

Advertising  
PLAY AGAIN



**Pinnacle**  
by GlobalScholar

**THE RIGHT GUIDE FOR YOUR SCHOOL IMPROVEMENT INITIATIVES**

[LEARN MORE](#)

**Countdown to Fall Conference**

Chicago, Ill.  
October 29-31, 2010  
**10 days, 20h : 46m : 54s**

[ASCD Edge](#) [HOME](#) [JOIN](#) [SHOP](#) [myASCD](#)



ASCD  Educational Leadership

**About ASCD**

[Join / Membership](#)

[Community](#)

[Research A Topic](#)

[Publications](#)

[Blog](#)

[Books](#)

[Educational Leadership](#)

[E-Publications](#)

[International Journal](#)

[Newsletters](#)

[Past ASCD Publications](#)

[SmartBrief](#)

[Professional Development](#)

[Online Learning](#)

[Conferences](#)

[Initiatives & Programs](#)

[Public Policy](#)

[ASCD Store](#)

**MISSION:** ASCD (formerly the Association for Supervision and Curriculum Development) is a membership organization that develops programs, products, and services essential to the way educators learn, teach, and lead.

**We are here to help!**

1703 North Beauregard St.  
Alexandria, VA 22311-1714

**Tel:** 1-800-933-ASCD (2723)  
**Fax:** 1-703-575-5400

8:00 a.m. to 6:00 p.m. EST  
Monday through Friday

Local to the D.C. area:  
703-578-9600, press 1

Toll-free from the U.S. and  
Canada: 1-800-933-ASCD  
(2723), press 1

All other countries  
(International Access Code):  
+1-703-578-9600, press 1

**PUBLICATIONS**



**EDUCATIONAL LEADERSHIP**

[Home](#) [Current Issue](#) [Archives](#) [Search](#) [Buy](#) [Contact](#)

November 2009 | Volume 67 | Number 3

**Multiple Measures** Pages 32-37

**The Big Tests: What Ends Do They Serve?**  
*Gerald Bracey*

**To measure the quality of our schools, we need more instruction-sensitive measures than NAEP, PISA, or TIMSS.**



I was recently interviewed by the editor of my local paper, the *Port Townsend Leader*, who expressed a pretty low opinion of tests. His wife teaches 3rd grade in a public school, and he can't imagine how *anyone* would think that a test could reveal more information about a child than a teacher collects as a matter of course. I agree.

In the last 50 years, the United States has descended from viewing tests first as a useful tool, then as a necessity, and finally as the sole instrument needed to evaluate teachers, schools, districts, states, and nations (Bracey, 2009). In a nation where test mania prevails, tests will occupy part of the education landscape until we can dig ourselves out of that 50-year hole. In the meantime, it's interesting to consider what some of the well-known testing programs measure and what their appropriate (and inappropriate) uses might be. Here I look at three testing programs—one domestic and two international.

**National Assessment of Educational Progress (NAEP)**

When U.S. Commissioner of Education Francis Keppel proposed the NAEP in the 1960s, he ran into a buzz saw of objections from virtually every education organization in the nation. "Local control" was sacrosanct then, and the groups feared that a national test would inevitably lead to a national curriculum. Opposition diminished only after Keppel agreed to house the program in a state policy institution, the Education Commission of the States, and to report results in no smaller unit than "region." (After fears subsided, both of these conditions were abandoned.)

Keppel and the NAEP's chief developer, Ralph W. Tyler, intended the assessment to be solely descriptive. Its purpose was to provide an indicator of the nation's general education health by determining what students knew and didn't know in the same way that a health survey determines what proportion of people have tuberculosis or low body fat.

In 1983, administration of the NAEP was put out for competitive bid and awarded to the Educational Testing Service. In 1988, Congress amended the NAEP law to permit state-by-state comparisons and to create the National Assessment Governing Board (NAGB), whose task was to decide what students of a certain age *should* know. The NAEP thus became prescriptive as well as descriptive.

In its prescriptive aspect, the NAEP reports the percentage of students reaching various achievement levels—Basic, Proficient, and Advanced. The achievement levels have been roundly criticized by many, including the U.S. Government Accounting Office (1993), the National Academy of Sciences (Pellegrino, Jones, & Mitchell, 1999); and the National Academy of Education (Shepard, 1993). These critiques point out that the methods for constructing the levels are flawed, that the levels demand unreasonably high performance, and that they yield results that are not corroborated by other measures.

In spite of the criticisms, the U.S. Department of Education permitted the flawed levels to be used until something better was developed. Unfortunately, no one has ever worked on developing anything better—perhaps because the apparently low student performance indicated by the small percentage of test-takers reaching Proficient has proven too politically useful to school critics.

For instance, education reformers and politicians have lamented that only about one-third of 8th graders read at the Proficient level. On the surface, this does seem awful. Yet, if students in other nations took the NAEP, only about one-third of them would also score Proficient—even in the nations scoring highest on international reading comparisons (Rothstein, Jacobsen, & Wilder, 2006).

Additional characteristics of the NAEP make it a poor accountability tool. First, because any given student would need hours to complete the whole test, no student ever takes the entire test, nor does any school

**MEMBER SIGN IN**

Username / Customer ID / E-mail

Password

[Forgot your Username or Password?](#)

[JOIN ASCD](#) | [MEMBER BENEFITS](#)

[Register for ASCD Edge](#)

[Print This Article](#)

[Send to a Friend](#)

**Advertising**

Special Conference on  
**TEACHING & LEARNING**

**Closing the Learning Gap Once and For All**

October 29–31, 2010  
Chicago, IL

**Online Store**

**ASCD's Top 5 Books**

1. *Classroom Instruction That Works* 
2. *Enhancing Professional Practice, 2nd Edition*
3. *The Art and Science of Teaching*
4. *Building Academic Vocabulary Student Notebook*
5. *Teaching with Poverty in Mind: What Being Poor Does to Kids' Brains and What Schools Can Do About It*

## Permissions and Translations

ASCD recognizes and respects intellectual property rights and adheres to copyright law. Learn about our rights and permissions policies.

[More](#)

have all its students participate. Neither districts, nor schools, nor individual students find out how they performed (although NAEP has conducted "trial" assessments in 11 large urban districts to explore the feasibility of reporting NAEP data at the district level). This can be taken as both a strength and a weakness. Students, especially older students, likely don't take the NAEP as seriously as they take the SAT, ACT, or high-stakes state tests, so their scores may underestimate their actual achievement. On the other hand, the fact that the NAEP is not a high-stakes test means that there are almost no test-gaming efforts to artificially increase scores. (This also applies to both of the international tests discussed later.)

Claims that recent gains in NAEP trends indicate the success of No Child Left Behind have been widely disputed—in fact, it appears that NAEP increases slowed after NCLB came into existence (Fuller, Wright, Gesicki, & Kang, 2007). Such claims would not be valid in any case, because the NAEP was not designed to measure the performance of schools. The assessment attempts to cover a broad range of knowledge and skills, but it doesn't rest on any specific curriculum or theory of learning. NAEP has nothing to say about education quality at the district or school level and little to say about the smallest reported unit, the state.

## Program for International Student Assessment (PISA)

PISA has tested 15-year-olds in reading, mathematics, and science every three years since 2000. It always measures all three topics, but each administration emphasizes one. The Paris-based Organization for Economic Cooperation and Development (OECD) administers PISA to students in the 30 countries that comprise the OECD and to a similar number of partner nations. The next PISA report, which will emphasize reading, will be published in 2010.

The United States usually scores below average on PISA tests. U.S. politicians and media often uncritically accept the tests as valid and point to U.S. schools as being at fault. Such conclusions are wrong on a number of counts.

In the first place, we should question the good sense of comparing a diverse, 300-million-person nation like the United States with tiny homogeneous city-states like Hong Kong and Singapore. In addition to size, other factors complicate the issues. In Hong Kong, schools concentrate on English, Chinese, and mathematics. Proposals to introduce "liberal studies," which looks like critical thinking to me, have stirred great controversy. In Singapore, schools serve a relatively small proportion of low-income students because many low-paying jobs are done by thousands of Malaysians who enter the country each day and return home in the evening or by "guest workers," mostly from Indonesia and the Philippines, who cannot bring their spouses or families.

Those who cite PISA results to criticize the U.S. education system also ignore a number of characteristics that keep PISA from being useful for comparing the quality of schools in different nations. One problem is the fact that PISA is administered only to 15-year-olds. Because different nations start formal schooling at different ages and have different policies about students repeating a grade, such a limited snapshot can hardly tell us much about a nation's overall success in educating students.

Another problem is the design of the test items. As PISA officials write, "The assessment focuses on young people's ability to use their knowledge and skills to meet real-life challenges, rather than merely on the extent to which they have mastered a specific school curriculum" (OECD, 2005, p. 12). Because the test purportedly measures students' ability to incorporate information that they might not have learned in school, PISA's design would seem to bias it toward affluent students whose homes and families have more resources.

The University of Oslo's Svein Sjøberg (2007) points out that PISA's "requirement that the text should be more or less identical [in different countries] results in rather strange prose in many languages" (p. 14), and that the translations of at least one PISA item word-for-word from English to Norwegian rendered it nonsensical. He is quite skeptical, as am I, that questions can be rendered free of cultural bias and translated into the many languages of PISA countries and still be the "same" questions. And some of the passages for science and math questions are so long and discursive that they obviously measure reading skills as well.

PISA reports contain the nations' average score, rank, and proportion of students reaching various levels of achievement. Virtually all the media and political attention goes to the average scores and ranks. But as Hal Salzman of the Urban Institute and Lindsay Lowell of Georgetown University observe (2008), the students scoring average are not likely to become national leaders in their chosen fields. Future innovators and leaders are more likely to come from high scorers—and the United States produces more than twice as many of these as any other OECD nation does. The bad news is that the United States also produces more low scorers than any other nation except Mexico.

## Trends in International Mathematics and Science Study (TIMSS)

TIMSS comes to us from the International Association for the Evaluation of Educational Achievement in The Netherlands, but most of the technical work is conducted at Boston College. It measures selected math and science skills in grades 4 and 8 using short, fact-oriented stems and mostly multiple-choice questions.

We have been through four rounds of TIMSS: 1995, 1999, 2003, and 2007. As with PISA, politicians and the public are quick to use TIMSS results to criticize the quality of U.S. schools. In his March 2009 speech to the Hispanic Chamber of Commerce, President Obama observed, "In 8th grade math, we've fallen to 9th place." Ninth place was indeed the U.S. rank (among 46 nations) for the 2007 TIMSS administration, but in 1995, the United States ranked 28th out of 41 countries. U.S. scores as well as ranks have actually risen for 8th graders, and they have been stable for 4th graders.

[Video](#) | [Audio](#) | [Downloads](#)

[More](#)

The TIMSS developers explicitly make a causal connection between high scores and a country's economic health and claim that "there is almost universal recognition that the effectiveness of a country's educational system is a key element in establishing competitive advantage in what is an increasingly global economy" (Mullis, Martin, & Foy, 2008). Even if this were true, the question would be, Does TIMSS measure that effectiveness? The answer is no. No test can do that, because no test can measure the many complexities of an "educational system," much less a test that measures only two subjects. To get some idea of the complexity of an "educational system," I suggest that readers glance through the 100-plus goals of public education in John Goodlad's 1979 classic, *What Schools Are For*.

### The Education/Economy Fallacy

Both politicians and the media have relentlessly linked scores on national and international assessments to economic health. Release of the PISA results in 2004, for instance, led to headlines like "Economic Time Bomb" (Kronholz, 2004) and "Math + Test = Trouble for the U.S. Economy" (Chaddock, 2004).

This notion is easily refuted by the example of Japan, which led the world in test scores and economic growth in the 1980s but saw its economy sink into the Pacific in the 1990s. Throughout this period, Japanese students continued to ace tests, but Japan's economy sputtered into the new century and slipped back into recession in 2007.

It is doubtful that the ability of 4th and 8th graders to bubble in answer sheets has any connection to the economy. In fact, although educators might not want to recognize it, the current economic calamity should drive home the reality that the economic forces at play in the world dwarf the effects of education. Iceland scores high on international assessments, but in the global crisis of 2008–2009 it became an economic basket case with a national debt equal to 850 percent of its gross domestic product.

Education, by itself, does not produce jobs. There are regions of India, for example, where thousands of applicants show up for a single job requiring moderate education. The people who noticed this phenomenon worry that overeducation in the absence of job production could destabilize India (Jeffrey, Jeffery, & Jeffery, 2008). Similar worries no doubt afflict the government of China, where 33 percent of 2008 college graduates are still looking for jobs (Johnson, 2009).

Those who decry the United States' rankings on international tests should note that the Institute for Management Development (2009) and the World Economic Forum (Porter & Schwab, 2008), two organizations that rate nations on global competitiveness, rank the United States as the most competitive nation in the world—especially in the area of innovation.

In an interview, Singapore Minister of Education Tharman Shanmugaratnam acknowledged that Singapore students score well on tests but often don't fare as well as U.S. students 10 or 20 years down the road. He cited creativity, curiosity, and a sense of adventure as some of the qualities tests don't measure, adding, "These are the areas where Singapore must learn from America" (Zakaria, 2006). Sadly for American students, as Robert Sternberg (2006) observed, "The increasingly massive and far-reaching use of standardized tests is one of the most effective, if unintentional, vehicles this country has created for suppressing creativity" (p. 47).

### Blunt Instruments

This brief look at several widely recognized assessments demonstrates that none of these tests are useful for comparing the quality of schools or teachers—especially in the United States, with its diverse population, high poverty rates (by far the highest among developed nations), and wide variety of pedagogical philosophies. As former Commissioner of Education Statistics Mark Schneider said, the tests are "blunt instruments. ... A dozen factors could be behind a nation's test score" (Cavanagh & Manzo, 2009, p. 16).

Nations vary greatly in the extent of their efforts to motivate students to do well on the assessments. In Germany, where PISA has likely received more attention than in any other country, PISA-prep books can be found in airports. Observers at a school in Taiwan reported that on PISA testing day, parents gathered with their children on the school grounds urging them to do well. The students then marched into the school to the national anthem and heard a motivational speech from the principal (Sjøberg, 2007).

Can low-scoring and middle-scoring nations learn anything from the high scorers? Mostly, no. After *A Nation at Risk* appeared in 1983, Secretary of Education Terrel Bell dispatched a team to Japan. The effort came to naught, no doubt in part because Japanese schools and U.S. schools are embedded in vastly different cultures.

W. Norton Grubb and an OECD team observing schools in Finland, which ranks at the top on PISA, found some things the United States could likely adopt—for example, the interlocking system in which teachers and specialists work to head off learning problems early on. But they also noted some things we could not adopt without also adopting other large segments of the Finnish social system, such as comprehensive health care and public housing. Grubb (2007) pointed out, "The Finns take it as axiomatic that both high-quality schooling and nonschool programs are necessary for equity" (p. 109).

### A Better Way

To be related to school quality, tests must be sensitive to instruction. Most of the tests used for accountability today aren't—in fact, the manner in which they are constructed *prevents* them from being sensitive to instruction. That means that schools under the gun to raise test scores increasingly rely on strategies that get immediate, but short-lived results. Evaluation based on instruction-insensitive tests cannot help but reduce the quality of teaching (and teacher morale).

The best assessment system, but a difficult one to bring off, begins with teachers rather than with external

measures that are imposed on them. The state of Nebraska developed such a system—the School-based Teacher-led Assessment and Reporting System (STARS)—based on instruction-driven measurement as opposed to the dysfunctional, measurement-driven instruction that predominates elsewhere. (Alas, it appears to have been almost eclipsed by the statewide program installed to meet NCLB requirements.) It is that kind of system—not NAEP, TIMSS, PISA, or similar tests—that will tell us what we need to know about our schools.

#### References

- Bracey, G. W. (2009). *Education hell: Rhetoric vs. reality*. Alexandria, VA: Educational Research Service.
- Cavanagh, S., & Manzo, K. K. (2009, April 22). International exams yield less-than-clear lessons. *Education Week*, 28(29), 1, 16–17.
- Chaddock, G. R. (2004, December 7). Math + test = trouble for U.S. economy. *Christian Science Monitor*. Available: [www.csmonitor.com/2004/1207/p01s04-ussc.html](http://www.csmonitor.com/2004/1207/p01s04-ussc.html)
- Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind? *Educational Researcher*, 36(5), 268–278.
- Goodlad, J. (1979). *What schools are for*. Bloomington, IN: Phi Delta Kappa.
- Grubb, W. N. (2007). Dynamic inequality and intervention: Lessons from a small country. *Phi Delta Kappan*, 89(2), 105–114.
- Institute for Management Development. (2009). *World competitiveness yearbook*. Lausanne, Switzerland: Author.
- Jeffrey, C., Jeffery, P., & Jeffery, R. (2008). *Degrees without freedom? Masculinities and unemployment in northern India*. Palo Alto, CA: Stanford University Press.
- Johnson, I. (28 April, 2009). China faces a grad glut after boom at colleges. *Wall Street Journal*, p. A1.
- Kronholz, J. (2004, December 7). Economic time bomb: U.S. teens are among the worst at math. *Wall Street Journal*, p. B1.
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report*. Chestnut Hill, MA: Boston College.
- Obama, B. (2009, March 10). President Obama's remarks to the Hispanic Chamber of Commerce. *New York Times*. Available: [www.nytimes.com/2009/03/10/us/politics/10text-obama.html](http://www.nytimes.com/2009/03/10/us/politics/10text-obama.html)
- OECD. (2005). *PISA 2003 data analysis manual*. Paris: Author.
- Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (Eds.). (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: National Academy of Sciences.
- Porter, M. E., & Schwab, K. (2008). *The global competitiveness report 2008–2009*. Geneva, Switzerland: World Economic Forum.
- Rothstein, R., Jacobsen, R., & Wilder, T. (2006, November 29). Proficiency for all is an oxymoron. *Education Week*, 26(13), 32, 44.
- Salzman, H., & Lowell, L. (2008). Making the grade. *Nature*, 453, 28–30.
- Shepard, L. (1993). *Setting performance standards for student achievement*. Stanford, CA: National Academy of Education, Stanford University.
- Sjøberg, S. (2007). *PISA and "real life challenges": Mission impossible?* Available: <http://folk.uio.no/sveinsj/Sjoberg-PISA-book-2007.pdf>
- Sternberg, R. J. (2006, February 22). Creativity is a habit (Commentary). *Education Week*, p. 47.
- U.S. Government Accounting Office. (1993). *Educational achievement standards: NAGB's approach yields misleading interpretations* (GAO/PEMD-93-12). Washington, DC: Author.
- Zakaria, F. (2006, January 9). We all have a lot to learn. *Newsweek*. Available: [www.fareedzakaria.com/ARTICLES/newsweek/010906.html](http://www.fareedzakaria.com/ARTICLES/newsweek/010906.html)

**Gerald Bracey** is an independent researcher and writer specializing in assessment and policy issues. He resides in Port Townsend, Washington; [gbracey@q.com](mailto:gbracey@q.com). His most recent book is *Education Hell: Rhetoric vs. Reality* (Educational Research Service, 2009).

Copyright © 2009 by Association for Supervision and Curriculum Development

[Login](#) | [Register](#)

Tags:

(Log in or register for ASCD EDge (free) to add a tag.)

YOUR COMMENTS

☆☆☆☆☆ (0 Ratings)

Log in to save to your profile!

(Log in or register for ASCD EDge (free) to add a comment.)

The following are comments from our users. This is not content created by ASCD

[Contact Us](#) | [Help](#) | [Copyright Information](#) | [Privacy Policy](#) | [Permissions](#) | [Terms of Use](#) | [Advertise with Us](#) | [Sponsorships](#) | [Site Map](#)

Copyright 2010 ASCD. All Rights Reserved. P2