea	72	Norway	51
Canada	72	Greece	51
Slovenia	72	Spain	49
Netherlands	71	Sweden	48
United States	71	United States	48
Czech Republic	71	Cyprus	47
		Portugal	44
Nations with Average Scores Significantly Lower Than the U.S.			
Austria	67	Iran, Islamic Republic	43
Latvia	67	Kuwait	38
Ireland	66	Colombia	29
New Zealand	66	South Africa	24
Hungary	66		
International*	64		
Iceland	63		
Israel	62		
Norway	58		
Greece	53		
Thailand	53		
Cyprus	53		
Portugal	52		
Iran, Islamic Republic	42		
Kuwait	36		

SOURCE: I. V. S Mullis, M. O. Martin, A. E. Beaton, E. J. Gonzalez, D. L. Kelly, T. A. Smith, *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study* (Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, Chestnut Hill, MA, 1997), p. 47. A. E. Beaton *et al.*, *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study* (Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, Chestnut Hill, MA, 1996), p. 41.

*Average of national averages of the participating countries.

Table 1. Performance of students in various countries on the TIMSS geometry achievement tests in the third and fourth, and seventh and eighth grades. Countries are grouped in comparison to the scores of the United States.

was the only nation not to show differences from one adjacent grade to the next that would rank it in the top quartile of countries in at least one area (*8*). This pattern echoes the U.S. pattern of dividing attention among many topics. The cumulative effect of such small gains is seen in the consistent decline of U.S. rankings.

Most recent criticisms of TIMSS results have focused on the end-of-secondary-school results. These included a general knowledge test for mathematics and science designed to draw on knowledge at about the eighth-grade level in all countries. These results showed still lower comparative rankings of U.S. students in both mathematics and science, which given the relatively average to poor standing at eighth grade and the general pattern of many U.S. students not continuing to study science or mathematics should hardly be surprising.

The end-of-secondary results must be viewed in the context of the results from the other grades tested and, in the case of the advanced topics, from the analysis of curriculum documents and textbooks. The results for physical science in fourth and eighth grade also hardly seem to justify surprise at weak comparative performance in physics at the end of secondary school. The end-of-secondary results are consistent both with achievement results at earlier points at the sub-area level and with curricular and textbook data.

Some Criticisms of TIMSS

One criticism of the TIMSS results has been that the differences from sub-area to sub-area of achievement results indicate that the results are not robust and should not be used to inform policy discussions. However, the differences in achievement are consistent with and seem to reflect curricular differences. Curricular and systemic differences are legitimate variables that policy can affect. This sensitivity of achievement to curricular factors is thus a demonstration of robustness rather than a basis for criticism.

A second criticism has been that the sampling at the end of secondary school is inadequate and invalidates the TIMSS results. Were those particular results to be inadequate, they would hardly invalidate the more robust and less problematic results in the earlier grades and the relationship of achievement differences to curricular differences. As it is, the sampling and population definitions were more problematic at the end of secondary school, and more exceptions had to be documented. However, the consistency of the findings with the curricular results and the achievement results from the earlier grades indicates that these data provide at least general indications of the cumulative effect of schooling in mathematics and science, even if the interpretation of those indications is not as straightforward.

The end-of-secondary populations were designed for a yield study of what comes out of the pipeline of precollege education in mathematics and the sciences. It was never assumed that the students would be similar in age, years of schooling, or in percent of age cohort still in school. Were those the criteria, then the end-of-secondary results would be of questionable use (*9*). However, from a policy perspective it seems worthwhile to compare the yield of national educational systems. What is known at the end of secondary school is relevant to establishing policies to enhance international economic competitiveness. These data are thus useful as one indicator of comparative strengths and weakness that can help to inform policy discussions.

References and Notes

1. W. H. Schmidt and C. C. McKnight, *Educ. Eval. Pol. Anal.* **17**, 337 (1995).
2. I. C. Rotberg, *Science* **280**, 1030 (1998).
3. TIMSS reports are issued from various sources. The international TIMSS Study Center is located at Boston College and has issued several reports, including I. V. S. Mullis *et al.*, *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Study* (TIMSS International Study Center, Boston College, Chestnut Hill, MA, 1998). They also issued separate reports for primary school and for middle school mathematics and science. All of these reports are available at www.csteep.bc.edu/TIMSS. The National Center for Education Statistics (NCES) of the U.S. Department of Education has issued a separate series of U.S.–oriented reports that also place descriptive results in an international context. This series includes *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context* (NCES 98-049, U.S. Government Printing Office, Washington, DC, 1998). Corresponding reports are available from NCES on third- and fourth-grade science and mathematics and on seventh- and eighth-grade science and mathematics at nces.ed.gov/timss. In addition, a series of monographs have been prepared at the U.S. TIMSS Research Center (*4–8*).
4. W. H. Schmidt, C. C. McKnight, S. A. Raizen, *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education* (Kluwer, Dordrecht, 1997).
5. W. H. Schmidt *et al.*, *Characterizing Pedagogical Flow: An Investigation of Mathematics and Science Teaching in Six Countries* (Kluwer, Dordrecht, 1996).
6. W. H. Schmidt, C. C. McKnight, G. A. Valverde, R. T. Houang, D. E. Wiley, *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics* (Kluwer, Dordrecht, 1997).
7. W. H. Schmidt, S. A. Raizen, E. D. Britton, L. J. Bianchi, R. G. Wolfe, *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science* (Kluwer, Dordrecht, 1997).
8. W. H. Schmidt, C. C. McKnight, L. Cogan, P. Jakwerth, R. T. Houang, *Facing the Consequences* (in press).
9. In the context of this question of "comparability versus yield," the naming of the NCES end-of-secondary report (*Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*) seems particularly unfortunate. The report focuses on the United States, where the end of secondary school is the 12th grade, and thus seeks to put U.S. 12th-grade results in an international context. It does not imply that the end of secondary school in all TIMSS countries is the same as the U.S. 12th grade.