# What Are Science & Math Test Scores Really Telling U.S.?

by Alan S. Brown and Linda LaVine Brown

MAYBE YOU REMEMBER the headlines: "Math and science tests find fourth and eighth graders in U.S. still lag many peers," proclaimed the *New York Times*. "No gain by U.S. students on international exam," reported the *San Francisco Chronicle*. "Economic time bomb: U.S. teens are among worst at math," warned the *Wall Street Journal*. And just in case you didn't sense the alarm, "World crushes U.S. kids in math, science," summarized the *Boston Herald*.

**"World Crushes U.S. Kids in Math, Science"**

**REALITY** Students in affluent suburban U.S. school districts score nearly as well as students in Singapore, the runaway leader on TIMSS math scores.

Of course, every commentator has a seemingly obvious solution. Spend less time motivating students and more time enforcing standards, proclaimed the *Cleveland Plain Dealer*. Give parents school vouchers, and open charter schools, demanded a *Newark Star-Ledger* columnist. Throw away the calculators, and get back to basics, insisted a coalition of conservative Californians.

Now the cycle is about to begin again. This spring, between 18,000 and 19,000 students in carefully selected school districts around the nation will take the fourth Trends in International Mathematics and Science Study (TIMSS). Given every four years, TIMSS assesses the math and science skills of students in fourth and eighth grades. Roughly 60 countries will participate.

It is hard to say whether U.S. students will improve their standings from the TIMSS given in 2003. Four years ago, U.S. fourth graders ranked twelfth in math and sixth in science among the students in 25 countries who took the test. By eighth grade, U.S. students had fallen to fifteenth in math and ninth in science among 45 nations measured.

American students lagged behind such economic competitors as Singapore, Taiwan, China, Korea, and the Netherlands. Although they outperformed some competitors (Italy and Norway), they achieved their middle-of-the-pack position at the expense of Serbia, Lebanon, Tunisia, Ghana, and other developing nations.

To put these numbers into perspective, only seven percent of U.S. eighth graders met the TIMSS benchmark for advanced math and 64 percent for intermediate math. In Singapore, the numbers are 44 percent advanced and 93 percent intermediate.

TIMSS tests knowledge and skills taught in specific grades. It is not the only way to measure science and math skills. The Organization for Economic Development and Cooperation (OECD), a group of 30 wealthy democratic nations, measures educational outcomes—whether 15-year-olds can use their knowledge of math to solve problems.

The performance of U.S. students on this test is not so easily brushed off. In 2003, the OECD's Program for International Student Assessment (PISA) assessed 15-year-olds, ranking each nation and grouping students into six levels based on math abilities.

Overall, U.S. students ranked 24th among 29 OECD countries. Twenty-six percent of U.S. students fell into the bottom two math ability levels, compared to the OECD average of 19 percent. Only 10 percent of U.S. students were in the top two levels, compared to 15 percent for the OECD average. Roughly three-quarters of students in Finland scored above the mean of U.S. students taking the test. The results from the 2006 PISA (given every three years) should be available in 2007.

### TOO MUCH SCIENCE?

PISA and TIMSS are two sides of the same coin. TIMSS measures how well students absorb their curriculum, the knowledge and skills imparted in a particular grade. PISA looks at whether they can put math and science to work solving every day problems, such as placing a call to a different time zone, exchanging currency, calculating heights and areas, and reading graphs.

The vast majority of these students will never become scientists, mathematicians, or engineers, so why does it matter? Because today's students will be looking for jobs in an increasingly technology-oriented service economy where manufacturing and other traditional jobs are vanishing. They will need math to estimate projects and to interpret graphical information. They will vote on policies—stem-cell research, nuclear power, global warming, and teaching evolution—that hinge on an understanding of science. TIMSS and PISA let us understand whether or not we have prepared them well for the future.

Yet *prepared* means different things to different people. When TIMSS and PISA scores are reported in 2007, politicians and pundits will undoubtedly use the results to push their agendas. Expect rousing editorials that call for higher standards, greater accountability, school vouchers, certifying science and math majors to teach, and more accountability testing—and equally passionate calls for less standardized testing, more money for public schools, and improved teacher training.

Yet buried underneath the TIMSS and PISA rankings are surprising data that shatter many myths about American education, as listed at right [Fig. 1].

In fact, a close reading of TIMSS and PISA data not only helps educators benchmark U.S. practices, but also sheds light on ways to improve how our nation teaches science and math.

## THE GAP

Patrick Gonzales, who heads the TIMSS analysis effort at the U.S. Department of Education's national center for education statistics, is quick to point to potential inaccuracies in TIMSS. "This is a common measure of curriculum, even though every nation's curriculum is different," he explains. "Some countries don't start science education until the third or fourth grade, while others teach basic geometry in seventh and eighth grade.

"If U.S. kids don't do as well in geometry as in other areas, it is because they are rarely exposed to geometry by eighth grade. How would you expect them to do well? That doesn't mean our kids can't do the work. They just haven't learned it yet."

Not that Gonzales is sanguine. What really worries him is the wide performance gap within the U.S. "When you break students into standard sociological groups—parents with college education, minorities—the gap between the top and bottom is greater within the United States than between U.S. and top-performing Dutch students.

"There are significant differences between boys and girls in math and science in fourth grade," Gonzales acknowledges. "But they pale in comparison with the differences between white and black or poor and wealthy."

The difference between fourth-grade boys' and girls' TIMSS math scores is only one-tenth the difference between the U.S. and Singapore. The difference between eighth graders in poor and wealthy U.S. school districts, however, is 1.5 times greater than the difference between Singapore and the U.S. at both grade levels. When the same children are compared in science, the differences within the United States are four to five times greater than the differences between the U.S. and Singapore.

The gaps are even noticeable in the number of eighth-grade students who scored at the TIMSS intermediate level in math. The gross numbers show 93 percent of Singapore students and 64 percent of U.S. students reached these levels; U.S. males scored 65 percent, and females 64 percent. While 75 percent of white students rated intermediate status, only 35 percent of blacks and

## *myth*/reality

**Myth:** U.S. students rank far behind students in the best-performing nations.

**REALITY** Students in affluent suburban U.S. school districts score nearly as well as students in Singapore, the runaway leader on TIMSS math scores. The gap between America's top-performing and low-performing schools is much greater than the gap between America and other nations.

**Myth:** We don't teach enough math and science.

**REALITY** We teach far more math and science topics than virtually any other nation. Critics argue that "mile-wide, inch-deep" science and math curricula hop from subject to subject and keep students from mastering key concepts.

**Myth:** U.S. teachers lack the math or science background to teach effectively.

**REALITY** There is surprisingly little difference in how teachers with and without math or science backgrounds deliver lessons. The real differences lie in how teachers in the U.S. and top performing countries plan and teach lessons.

**Myth:** The numbers are rigged; our competitors cherry-pick their students and schools to look good.

**REALITY** TIMSS and PISA have grown more scientific and more accurate over time. Many people resist hearing the bad news. Close reading of TIMSS and PISA data can help educators benchmark U.S. practices, and improve how our nation teaches science and math.

**Fig. 1**

45 percent of Hispanics did so. Wealthier school districts scored 86 percent, while poor districts rated 32 percent.

"Shave off the middle-class suburbs and we're in the range of competitiveness," says Roger Bybee, a well-known educator who heads BSCS, a nonprofit science-curriculum development organization in Colorado Springs, CO.

But he also points to an issue as serious as the gap between rich and poor. "TIMSS is a curriculum-based test, and our better schools are competitive at demonstrating knowledge of the curriculum. But if we go back to innovation—critical thinking, reasoning, invention, and discovery—it's a little less clear that we are competitive.

"The scariest part of it," Bybee continues, "is that PISA also asked students about their educational aspirations, and U.S. students ranked second after those in South Korea. We have the highest aspirations, but we are near the lowest in terms of problem-solving skills. Our skills are not commensurate with aspirations."

## AN INCH DEEP

Yet the gap between rich and poor is not Bybee's only concern. Like many who have studied TIMSS and PISA results, he sees problems with the science and math curricula taught in the United States.

"The countries that we compete with, like Japan, teach six or seven major ideas per grade," Bybee explains. "We teach something like 75. For our kids, it is like sitting in front of a television when somebody who has the clicker is changing channels one after the other. You don't know the score of the football game, the plot of the story, or the guest on the talk show. It's just click, click, click—but in the end, you're going to have a quiz on this information. That's what it's like sitting in an eighth-grade classroom."

Bybee points to the work of William Schmidt, the noted researcher whose Michigan State University center coordinates U.S. participation in TIMSS. Schmidt was the first to apply the expression "a mile wide and an inch deep" to U.S. math and science curricula that jump from topic to topic without providing a coherent picture of how the topics fit together.

Schmidt often publishes graphs that show the sequence of mathematics topics studied in first through eighth grades. The vertical axis lists math topics from simplest (whole number meaning) on top to hardest (slope and trigonometry) on the bottom. Grade levels one through eight appear from left to right along the bottom.

For the top-achieving countries, the graph shows a diagonal progression from the top left (easiest topics, first grade) to bottom right (hardest topics, eighth grade). According to Schmidt, these nations teach first and second graders whole number meaning, whole number operations, and measurement units—and nothing else.

Schmidt's chart of three sample U.S. states shows no slope whatsoever. These states cover nearly every topic in nearly every grade. In addition to whole number meaning, operations, and measurements, first- and second-graders in all three states learn data representation and analysis, polygons and circles, estimating computations, measurement estimation and errors, 3D geometry, and patterns, relations, and functions. All told, the three states introduce 13, 20, and 30 different topics, respectively, in grades one and two.

"Compared with higher-performing countries, our curriculum is incoherent," adds Bybee. "It lacks focus and rigor. We tend towards an emphasis on terms and facts—and I'm not opposed to facts—but we know from contemporary models of learning that students also need conceptual ideas to hang those facts on."

Francis (Skip) Fennell, president of the National Council of Teachers of Mathematics and professor of education at McDaniel College in Maryland, agrees. He also underscores how hard it is to teach such a fractured curriculum.

"In my job, I get to hear the frustration of elementary classroom teachers around the country," he says. "In 49 of 50 states, there are state curriculum frameworks, and their requirements are all over the place. They have 20 to 30—and sometime *hundreds*—of objectives. How is a teacher going to achieve 100 objectives in 181 days of school?

"This sends a signal to a fourth-grade teacher who may not have a degree in math that all these 100 objectives are equally important. But they've never been equally important. Experienced teachers know this, but many teachers worry about covering all the topics because their students are going to be tested on them.

"We're not saying that experience with [a broad range of math topics] is necessarily bad, but we'd like to have deeper, more grounded experiences at lower grade levels," says Fennell. "Some states present probability as early as first grade. I don't know why a six-year-old kid needs probability that early."

The same process is at work in science. Here, teachers usually have even less of an academic background than they do in math. As a result, they tend to rely on textbooks and the lesson plans, pretests, quizzes, and tests that come with them. "The book should be a supplement or complement to the curriculum, but at least as far as science is concerned, the book is the *de facto* curriculum," says Bybee.

Like math, publishers try to cram as many topics into their science textbooks as possible, so they can sell the same book to several states. The results are often disconnected. One chapter follows another, but each exists in a vacuum with few strands tying one topic to another.

Also like math, state curricula often call for too much too soon. "What is the value of introducing the periodic table in third grade or plate tectonics in the sixth grade," asks Elizabeth Stage. She is director of the Lawrence Hall of Science, a combined museum and science teaching center at the University of California, Berkeley.

"What we know about teaching and learning is that we should teach with more depth and more time," says Stage. "So teachers have to take it on faith that if they take the time to teach the important ideas in depth, kids won't go belly up on accountability tests."

"To have the time for this, it is essential to make some choices that confront the essential purpose of schooling. Is it to prepare knowledgeable citizens, make students to be effective in the workplace, or enable them to become scientists?"

Not surprisingly, Stage favors national curriculum standards that would limit topics, provide adequate time to fully explore ideas, and enable publishers to produce thinner, more focused textbooks and teaching materials.

Bybee agrees: "In other countries, there is a ministry of education or some group that says, 'This is what curriculum must be.' Here we have the paradox that government can't plan curriculum so it defers to all 50 state departments of education, and they defer to school district. We're building incoherence into our curriculum."

## TEACHING

"Curriculum is important," admits Jim Stigler, professor of psychology at the University of California, Los Angeles, and senior vice president for research at Pearson Education. "But nothing impacts student learning unless it can transform the quality of teaching. A better curriculum will have to somehow change the quality of what happens in the classroom."

Stigler is the expert in how teachers actually teach their curriculum. In 1995 and 1999, he spearheaded a project to film how teachers in different countries taking TIMSS actually taught math and science. His team then analyzed the films, breaking down teaching in specific actions and comparing them to one another.

One key finding—obvious on the surface but with profound implications—was that each nation had its own distinct style of teaching. He found, for example, similar teaching styles among U.S. teachers who had majored in math while in college and those who had not. Even in a nation with 50 different state curricula and thousands of local modifications, the variation between teachers was much smaller than Stigler expected. It was certainly much small than the differences between U.S. teachers and those in Japan or Germany.

"This doesn't mean teachers don't need to know mathematics," he says. "It means that if we want to change teaching to engage students, just training people better in math or recruiting people with math backgrounds won't do it."

"Most Americans," he continues, "understand what it means to teach a skill. We understand what it means to solve problems by applying procedures to a situation. But we don't have a tradition of teaching concepts directly or making connections between concepts, facts, and procedures."

This assertion grows from an analysis of the type of problems presented in eighth-grade math classes. Most of the problems given to students fell into two categories: (1) basic computational skills and procedures; and (2) concepts and connections among mathematical ideas.

At first glance, two of the highest-scoring countries took diametrically opposed approaches in their classrooms. In Hong Kong, 84 percent of the problems were procedures, and only 13 percent connections. In Japan, 41 percent were procedures, and 54 percent connections. The U.S. fell somewhere in the middle, with 69 percent procedures and 17 percent connections.

Then Stigler looked at how teachers and students actually worked on these problems. For example, a teacher delivering a connections problem could ask students to try to figure how to find the area of different triangles on their own. Or he or she could transform it into a procedures problem by giving them the formula for area (1/2 base x height) and asking them to apply it to different triangles.

Hong Kong and Japan look more alike when viewed in terms of implementation. Although Hong Kong gave far fewer connection problems than Japan, it implemented 46 percent of them as connections problems compared with 48 percent for Japan.

The real shocker was the United States. It implemented *none* of the connection problems as connection problems. Not a single U.S. teacher taped actually taught a connection problem the way it was intended. Instead, they transformed 59 percent of connection problems into procedures problems.

Stigler's conclusion: "U.S. eighth graders spend most of their time in mathematics classrooms practicing procedures. They rarely spend time engaged in the serious study of mathematical concepts."

Science, he says, suffers from the same disconnection. "We see a lot of lab activity in American classrooms, but teachers never seem to get around to connecting the activity to science."

Stigler goes on to recommend that instead of focusing on improving teachers, it is time to rethink the culturally ingrained methods teachers use in their classroom.

This starts with curriculum. "Math," he points out, "is an elegant system with a few key principles that tie it together. We treat everything as if it is separate. The types of problems teachers present and how they implement them have the potential to tie these things together."

Changing America's teaching culture will take time. "First, we need to show teachers how to recognize what they're doing in the classroom and how it affects what students learn," says Stigler. "They're never asked to do that, not in their education classes, not in pre-teaching experiences, not as teachers.

"When we set teachers down in front of a videotaped lesson, their analysis is usually very shallow. Only if they learn to analyze what's actually happening during a class can they make intelligent changes in how they teach."

Stigler always wants to give teachers a sampling of alternative strategies. "There's very little variability in our own culture about how we teach math. We need to provide examples of how they teach in Czech Republic or Switzerland so they have something really different for comparison," he says.

Finally, he thinks teachers need a stable time and place where they can improve their teaching skills. "In Japan, there's a form of professional development called 'lesson study' that probably came out of total quality movement. Teachers, get together, identify a student problem, and plan a lesson as a group. Then one of the teachers teaches it while others observe. Then they debrief, revise the lesson and teach it again," Stigler explains.

Any resemblance between that and the approach taken in any U.S. school is entirely coincidental.

## CHANGES

Changes, however, are on the way. Curriculum, in particular, has received renewed attention. "What we can learn from other high-achieving countries is that a coherent curriculum is very important," says Bybee.

A stronger, more focused curriculum would also make it easier for poorer districts to teach science and math effectively. The Japanese, after all, do not divide students into high- and low-ability groups until high school. This suggests that a more focused curriculum would improve the performance of poorer as well as middle-class students.

Developing more coherent and focused curricula will require choices. Educators will have to decide which core skills students need and eliminate other topics so students have enough time to truly master them. It means creating textbooks and lesson plans that link math and science topics with one another, so students see how skills and concepts fit together into an elegant and powerful structure.

Bybee, Stage, Schmidt, and Stigler believe that this is best done by setting a single national curriculum standard. Yet they acknowledge that this is an uphill battle. "The current administration is funding math and science partnerships where it sends money to states and local school districts and they decide what's important to them," says Bybee. "That sounds good on the surface, but it builds incoherence into the curriculum. Since we're not going to take away state rights, we have to ask what we can do to influence the discussion in positive ways," asks Bybee.

One of those positive things is to push for voluntary national standards and curriculum models. "They're voluntary, not federally mandated," says Bybee. "But as states revise their standards, they act like the weak force in physics to bring curricula into alignment as they get closer."

This is already happening. In September, the National Research Council released a report urging changes in the science curriculum, *Taking science to school: learning and teaching science in grades K- 8*. "To be successful in science," the report stated, "students need carefully structured experiences, instructional support from teachers, and opportunities for sustained engagement with the same set of ideas over weeks, months, and even years."

Also in September, the National Council of Teachers of Mathematics issued a 41-page report (including appendices) that lays out a more coherent, tightly focused mathematics curriculum.

"We've already had eight different states—as varied as Florida, Pennsylvania, and Utah—contact us about revising their existing standards to align them with this document," says Fennell. "New York wants to talk to us, even though it just revised its curriculum in 2005. They see the need to focus on fewer topics, and there's immediacy to their request."

Simplifying and focusing curricula may give teachers the time they need to think about how they teach math and science. "It's very hard to teach differently from the way you were taught," says Stage. "It's very hard if you're a science teacher to change children's perceptions as learners. They expect you to teach one way.

"It's as if the weatherman suddenly starts to teach you how El Niño works instead of giving you the weather. You just want the weather. If a teacher says, 'I will guide you to figure this out,' the students will say, 'just give me the answer.' This has always been challenging, and it is more challenging now than before accountability testing."

Yet Bybee, for one, remains optimistic. Changes will take time, but they are starting to happen now. "These ideas are like a virus," he says. "We're infecting the system. Little changes can have big results if picked up by people who disseminate them to others. If we can infect the system, it can catch on in a big way."

The key is not to lower standards, but to make sure accountability tests actually test the type of changes Bybee and others want to make in the curriculum.

"Science," says Gerry Wheeler, head of the National Science Teachers Association, "has an interesting advantage: It motivates students because it is very engaging from a problem-solving point of view. Yes, there's some basic knowledge we want students to have, but what we really want to do is train their problem-solving and thinking skills."

This may be why so many educators worry about placing too much emphasis on accountability, TIMSS, PISA, and other test scores. The scores, by themselves, tell only part of the story. Instead of using a single number as a bludgeon to hammer home a point or buttress a point of view, it is better to peek behind the numbers and see what they tell us about our teaching practices.

"Otherwise," says Wheeler, "we may end up valuing what we measure, rather than measuring what we value."

**Alan S. Brown** of Dayton, NJ, has been a free-lance writer and editor for more than 25 years and is associate editor at *Mechanical Engineering*. A member of the National Association of Science Writers and an officer of Science Writers in New York, he graduated New College at Hofstra *magna cum laude* in 1974 (*insight@comcast.net*).

**Linda LaVine Brown** has taught special education for 30 years since her graduation from William Paterson College in 1977.