

**Public Hearing  
Regents Learning Standards  
and High School Graduation Requirements**

**NY State Standing Committee on Education  
NYC, October 15, 2003**

**Testimony submitted by Dr. D. Monty Neill  
Executive Director  
National Center for Fair & Open Testing (FairTest)**

---

**High Stakes Testing and Better Alternatives: A National Perspective**

Thank you for this opportunity to speak with you.

My name is Monty Neill. I am the Executive Director of the National Center for Fair & Open Testing, also known as FairTest, based in Cambridge, Massachusetts. We are the nation's leading assessment reform organization. Our mission is to end the misuses and flaws of standardized testing and ensure that evaluation of students, teachers and schools is fair, open, valid and educationally beneficial.

FairTest has taken up this task because the misuse of standardized testing causes grievous harm to education, and particularly to our most vulnerable students - low income, minority group, limited English proficient, and special needs - who are often described as those who will most benefit from intensified high-stakes testing. There are documented better ways to assess students and establish accountability than by relying on standardized tests such as the Regents exams.

New York state now faces a crisis of high-stakes testing that is similar to crises emerging across the nation. The crisis has been caused by the use of standardized tests as mandatory hurdles to be leaped by students before they can obtain a high-school diploma or be promoted to the next grade, and by the use of those exams as the near-sole basis for evaluating schools and districts.

FairTest concludes that the evidence of the past decade of high-stakes testing from across the nation shows that there are multiple and unavoidable factors which guarantee there will be continuing major problems should high-stakes use of standardized tests continue. In particular, high-stakes testing undermines good education where it does exist and inhibits progress toward improving education. Such problems are inevitable with high stakes standardized testing and so cannot be solved by tinkering with the tests. Assessment is, despite the problems with standardized tests, a critically important element of school reform. Fortunately, there are other, better means to assess student learning.

In this testimony, I will quickly describe the basic problems. Far more information and evidence is available, for example on the FairTest website ([www.fairtest.org](http://www.fairtest.org)) and in many journals and

articles, of which I will reference only a few. I will leave to other people to provide testimony on the particulars of the problems caused by testing in New York state in order to use my time to provide a national perspective to help illuminate the issues faced by New York State.

## **I. Problems with the technology of testing**

Clearly, many children pass through schools and leave with only a weak education. These children are usually low-income and disproportionately children of color, recent immigrants, or have disabilities. Equally clearly, and as New York's highest court has affirmed, states have a moral and legal responsibility to change this situation. States must ensure that all children receive an education that enables them to participate fully as citizens in a democracy, to be lifelong learners, and to succeed in further education and well-paid employment.

Across the nation, the primary approach to addressing the problem of inadequate education has been through the use of standards and tests. Standards describe what students should know and be able to do, and they establish how much of the specified knowledge and skills is sufficient to warrant awarding a high school diploma. Assessments are intended to measure whether students have met the standards. In 19 states, a single standardized exam is used to make that determination. With few exceptions, in those states, students who do not pass one or more tests do not obtain a diploma regardless of their academic record in high school.

The critical underlying claims are first that standards and tests will cause schools to focus on what is most important, with the result that students will be better educated; and second that testing can fairly distinguish between those whose academic achievement says they deserve a diploma and those who do not. Numerous flaws have been identified that disprove the hypothesis that high-stakes testing induces high-quality schooling. They include:

- Many states have far too many standards that are too complex or too vague, as well as often developmentally inappropriate. The result is that teachers do not know what to focus on. Tests have therefore served the purpose of defining what is most important. This means that, in effect, what is most important is decided by what can most readily be measured on a few-hour exam that relies on multiple-choice and short answer questions. The necessary corollary is that much of the content of reasonable standards is not measured by state exams. A series of independent studies have confirmed the inadequacy of actually existing tests to measure the standards (Achieve, 2001; WCER, 1999).

- Standardized tests are weak measures for many important areas of learning. Many of the tests themselves suffer from poor construction and errors. (I will address this point in more detail below.)

- Some schools find ways to improperly narrow instruction and to push low-scoring students out. (I will address this point in more detail below.)

- Many schools are inadequately funded. High-stakes testing holds students and educators responsible for conditions which they do not control. Until all students receive a fair opportunity to learn and all schools are given the necessary financial and other support to provide such opportunities, high stakes accountability unfairly punishes the victims of inadequate opportunity.

***1A - Limits to testing: Why measurement experts caution against high-stakes uses***

This use of a standardized test to deny graduation regardless of other evidence of student learning has been warned against by the measurement profession itself:

The *Standards for Educational and Psychological Testing* promulgated by the American Education Research Association (AERA), the American Psychological Association, and the National Council on Measurement in Education states at Standard 13.7: “In educational settings, a decision or characterization that will have a major impact on a student should not be made on the basis of a single test score” (AERA, 1999).

Not graduating is surely a “major impact.” Elaboration on this issue in the *Standards* and in separate statements by the AERA (AERA, 2000) make clear that this statement means that this standard is not met by providing multiple opportunities to take a test. Rather, in the case of graduation, a student who does not pass an exam should be able to use other evidence to show that she or he has met the graduation standards.

The reasons for this warning from the measurement profession include:

1. The error margin on any standardized test is too high to comfortably make a major decision based solely on the results of the one exam. Students will unfairly be denied graduation and schools will unfairly be deemed “in need of improvement” due to measurement error rather than to actual achievement. Multiple opportunities to take an exam reduces the chance that measurement error is to blame, but cannot eliminate that chance or reduce it to an acceptable level.
2. Any test provides one or two means of measuring, most commonly multiple-choice and short-answer. Yet, many students do not perform well on these types of measures, or under standardized exam conditions in general, and thus are not provided a fair opportunity to demonstrate what they in fact know.
3. In education, much evidence of student achievement is developed, most of it in the classroom. Thus, tests need not stand alone as a criterion for graduation. Some argue that this problem is addressed by the requirement that students must also pass their courses. However, so long as the test represents a necessary hurdle students must leap, regardless of grades or other relevant evidence, the test is being misused.

### ***1B - Inadequate test quality***

The second major problem with the reliance on tests is that test quality is itself not reliable. Dozens of errors in standardized tests have been reported in the last few years. In a recent report, researchers from Boston College documented many of them. (Rhoades and Madaus, 2003). These include denial of a diploma based on flawed exams and forced attendance in summer school. Study of the errors reveals that errors are inevitable, that many are caught only by chance or by unusual persistence by a few people, and that many errors are almost certainly not caught. Neither New York nor any other state should attach high stakes for students to such chronically flawed measures.

Exams also can be too difficult. The proclivity of subject matter experts to believe that all people

should know and be tested on a great deal of detail in their subject is compounded by the use of testing technology that builds in a process of excluding prospective test items that most students actually know. The technology was developed for use with norm-referenced tests that are intended to sort and rank. However, the technology is employed for use with most supposedly criterion-referenced tests. This procedure guarantees many students will fail (Haney, 2002).

Thus, one major set of problems that guarantees crisis is that tests are limited, narrow tools, rife with errors, that cannot fairly be used for high-stakes decision-making about individuals or schools.

## **II. Over-reliance on testing undermines educational quality**

The second major set of problems is that use of the tests undermines educational quality. Much of this problem is inherent in the limitations of the tools and cannot be solved by tinkering with the tests.

Much of value that students should learn and that is often included in state standards cannot be tested with any paper-and-pencil test of a few hours duration. In a high quality education, students conduct science experiments, solve real-world math problems, write research papers, read novels and stories and analyze them, make oral presentations, evaluate and synthesize information from a variety of fields, and apply their learning to new and ill-defined situations. It is rather self-evident that standardized tests do not measure these forms of knowledge and skill. At best they indirectly measure a limited slice of the knowledge and skills that make up the richer, more comprehensive subject areas.

Standardized exams also offer few genuine opportunities to display the attributes of higher-order thinking, such as analysis, synthesis, evaluation, and creativity. Higher order thinking is encouraged and revealed by in-depth, open ended and extended work, not by one-shot tests (Shepard, 2000; Pellegrino and Chudowsky, 2003; National Research Council, 2002; Case Against). Not only does the public therefore not know whether students have attained these valuable attributes, but also, if instruction focuses on the test, students will not learn these skills, which are needed for success in college and life.

It is important also to understand that the testing technology in widespread use relies on archaic views of how humans learn and think (National Research Council, 2002; Pellegrino and Chudowsky, 2003; Shepard, 2000). In the past few decades, there has been an explosion of knowledge in cognitive and developmental sciences. The very limited theoretical constructs underlying the development of standardized testing have been demonstrated false or superceded by the new knowledge.

The problem is not only inadequate measurement tools that fail to provide sufficient high-quality useful information. It is that such inadequate measures which do not correspond to how people learn foster methods of teaching that fail to engage students, fail to match the general and varied ways in which they actually learn, and therefore promote and support ineffective educational practices. Research conducted on Chicago students found that teachers who utilized challenging rather than remedial curriculum and who engaged in a mix of interactive and didactic instruction

rather than primarily didactic instruction ensured substantially greater learning gains even as measured by the Iowa Tests of Basic Skills (Newmann, Bryk & Nagaoka, 2001; Smith, Lee & Newmann, 2001). The Iowas do not even measure the richer forms of learning that are gained through challenging and interactive instruction. The greater learning gains happened for low as well as higher achievers, and students of all racial-ethnic groups.

When standardized tests are the primary factor in accountability, the often overwhelming pressure is for schools to use the tests to define curriculum and focus instruction. What is not tested is not taught, and what is taught does not include higher-order learning. How the subject is tested becomes a model for how to teach the subject. At the extreme, school becomes a test prep program – and this extreme already exists in too many schools across the nation.

The evidence that focusing on tests fails to induce high-quality schooling or learning comes from many sources. Texas has been touted as the great success story of high-stakes testing, largely based on rapid gains on its TAAS tests and in particular on the closing of the score gaps between African American and Latino students on one side and white students on the other. However, when the new TAKS test was introduced in Texas, it revealed that the score gap remained (Peabody, 2003). Closing the gap was purely an artifact of narrowly teaching to the test, of turning instruction for African American and Latino students into mere test preparation. We at FairTest has heard the stories of how reading instruction became reading short, decontextualized segments, followed by multiple-choice questions, and how students were taught to “read” by scanning test answer options and looking for key matching words in the preceding text.

Some have pointed to results on the National Assessment of Educational Progress (NAEP) to tout high-stakes testing. However, NAEP reading results show no closure of the racial gap in Texas (Schrag, 2000). (In New York, it is true that grade four reading results rose – but they have not increased at grade 8. Of course, it makes little sense to attribute gains on grade 4 NAEP in one snapshot year to the Regents exams.) NAEP itself is a limited, though well-crafted test, which relies entirely on multiple-choice and short-answer questions. Thus, it too cannot measure many important educational goals.

In line with the claim that high-stakes testing will improve schooling for the disadvantaged is the claim that such testing will lead to instruction that better prepares students for college. However, the Texas higher education system has reported an increase in the need for remediation since the advent of the tests, a need that cannot be explained by the rather modest growth of the system (Caballero, 2002).

More generally, research by professors Audrey Amrein and David Berliner (2002) of Arizona State showed that gains states report on their own high-stakes tests do not correlate with results from other exams, such as the National Assessment of Educational Progress or SAT and ACT college entrance tests. In states with graduation tests, scores on these other exams often declined or grew less quickly relative to the nation as a whole.

Lastly, for this section, it is vital to understand that the increase in testing has been extensively

associated with increases in the dropout rate (Amrein & Berliner, 2002; Jacob, 2001). At a minimum, across the nation, education reform programs that have focused on high stakes testing have failed to lower the dropout rate. As Professor Walt Haney (2003) has shown, the national dropout rate is unacceptably high and New York State suffers one of the very worst dropout rates.

It is of course possible to give a standardized test and not let it control curriculum and instruction. However, if the tests have high stakes attached, this can result in a school putting itself at risk for producing lower test scores. Worse, they can put students at risk of not graduating even if they have solid academic achievement and are capable of enrolling and succeeding in college. Therefore, schools are compelled to focus on testable content to the detriment of better academics.

What I have argued thus far is that the crisis of standardized testing resides in the misuse of narrow and flawed instruments for making high-stakes decisions and in the damage to curriculum and instruction that follows in the wake of high stakes.

Before turning to solutions, I want to address one further issue: the question of public opinion and support.

### **III. Public Opinion**

Numerous surveys have been conducted exploring support for standards and tests among parents, educators and the general public. We think it is fair to conclude the following (Bostrom, 2003; Rose & Gallup, 2003):

- 1) There is wide support among all groups for the idea of standards.
- 2) There is significant support for the use of some standardized testing.
- 3) There is significant opposition, from more than half up to two-thirds of respondents, depending on the poll, to the use of high-stakes tests for making decisions about students or schools.
- 4) A sizeable majority oppose teaching to the test.

Concerns over the impact of high-stakes testing is greatest among educators, but those concerns seem to be shared by a majority of parents and citizens, given that the same findings occur in differently worded polls conducted by different polling groups over the past several years.

FairTest concludes that changing from a one-size-fits-all high-stakes approach to the use of multiple measures would represent the will of the majority.

### **IV. Better alternatives**

A testing program that inadequately meets educational needs, promotes instruction that does not match how humans learn, cannot be shown to meaningfully increase student learning or future success, and lowers or fails to improve the graduation rate, should be viewed as a failure, a disaster. In the face of such a major mistake, it is hard for policymakers to admit the error and make the fundamental changes that are needed. But you must display the necessary courage for the sake of the children and the future of the state.

Testing proponents often claim we must choose between leaving things as they are and relying on high-stakes tests. The evidence shows that relying on high-stakes testing is in fact often worse than doing nothing. However, states are not constrained to choose between two poor options. Other things can be done.

First, a great deal of knowledge has been developed about how to engage in high-quality assessment. The same research that has shown how standardized testing fails to adequately reflect what we now know about how humans learn has begun to lead to development of assessments that do reflect such knowledge (Shepard, 2000; Pellegrino and Chudowsky, 2003; National Research Council, 2002). New electronic technologies are increasingly making feasible the compilation and analysis of much more detailed classroom-based assessment information. Professional development practices have found that teachers can indeed learn to use sophisticated assessments to inform their teaching.

This leads not only to improved evaluation of student achievement, but more importantly to improved instructional practices and improved student learning. The landmark study by Paul Black and Dylan William (1998) found that high-quality formative assessment – assessment used by teachers and students in the classroom to shape instruction – results in learning gains that have never come close to being matched by programs which focus on teaching to standardized tests. Significantly, lower-achieving students made the greatest gains, thereby closing the gap with students who started out ahead.

In addition to a growing research base, several states have taken steps toward establishing assessment processes that simultaneously strengthen standards and allow local schools and districts substantial flexibility in how they assess student achievement. For example:

- Rhode Island has established new graduation requirements in which students will have to demonstrate they have met state standards (FairTest, 2002). This proficiency must be demonstrated through at least two of the following: departmental end of course exams, a Certificate of Initial Mastery, portfolios, extended ‘capstone’ projects, public exhibitions, and the use of technological tools. Schools may use results from the state assessment as part of the graduation determination, but if they do the exams may not count for “more than 10 percent of all the weighted factors contributing to promotion or graduation.” The state Education Commissioner will approve local graduation requirements.

- Maine is developing a system which includes state standardized tests in three grades that do not carry high stakes, and local assessments in all grades that will be used to evaluate and improve

local schools and be used to determine readiness to graduate (FairTest, 2002, Spring). (The state website provides a great deal of information about this program;  
<http://www.state.me.us/education/homepage.htm>

- Nebraska has developed a system that relies on local assessments for all accountability uses (FairTest 2002, Fall; Roschewsk, 2003). It does not require the use of such assessments to determine graduation. Local districts must adopt the state standards or develop their own which the state must approve. The state has a set of criteria which local assessments must meet. Independent reviews by the highly respected Buros Institute for Mental Measurement have found that the local assessments do assess the standards, are fair and unbiased, and do meet technical requirements for valid and reliable measurement. (See state website at:  
<http://www.nde.state.ne.us>.)

In all three cases, local assessments can include compilations of students' schoolwork, for example, through the evaluation of portfolios.

In the 1990s, a great deal of work was done by a commission headed by Linda Darling-Hammond to design an accountability system that would provide valuable data for improving schools while avoiding the pitfalls of relying on inadequate tests, applying high-stakes exams, and narrowing and dumbing-down education. It is nothing less than a tragedy that then newly-arrived Commissioner Richard Mills chose to discard all that work in favor of an overly-centralized, absurdly bureaucratized, dangerously narrow exam system.

The New York State Legislature should examine the evidence. FairTest believes that when it does so, it will come to the conclusion that the state should shift from a test-based program to one based on more comprehensive assessments that can include limited and low-stakes standardized testing. The Legislature should then take steps to curtail reliance on the Regents exams. It should ensure the design of a very different system of assessment and accountability, building on the work of Dr. Darling-Hammond's commission, the extensive work that has been done in assessment research and development, and the successes in other states. And it should allow qualifying schools to use alternative measure to determine graduation in the interim period before a new system is designed.

## References

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>

American Educational Research Association (AERA). (2000, July). Position Statement Concerning High-Stakes Testing in PreK-12 Education.

<http://www.aera.net/about/policy/stakes.htm>

American Educational Research Association (AERA), American Psychological Association,

National Council on Measurement in Education (1999) *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

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/>

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

Bostrom, M. (2003, April). Fulfilling the promise of No Child Left Behind: A metaanalysis of attitudes toward public education. Severna Park, MD, Public Knowledge LLC.

Caballero, P. (2002, Dec. 22). Students' college readiness is lagging. *Dallas-Ft. Worth Star-Telegram*. <http://www.dfw.com/mld/starttelegram/news/local/47.../printstory.js>) accessed Jan. 23, 2002.

Case Against High-Stakes Testing <http://www.fairtest.org/arn/caseagainst.html> is a section of the FairTest website that contains voluminous material on the problems associated with high-stakes standardized testing.

*FairTest Examiner*. (2002, Fall). RI adds varied graduation requirements; p14.

*FairTest Examiner*. (2002, Spring). Nebraska and Maine assessment models; pp 16, 12-13.

Haney, W. (2003, Sept. 23). Attrition of students from New York schools. Invited testimony at "Regents Learning Standards and High School Graduation Requirements" before NY Senate Standing Committee on Education.

Haney, W. (2002, July 19). Ensuring failure: How a state's achievement test may be designed to do just that. *Education Week*, 21:42, pp. 546, 58.

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/>

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

National Research Council. (2001). *Knowing what students know: The Science and Design of Educational Assessment*. Committee on the Foundations of Assessment, Board on Testing and Assessment. Washington, DC: National Academy Press.

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.

Peabody, Z. (2003, July 24). Test shows HISD 'achievement gap' back. *Houston Chronicle*.  
<http://www.chron.com/cs/CDA/story.hts/metropolitan/2009440>

Pellegrino, J. W., and Chudowsky, N. (2003). The foundations of assessment. *Measurement: Interdisciplinary Research and Perspectives*, 1(2), 103-148.

Rhoades, K., and Madaus, G. (2003, May). Errors in Standardized Tests: A Systemic Problem. National Board on Educational Testing and Public Policy.  
<http://www.bc.edu/research/nbetpp/statements/M1N4.pdf>

Roschewsk, P. (2003, March). Nebraska STARS line up. *Phi Delta Kappan*, 84(7), pp. 517-520.

Rose, L. C., and Gallup, A.M. (2003, Sept.). The 35th annual Phi Delta Kappa/Gallup Poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, pp. 41-56.

Schrag, P. (2000, January). Too good to be true," *The American Prospect*, 11: 4.  
<http://www.prospect.org/print/V11/4/schrag-p.html>

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

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.

---

**Dr. D. Monty Neill**  
Executive Director  
National Center for Fair & Open Testing (FairTest)  
342 Broadway  
Cambridge, MA 02139  
(617) 864-4810  
email: [Monty@fairtest.org](mailto:Monty@fairtest.org)