

FairTest

National Center for Fair & Open Testing

How Standardized Testing Damages Education

How do schools use standardized tests?

The No Child Left Behind (NCLB) era has seen an unprecedented expansion of standardized testing and test misuse. Despite ample evidence of the flaws, biases and inaccuracies of standardized exams, NCLB and related state and federal policies, such as Race to the Top (RTTT) and the NCLB waivers, have pressured schools to use tests to measure student learning, achievement gaps, and teacher and school quality, and to impose sanctions based on test scores. This is on top of using tests to determine if children are ready for school; track them into instructional levels; diagnose learning disabilities, retardation and other handicaps; and decide whether to promote, retain in grade, or graduate. School systems also use tests to guide and control curriculum content and teaching.

Aren't these valid uses of test scores?

Measurement experts agree that no test is good enough to serve as the sole or primary basis for any of these important educational decisions. A nine-year [study](#) by the National Research Council (2011) concluded that the emphasis on testing yielded little learning progress but caused significant harm. *NCLB* demonstrated what happens when tests are misused. Negative consequences include narrowing the curriculum, teaching to the test, pushing students out of school, driving teachers out of the profession, and undermining student engagement and school climate. *High school graduation tests*, used by 25 states, disproportionately penalize low-income and minority students, along with English language learners and the disabled. They do not promote the knowledge, skills and habits needed for success in college or skilled work. *Tracking* generally hurts slower students but does not help more advanced students. Too often, the assumption is that low-scoring students need low-level remediation rather than enrichment, challenge and support. *Retention in grade*, flunking or holding a student back, is almost always academically and emotionally harmful. It generally does not lead to sustained academic improvement, lowers student self-esteem, and leads to dropping out. *Screening* and *readiness* tests are frequently inaccurate and can lead to misdiagnosis of student learning needs.

Who is most often hurt by these practices?

Students from low-income and minority-group backgrounds, English language learners, and students with disabilities, are more likely to be denied diplomas, retained in grade, placed in a lower track, or unnecessarily put in remedial education programs. They are more likely to receive a “dumbed-down” curriculum, based heavily on rote drill and test practice. This ensures they will fall further and further behind their peers. Many drop out, some ending up in the “school-to-prison pipeline.” On the other hand, children from white, middle and upper income backgrounds are more likely to be placed in “gifted and talented” or college preparatory programs where they are challenged to read, explore, investigate, think and progress rapidly.

How do tests control curriculum and instruction?

In many districts, standardized exam results have become the single most important indicator of school performance. As a result, teachers and administrators feel enormous pressure to ensure that test scores

P.O. Box 300204, Jamaica Plain, MA 02108
fairtest@fairtest.org 617-477-9792 <http://fairtest.org>

consistently rise. Schools narrow and manipulate the curriculum to match the test, while teachers tend to cover only what is likely to be on the next exam. Methods of teaching conform to the multiple-choice format. Education increasingly resembles test prep. It is easy to see why this could happen in low-scoring districts. But some high-scoring schools and districts, striving to keep their top rank, also succumb. The pressure is so great that a growing number of administrators and teachers have engaged in various kinds of cheating to boost scores.

Are test results a good way to measure teacher quality?

Student tests cannot reliably, validly and fairly be used to judge educators. Researchers looked at popular value-added methods of teacher evaluation and found them fraught with errors and unreliable. One researcher concluded that “a teacher’s performance evaluation may pivot on what amounts to a statistical roll of the dice.” The negative consequences for teaching and learning will only intensify when educators are judged “in significant part” by student test scores, which is a requirement in both RTTT and NCLB waivers. Knowledge of the arbitrary and inaccurate consequences will deter some strong young candidates from becoming teachers or principals, and drive good, experienced educators away from working in the most high-need schools.

Don’t standardized tests provide accountability?

No. Tests that measure as little and as poorly as multiple-choice exams cannot provide meaningful accountability. Instead of being accountable to parents, community, teachers and students, schools become “accountable” to an unregulated testing industry. “Score inflation” results when narrow test preparation replaces more in-depth and comprehensive instruction. Not only do students get an inferior education, but the public gets the mistaken impression that education is improving.

If we do not use standardized tests, how will we know how students and programs are doing?

Standardized tests can be one part of a comprehensive assessment system. However, they offer just a small piece of the picture. Better methods of evaluating student needs and progress already exist. Careful observation and documentation of student work and behaviors by trained teachers is more helpful than a one-time test. Assessment based on student performance on real learning tasks is more useful and accurate for measuring achievement – and provides more information for teaching – than multiple-choice achievement tests.

Are other methods of assessment reliable?

Trained teams of judges can be used to rate performance in many academic areas. Studies have shown that, with training and clear guidance, the level of agreement among judges (“inter-rater reliability”) is high. At the Olympic Games, for example, gymnasts and divers are rated by panels of judges. Advanced Placement essays and its Studio Art assessment are scored entirely by teams of trained educators. Independent evaluators have consistently judged collections of student classroom work (portfolios and learning records). A process of sampling from classroom-based evidence can provide richer information, be adequately reliable, and help stop teaching to the test. As with multiple-choice exams, safeguards are needed to ensure that race, class, gender, linguistic or other cultural biases do not affect evaluation.

How do other nations evaluate their students?

The U.S. is the only economically advanced nation to rely heavily on multiple-choice tests. Other nations use performance-based assessment to evaluate students on the basis of real work such as essays, projects and activities. Ironically, because these nations do not focus on teaching to multiple-choice and short-answer tests, they score higher on international exams.