

May-2017

# International Standardized Testing: The Measurement Problem

J. Michael Bodi

*Bridgewater State University*, [jbodi@bridgew.edu](mailto:jbodi@bridgew.edu)

---

## Recommended Citation

Bodi, J. Michael (2017). International Standardized Testing: The Measurement Problem. *Bridgewater Review*, 36(1), 13-16.  
Available at: [http://vc.bridgew.edu/br\\_rev/vol36/iss1/6](http://vc.bridgew.edu/br_rev/vol36/iss1/6)

This item is available as part of Virtual Commons, the open-access institutional repository of Bridgewater State University, Bridgewater, Massachusetts.

---

# International Standardized Testing: The Measurement Problem

J. Michael Bodi

*A specious argument or statement is one that is superficially plausible, but actually wrong. It's misleading in appearance, and misleadingly attractive...*

*(Merriam-Webster Dictionary)*

From day one, public education in America has been under siege. Among the first to establish common schools, many Americans, since the middle decades of the 19<sup>th</sup> century, have championed widespread learning for our people. But others have argued just as forcefully against it. Today's headlines about American public schools show that education is just as fraught as it was 150 years ago. Among the most troubling issues has been the standardized testing of students.

## What Generated this Enthusiasm for Testing?

Standardized testing in the U.S. is as old as it is controversial. How did we get into this current standardized testing mania? It has an interesting history. Prior to the 1950s, use of standardized tests in public schools was haphazard, with the exception of the Army using them to test recruits during World War II. In the wake of the assassination of President John F. Kennedy, President Lyndon Johnson enthusiastically took up some of the legislative agenda and plans for economic expansion initiated by Kennedy that led to an increase in federal funding for Johnson's War on Poverty, which included public schools. The Cold War had an important effect, too. Fearing that the Soviet Union, which launched Sputnik in 1957, was poised to pull ahead of the U.S. and create a missile gap, and later, triumph in the "space race," the federal government decided to become more

involved in our public school system so that future generations would be competitive with the Soviets. Though it was originally intended to be a jurisdiction governed solely by the states, 1960s-era politics pushed and pulled the federal government into school policy as education became a *national* matter, too important to be left to the states alone. Since then, scholars Laura Holden and Jeff Biddle

write, "It is now widely agreed that the Federal government has a responsibility to provide funding for education, and increasingly, in practice if not in principle, a right to exercise control over education" ([https://liberalarts.utexas.edu/\\_files/ms37643/Holden-Biddle\\_for\\_Hamerama.pdf](https://liberalarts.utexas.edu/_files/ms37643/Holden-Biddle_for_Hamerama.pdf)).

Federal interests since the 1960s have called for national education standards, measurable through standardized tests. In the 1960s, standardized testing began in earnest, but it was nothing like it is today. When the Elementary and Secondary Education Act (ESEA) was signed by President Johnson in 1965, the use of norm-referenced standardized testing in the public schools became mainstream. Until the 1990s, the types of standardized tests that school districts could purchase and administer were provided by test-making companies. Some states did not mandate testing, some did; and those that did didn't always determine which test should be used by their school districts. It was pretty hodgepodge. In the early 1990s, the "educational reform movement" took root, which led many states to employ a "high-stakes test" based on standards created within their states. "High stakes" simply means that students who don't pass the test can't graduate from high school. It kicked off a mania for "accountability."

In 2001, the federal government instituted No Child Left Behind (NCLB), a continuation of the ESEA

---

The most glaring oversight made by the researchers who manage the test data collection is that the students who are taking the test are not the same kinds of students among countries; the test results compare apples to oranges.

from the 1960s but expanded to include state-mandated standardized testing nationwide. There was no nationally required test per se (though an elective test, the National Assessment of Educational Progress [NAEP], has been administered to many fourth- and eighth-grade public-school students every year since 1964; [www/nces.ed.gov](http://www/nces.ed.gov)). With the NCLB, each state was directed to create its own exam, which led to widespread competition by testing companies to secure state contracts. In Massachusetts, the test is called the Massachusetts Comprehensive Assessment System, or MCAS.

## International Standardized Testing

Standardized testing has succeeded in providing benchmarks that American educators and legislators have used to identify strengths and weakness in student learning. But they have also created the possibility for American students' scores to be held up against their counterparts in other countries. In recent international comparisons using standardized tests, U.S. students have done remarkably poorly. The argument that the U.S. lags behind other countries in education is a routine headline,

most often trumpeted immediately after the test results come out. How did we get to a position of comparative weakness? The answer to that question is rooted in both the disparate methods we use to measure proficiency and, more basically, how we define the purpose of education. With regard to the issue of international standardized testing, those who disparage American public schools don't understand the fundamentals of subject sampling.

I believe that there is such a thing as a good standardized test. My qualifications are that I teach assessment

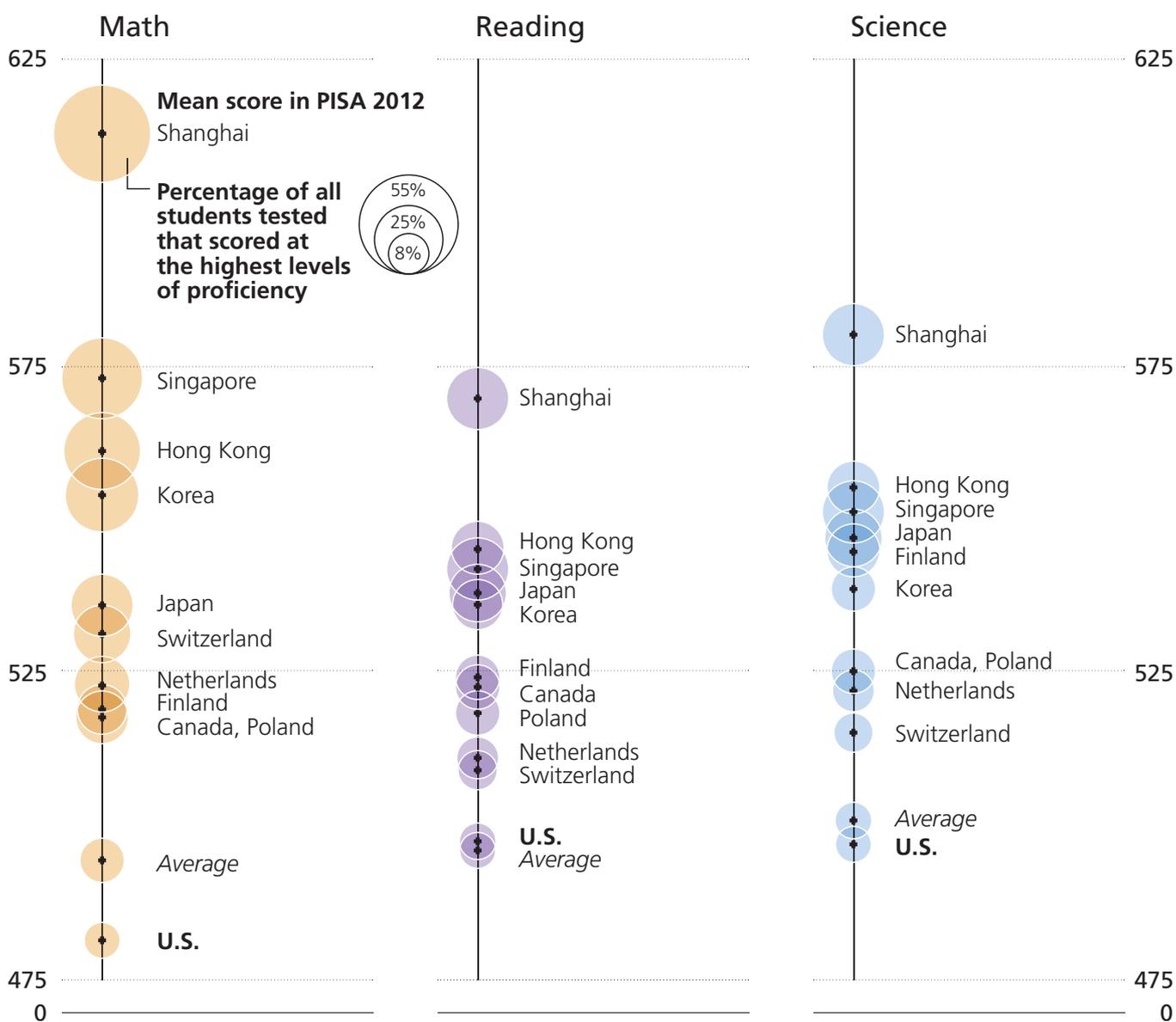


Figure 1. Results of the PISA (Programme for International Student Assessment) Test 2012 (Sources: Organization for Economic Cooperation and Development [2013] and PISA 2012 Results, Washington Post, December 3, 2013).

---

(including standardized tests and testing protocols) to educators at both the undergraduate and graduate levels, and educational research methods at the graduate level. Some of the best-known international standardized tests (IST) include the Trends in International Mathematics and Science Study (TIMSS), which is a series of international assessments of the mathematics and science knowledge of students around the world. Another is the Programme for International Student Assessment (PISA), which is a world-wide study done by the Organisation for Economic Co-operation and Development (OECD). PISA is a test administered to 15-year-old school pupils to discover their scholastic performance on mathematics, science, and reading. These two are good, valid and reliable tests. Each of them assesses accurately what it was designed to test and the results have been reliable over time.

But they do have problems. The first of these is that their results are discussed by public officials and scholars as if they somehow *prove* that American students, their teachers, and our public schools are lagging behind at least half of the rest of the world's schools. IST is fraught with complicated variability due to language, cultural and class differences, sampling procedures, and more. According to Ludger Wößmann (*Oxford Bulletin of Economics and Statistics*, 2003), "international differences in student performance cannot be attributed to resource differences but are considerably related to institutional differences." On the 2012 PISA assessing students from 65 countries, the *Washington Post's* Lyndsey Layton wrote, "students from the United States posted mostly average scores" (*WP*, 3 December 2013). Figure 1 speaks for itself.

However, the most glaring oversight made by the researchers who manage the test data collection is that the students who are taking the test are

No Child Left Behind, with all its problems, put a safety net under the lowest-achieving learners. At the same time, it put a cap on our highest-achieving students as the bulk of resources go to those most in need.

---

not the same kinds of students among countries; the test results compare apples to oranges.

The students being tested are 15 years of age (using the PISA data). But who are these 15-year-olds? They are not everywhere the same. Fifteen-year-olds in Shanghai, Singapore, Japan, Netherlands—literally every place in the world except the U.S.—are from groups of already-selected students who, prior to age 15, have passed their national tests and gone on to high school, while those who didn't pass went into some sort of training or other non-academic path. In Germany, 10-year-old students have taken a national test; 12-year-olds do so in Japan. In the U.S., there is no national test, and virtually all go on to high school. By the time the sample populations sit the IST, only the highest-performing students (academically speaking) of every country except the U.S. take the test.

A look at the way sample populations are assembled for PISA testing is instructive. In each country tested, there is a random sampling of schools with usually 150 chosen. Within those schools, there is a random sample of students selected, usually numbering 4000. Excluded from testing are students with cognitive impairments, those from small schools, and non-native-language speakers. Samples are stratified according to

geographic regions, school types (public and private—the U.S. does not include private schools), languages of instruction, levels of urbanization, socio-economic indicators, and schools' performance on national exams. But there are some problems here with these sampling procedures. PISA's national scores are averaged without disaggregating those scores by social class. Also, as Economic Policy Institute data (2013) show, student performance is strongly correlated to institutional performance, which can't be seen in PISA's presentation of test results (Economic Policy Institute, 2013). Most importantly, because the U.S. does not have a centralized education system, any sampling that is done comes from institutionally uneven terrain.

The United States, with the most diverse population among participating nations, includes far more disadvantaged students in the ISTs. There is a glaring difference as to how education is organized in the U.S. as against all other countries. Almost every country in the world has a centralized education system. In Japan for example, every child in the third grade is learning the same content at the same time during the school year; so, if a family moves, the child resumes the third grade and won't miss a lesson. In the U.S., each state has its own approach to conducting education in the public schools; if a family moves from Massachusetts to Idaho, its third grader could very well

be behind or bored silly in his new school placement. No smooth transition is guaranteed, or likely. The right to an education is not included in the U.S. Constitution, making it the states' right. Therefore, we might expect a great deal of variation among states in how much public education is valued, in how curriculum is delivered, and in how well students across the nation learn. And that expectation is borne out by the IST results. When we

our great achievement in being first among the states. But it's important to consider that these scores mean only that our students are number one in scoring high on standardized tests. I'd be much more impressed if a survey of our students' attitudes about learning in school was also extremely high. A positive correlation between high achievement and high interest in learning would really be a significant accomplishment.

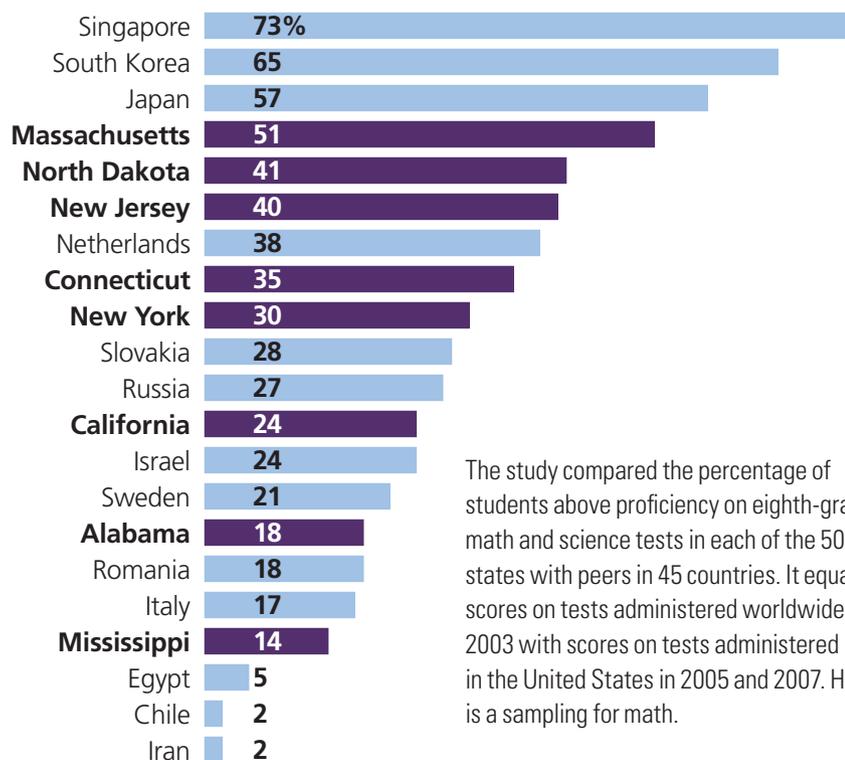
students as the bulk of resources goes to those most in need. If we truly believe in and support the notion that all of our children should be as fully educated as possible, then we can't ignore those who, as the National Association for Gifted Children ([www.nagc.org](http://www.nagc.org)) reminds us, should be accelerated.

Increased federal funding is critically important to supporting the educational bedrock of our society, but it can't level the playing field divided by state jurisdictions. And economic disparities—among states, among regions—persist. And as education scholars Stephen Gorard and Emma Smith (*Comparative Education*, 2004), write: "Nationally comprehensive systems of schools tend to produce narrower social differences in intake and outcomes... Systems with more differentiation lead to greater gaps in attainment between social groups."

The experience of U.S. students with International Standardized Tests underlines the fundamental challenge of American schooling. So, then, we face two options. Do we educate all our children, from low to high achievers, thereby causing us to remain somewhere in the middle in international standardized testing? Or do we segregate the students in schools, tracking them by academic achievement so that our highest achievers can be in the sample of students who take the IST?

I'll take the first, and damn the naïve interpretations of the IST data.

## States vs. Countries in Math



The study compared the percentage of students above proficiency on eighth-grade math and science tests in each of the 50 states with peers in 45 countries. It equates scores on tests administered worldwide in 2003 with scores on tests administered in the United States in 2005 and 2007. Here is a sampling for math.

Figure 2. *States versus Countries in Math*  
(Source: American Institutes for Research and the New York Times, 14 November 2007).

disaggregate the PISA data (Figure 2), when each of the 50 states in the US is represented as a country, the test results give us a very different picture.

Once separated, the various 50 states score from the highest to the lowest among the list of U.S. states and countries on the PISA. So when one aggregates and averages those scores, the U.S. students regress to the mean, which is literally in the middle. Here, in Massachusetts, we can be proud of

## Conclusion

The United States has the most diverse society on the planet. Its educational system must provide, by state law, access to a free public education. We have made great strides in raising the academic standards for all children. No Child Left Behind, with all its problems, put a safety net under the lowest-achieving learners. At the same time, it put a cap on our highest-achieving



J. Michael Bodi is Professor and Chair in the Department of Secondary Education and Professional Programs.