

THE FOLLOWING ARTICLES PRESENT DIFFERENT VIEWS ON STUDENT ASSESSMENT, OR EVALUATION, AS MOST OF US LIKE TO CALL IT. THESE ARTICLES WERE PUBLISHED IN AN EDUCATIONAL MAGAZINE ENTITLED "EDUCATIONAL LEADERSHIP", IN FEBRUARY 2003. YOU WILL NOTICE THAT THIS IS AN AMERICAN PUBLICATION, YET YOU WILL ALSO SEE THAT THERE ARE VALID ARGUMENTS PRESENTED IN BOTH OF THE ARTICLES. A LITTLE FOOD FOR THOUGHT...

The Dangers of Testing

MONTY NEILL

Educators should reject test-based reform and focus instead on formative assessment practices that encourage skilled teaching and higher-level learning.

Driving instruction with high-stakes tests will not improve schools. A large body of research demonstrates that high-stakes testing narrows curriculum and dumbs down instruction. It causes students to turn off, tune out and often drop out; induces schools to push students out; increases grade retention; propel teachers to leave; and inhibits needed improvements. In the end, high-stakes testing will hurt students – particularly those students who most desperately need better schools.

A POOR RECORD

High-stakes testing has not produced improvements in educational outcomes, even as measured by results on other tests. High school graduation tests started proliferating in the early 1980s, along with much-expanded state testing programs. By the late-1990s, high stakes for schools had become common. Over this time, research shows that:

- States that did not have high-stakes graduation exams were more likely to improve average scores on the National Assessment of Educational Progress (NAEP) than were states that did have such exams (Neill & Gaylor, 2001). At the same time, NAEP score gaps between low- and high-income students did not narrow (Barton, 2002).
- States without graduation tests were more likely than states with

such exams to show improvement or to improve at a faster rate on a variety of tests, including the NAEP, the SAT and the ACT (Amrein and Berliner, 2002).

- Data from the National Educational Longitudinal Study indicated that high-stakes testing was not associated with improved scores but was associated with higher dropout rates (Jacob, 2001)

These reports and many others show that the focus on tests has not produced the promised results. Claims of improvement typically rest on inflated scores on state exams. Texas, the model for the new Elementary and Secondary Education Act, provides a good example. As Texas Assessment of Academic Skills scores rose dramatically, the state's NAEP reading results did not change significantly. At the same time, the racial score gap in Texas widened (Neill & Gaylor, 2001). Meanwhile, the test-driven approach in Texas led to a much higher dropout rate, particularly for Latinos and African Americans (Haney 2000, 2001).

AN EQUALLY POOR OUTLOOK

If high-stakes tests have not led to improvement by now, will they ever? Two important factors suggest that even if they could in theory, they won't in practice.

Socioeconomic effects. Low-income and African American and Latino students enter school substantially less

well prepared to do academic work than their middle-income (never mind wealthy) or white or Asian peers (Lee and Burkam, 2002). In addition, the first group attends schools that are far less well prepared, in terms of teachers and physical resources, to teach them.

Rothstein (2002) points out the vast disparities in housing, health care, and other supports available to children. Those who start with less, get less, and as a result, they either fail to catch up or fall even further behind. Without major social investments in both classroom and out-of-school supports for low-income children, it is absurd to believe that more tests will enable schools to overcome the gaps in academic learning. Several reports over the past few years have purported to show that large numbers of schools serving low-income students have succeeded in overcoming the effects of poverty. More careful studies, however, have shown these claims to be at best wildly inflated and often completely misleading (Krashen 2002). Simply demanding higher scores, even with rewards and sanctions attached, will not do the job.

Teaching to the test. Any exam can only sample the curriculum that students should learn. For test results to be valid measures of real learning, students must have been taught a full curriculum. If instruction narrows to focus on the limited sample covered by the test, scores become inflated and misleading. This largely explains the

difference between results on state-specific tests and those on such neutral measures as the NAEP.

Many supporters of test-driven «reform» argue that it will at least guarantee that teachers teach the basics (which the test will supposedly cover), and will thereby initiate a positive reform process. Unfortunately, test-driven schooling often fails to provide even the basics. For example, some reading teachers teach students to read by looking over the answer options to questions attached to short reading passages and then searching the passage to find a clue for selecting the answer. Others drill phonics, but never get to comprehension. As a result, independent evaluators find that the students cannot explain what they have just read - meaning they actually cannot read (McNeil and Valenzuela, 2001).

A FLAWED CONCEPT

Testing proponents believe that focusing on a limited set of skills and facts will prepare students for further education - a theory based on profound misunderstandings of how humans learn. Shepard (2000) has demonstrated the mismatch between tests and learning, a point also developed in rich detail in several books from the National Academy of Sciences (Bransford, Brown, & Cocking, 2000; Pellegrino, Chudowsky, & Glaser, 2001). In short, humans learn best through active thinking. «Learning» while not thinking is like remembering lists of phone numbers one will never call. Memorization of facts and procedures has its place deep learning must engage the brain and spur thinking. Teaching to the test rarely accomplishes either.

Overall, state tests do not do a good job of measuring state standards, particularly higher-level thinking. Researchers at the University of Wisconsin found that state exams poorly matched state

standards and measured mostly lower-level thinking (Wisconsin Center for Educational Research, 1999).

In a high-quality education, students conduct science experiments, solve real-world math problems, write research papers, read and analyze books, make oral presentations, evaluate and synthesize information from a variety of fields, and apply their learning to new and ill-defined situations. This work provides them with both deeper substantive knowledge and higher level thinking skills. Standardized paper-and-pencil tests are poor tools for evaluating these important kinds of learning. If instruction focuses on the test, students will not learn the skills they need for success in college and beyond.

WHAT'S THE ALTERNATIVE?

If focusing instruction on large-scale, high-stakes tests will not lead to genuine improvements in our nation's schools, what should schools do in response to high-stakes testing mandates?

Educators can take some comfort from the knowledge that employing richer curriculum and instruction is likely to somewhat improve standardized test scores (Center for Collaborative Education, 2001; Institute for Education and Social Policy, 2001). In Chicago, students of teachers who used more interactive rather than exclusively didactic instructional approaches gained more than their peers on scores on the Iowa Tests of Basic Skills (Newmann, Bryk, & Nagaoka, 2001; Smith, Lee & Newmann, 2001).

One can also assess higher-level attributes. Teachers do it all the time. Studies of a group of small schools in New York City and similar schools in Boston described these successful schools as having high standards and rigorous as-

essments (Center for Collaborative Education, 2001; Institute for Education and Social Policy, 2001). The high schools often require students to present and defend their work in a number of subject areas to earn a diploma. A committee that typically includes outside experts reviews the work. As Meier (2002) says, these high standards do not entail the standardization or drill-and-kill imposed through high-stakes exams.

These high-quality schools in New York and Boston generally do not focus on teaching to the test; indeed, many are struggling to avoid being coerced into becoming test-prep programs. They demonstrate their success by far more than test scores. Deborah Meier's Central Park East Secondary School, for example, prepares low-income students not only to enroll in college but to succeed there - a far more important goal than merely increasing their test scores (Meier, 1995). Indeed, Meier has often pointed out that although her students learned to succeed, their standardized test scores did not necessarily show substantial gains. These schools succeed by teaching for deeper, important learning.

Some have argued that such schools cannot be replicated. They presume that test-driven improvements can. However, researchers from the Dana Center went hunting for successful Texas school districts that had improved scores and closed racial score gaps (Skrla, Scheurich, & Johnson, 2000). Using modest criteria that were overly dependent on TAAS scores - enrollement of 5,00 or more students; high poverty levels; and 50 percent of the high-poverty schools in the district categorized as Recognized or Exemplary on the basis of their test scores - they studied data from all Texas districts. They found 11 districts to examine closely, which they reduced in the end to 4 suc-

cessful districts. In other words, after nearly a decade in which the state had focused on high-stakes testing, sympathetic researchers found only a few districts in the state to meet their narrow criteria for success.

Unlike high-stakes testing, which undermines good schools and prevents real improvement, better forms of assessment can play a powerful role in improving schools. Most important is assessment that provides meaningful and usable feedback to students and engages them in self-evaluation. A research summary of more than 250 articles and chapters by Black and William (1998) found that formative assessment can contribute more to improving outcomes (primarily as measured by test scores) than any other school-based factor, benefiting low achievers more than high achievers. In other words, it helps raise everyone's achievement while also closing learning gaps – which test-driven reform has not done. This approach, however, requires treating teachers as professionals, improving professional development, and spending more money.

Rather than chasing the illusion that test-driven change will produce significantly improved learning, policymakers need to shift attention to practices and models that emphasize serious thinking and skilled teaching. If we do less, we will consign far too many students to a continued second-class education. At the moment, few policymakers recognize this. It appears that, to get on the path of genuine improvement, educators and parents will have to join together to beat back test-based «reform.»

Monty Neill, Ed.D., is Executive Director of the National Center for Fair & Open Testing (FairTest) in Cambridge, MA.

Bibliography

Additional material can be found at www.fairtest.org

Achieve, Inc. (2001, October). Measuring up: A report on education standards and assessments for Massachusetts. <http://www.achieve.org/achieve.nsf/StandardForm3?openform&parentunid=FCCF5B09747347D285256AE700733993>

Amrein, A. & Berliner, D. (2002, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives* 10(18). <http://epaa.asu.edu/epaa/v10n18/>

Barton, P.E. (2002, January). Raising achievement and reducing gaps. Washington, DC: National Education Goals Panel.

Black, P. & William, D. (1998, October). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 139-148

CARE (Coalition for Authentic Reform in Education). (1999b, December). Call for an authentic statewide assessment system. <http://www.fairtest.org/care/accountability.html>

CARE. (1999a, 2000). Reviews of MCAS tests. Cambridge, MA: unpublished; available from FairTest.

Center for Collaborative Education (CCE). (2001). How Are Boston pilot schools faring? An analysis of student demographics, engagement, and performance. http://www.ccebos.org/quant_report_final.pdf

Haney, W. (2000, Aug 19). The myth of the Texas miracle in education. *Education Policy Analysis Archives* 8(41). <http://epaa.asu.edu/epaa/v8n41/>

Haney, W. (In press). Revisiting the myth of the Texas miracle in education.

Institute for Education and Social Policy. (2001, December). Final report of the evaluation for New York Networks for School Renewal. New York: Author.

Jacob, B. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis* (23)2, 99-121.

Kohn, A. (2000). The case against standardized testing: Raising the scores, ruining the schools. Portsmouth, NH: Heinemann.

Krashen, S. (2002, February). Poverty has a power-

ful impact on educational attainment, or, don't trust Ed Trust; Chicago, Substance; on web at <http://www.fairtest.org/k12/krashen%20report.html>

Lee, V. E. & Burkham, D.T. (2002). Inequality at the starting gate. Washington, DC: Economic Policy Institute.

McNeil, L. & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In Orfield and Kornhaber (2001).

Meier, D. (2002). In schools we trust. Boston: Beacon Press.

Meier, D. (1995). The power of their ideas. Boston: Beacon Press.

Neill, M., with Gaylor, K. (2001). Do high-stakes graduation tests improve learning outcomes: Using state-level NAEP data to evaluate the effects of mandatory graduation tests. In Orfield and Kornhaber (2001).

Newmann, F. M., Bryk, A.S. & Nagaoka, J.K. (2001, January). Authentic intellectual work and standardized tests: Conflict or coexistence? Chicago: Chicago Consortium on School Research.

Orfield, G. & Kornhaber, M.L. (2001). Raising standards or raising barriers? Inequality and high-stakes testing in public education. New York: Century Foundation.

Rothstein, R. (2002). Out of balance: Our understanding of how schools affect society and how society affects schools. Chicago: Spencer Foundation.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher* 29(7), 4-14.

Skrla, L., Scheurich, J.J. & Johnson, Jr., J.H. (2000, September). Equity-driven achievement-focused school districts. Austin: Charles A. Dana Center, University of Texas.

Smith, J.B., Lee, V.E. & Newmann, F.E. (2001, January). Instruction and achievement in Chicago elementary schools. Chicago: Chicago Consortium on School Research.

WCER. (1999, Fall). Are state-level standards and assessments aligned? WCER Highlights, 1-3. Madison: Wisconsin Center for Educational Research.