

EDUCATION WEEK

Published Online: January 18, 2013

Published in Print: January 23, 2013, as **International Tests Reveal Surprises at Home and Abroad**

International Tests Reveal Surprises at Home and Abroad

By **Martin Carnoy and Richard Rothstein**

Policymakers and pundits raise alarms whenever international test results are announced. In December, upon release of new scores from the Trends in International Mathematics and Science Study, or TIMSS, U.S. Secretary of Education Arne Duncan called them "unacceptable," saying they "underscore the urgency of accelerating achievement ... and the need to close large and persistent achievement gaps."

It was no different a little over two years ago, when the Program for International Student Assessment, or PISA, released its latest scores. Secretary Duncan said they showed Americans "napping at the wheel. ... As disturbing as these national trends are for America, enormous achievement gaps among black and Hispanic students portend even more trouble for the United States in the years ahead."

Such conclusions are oversimplified, frequently exaggerated, and misleading. They ignore the complexity of testing and may lead policymakers to pursue inappropriate and even harmful reforms.

Both TIMSS and PISA publish not only average national scores, but also a rich database from which analysts can disaggregate scores by students' socioeconomic characteristics. Examining these can lead to more nuanced conclusions than those suggested from average national scores alone.

Yet, for some reason, although TIMSS published average national results this past December, it only released its underlying database last week. This puzzling procedure ensures that commentators draw quick but ill-informed interpretations. Analysis of the database takes time, and headlines from the initial release are codified before scholars can complete more careful study.

Since the last PISA release in 2010 (of a test given in 2010), we have been digging deeper into its database and examining older databases for TIMSS and for our domestic National Assessment of Educational Progress. We concentrated on adolescent scores—8th graders on TIMSS and NAEP, 15-year-olds on PISA—in the United States; three top-scoring countries (Canada, Finland, and South Korea); three

[← Back to Story](#)

EDUCATION WEEK
**Multi-User
LICENSING**



**UNLOCK
ESSENTIAL
CONTENT**
Open your
team's potential

+ Click here for more info

similar postindustrial countries (France, Germany, and the United Kingdom); seven American states; and three Canadian provinces for which trends are available.

The TIMSS executive director deemed the **report of our findings** helpful (without endorsing our analysis in detail), but officials of PISA and of the U.S. Department of Education were harshly critical. We have posted their criticism, and **our response to it**, online.

Some of our conclusions are obvious; some are counterintuitive or startling. Here are a few:

- **A larger proportion of students in the United States is disadvantaged than in any comparison country.**

The number increased rapidly over the last decade, while in comparison countries it did not. Nonetheless, reading and mathematics achievement of lower-social-class U.S. students improved substantially, while achievement of similarly disadvantaged students declined in countries to which the United States is frequently unfavorably compared.

Thus, while the reading achievement on PISA of the lowest-social-class students in the United States grew by more than 0.2 standard deviations from 2000 to 2009, it fell by an even larger amount in Finland. In math, U.S. students from the lowest social class also gained substantially, while scores of comparable Finnish students declined. This is surprising because the proportion of disadvantaged students in Finland also fell, and we might expect this to make the task of devoting resources to them easier.

Certainly, even for the lowest social class of students, Finland's scores remain higher than ours, but examination of trends (performance changes over time), as well as levels (performance at a single point in time), challenges the easy assumption that simply imitating Finnish education is a recipe for U.S. success.

- **Unsurprisingly, every country has an achievement gap between its most- and least-disadvantaged students.** But unexpected is that this gap is smaller in the United States than in similar postindustrial countries, and often only slightly larger than gaps in top-scoring nations.

In some comparisons, the U.S. achievement gap is relatively small, while the scores of both disadvantaged and advantaged students are relatively low. But in other comparisons, the relatively small gap is attributable to our disadvantaged students' scores being relatively high, while the scores of our advantaged students are relatively low, especially in math.

This brings into question our widespread policy consensus that American educational problems are concentrated among disadvantaged students.

- **Creating truly representative national samples is difficult for test-makers, and errors can have big consequences.** For example, in 2009, PISA apparently oversampled low-income U.S. students who attended schools with very high proportions of similarly disadvantaged students, artificially lowering the average U.S. score. While 40 percent of the PISA sample in 2009 was drawn from schools in which half or more of the students were eligible for the subsidized-lunch program, only 32 percent of students nationwide attended such schools, according to National Center for Education Statistics data.



—iStockphoto.com/Nikolass Boden

A top PISA official disputes our claim. While acknowledging that 40 percent of the PISA sample for the United States was, indeed, composed of students who attended schools where more than half qualified for subsidized lunches, he suggests that free-lunch participation must have jumped because of the recession.

This explanation is theoretically possible, but not plausible, considering that the NCES database shows that for the same year (2009) a considerably smaller proportion of all secondary school students (not a representative sample) attended schools with such high poverty concentrations.

So we stand by this conclusion: If we make two reasonable adjustments to the reported U.S. average, our international ranking substantially improves. The first adjustment reweights the social-class composition of U.S. test-takers to the average composition of similar postindustrial countries. The other reweights the distribution of lunch-eligible students by the actual concentration of such students in schools. Once we further eliminate ranking differences based on tiny and meaningless statistical differences, these adjustments raise the U.S. international ranking on the 2009 PISA from the publicly reported 14th to sixth in reading, and from the publicly reported 25th to 13th in mathematics. While there is still room for improvement, these are more respectable showings.

- **We should be cautious about drawing conclusions about international comparisons from any single test.** From 2000 to 2006, U.S. math scores on PISA fell substantially, causing great alarm among U.S. policymakers and pundits. But few noticed that during roughly the same period, U.S. math scores on TIMSS were rising, as they did on NAEP. We cannot attribute this to alleged superior alignment of TIMSS to U.S. curricula because in the next period, the pattern reversed: From 2006 to 2009, PISA and NAEP scores both rose, while TIMSS scores were flatter.

- **Just as population sampling is complex, so is the choice of topics to test.** New TIMSS results show that the United States' and Finland's scores are now about the same overall in mathematics. But underlying data show that the United States does better than Finland in algebra but worse in number properties (e.g., using fractions). Algebra and numbers now each constitute 30 percent of test items (proportions have changed over time, adding another challenge to interpretation). But in the United States, we have made a big policy push to expose all students to algebra in 8th grade. If algebra had a bigger timss weight, our average score would be higher.

When it comes to domestic tests, U.S. analysts are relatively sophisticated. Federal law requires that average scores be disaggregated by race, ethnicity, and family income. But in our rush to condemn U.S. international performance, we ignore this common-sense approach. It is time for analysts to apply the same sophistication to international scores that is now second nature to us domestically.

Martin Carnoy (carnoy@stanford.edu) is a professor at the Stanford Graduate School of Education. Richard Rothstein (rrothstein@epi.org) is a research associate at the Economic Policy Institute, which is based in Washington. They are the co-authors of What Do International Tests Really Show About U.S. Student Performance? (Economic Policy Institute, 2013). Their previous collaboration was The Charter School Dust-Up: Examining the Evidence on Enrollment and Achievement (EPI, 2005).