

Test Anxiety: Are Students Failing Tests –Or Are Tests Failing Students?

©David Sadker and Karen Zittleman

Please do not reprint without authors' permission

(This article was originally published in *Phi Delta Kappan*, June 2004)

Ah, the beauty, simplicity, and allure of standards! What can be more rationale (let's face it, even honorable) than setting educational standards? Standards are like guiding stars that enable us to navigate a successful academic voyage. Students can see what they need to learn, educators can alter course to improve performance, and the public gets a report card on just how well their educational ship of state is performing. It could be a comforting picture.

While the standards current is becalming, high stakes tests are the menacing clouds on the horizon that portend an ominous change in the weather. The second Bush administration's *No Child Left Behind Act (NCLB)* is all about testing. Every third- through eighth-grader in U.S. public schools will be tested in reading and math (with science to follow in 2008) and for students, teachers and schools, poor performance means severe penalties.

We know too well the alphabet soup of high stakes tests - IQ, SAT, ACT, and GRE – scores that brand us for life creating hopeful futures or a lost opportunities. But *NCLB* raises even these high stakes. In this article, we shall detail the pitfalls of the current reign of “test or consequences” that lead critics like Gerald Bracey to call the misguided venture *The No Child Left Act*. We will describe seven reasons why *NCLB* will not work, and offer alternative standards that empower teachers and students and reflect more meaningful learning.

Problem 1: *Knowing What's Worth Knowing*

Sounds like *Philosophy 101: What Is Knowledge?* Although testing sounds logical, knowing what's worth knowing is not always so clear. Try your hand at the following test: ¹

Directions: The penmanship of applicants will be graded from the manuscripts.

GRAMMAR

1. Give nine rules for the use of Capital Letters.
2. Define Verse, Stanza, and Paragraph.

ARITHMETIC

1. District No. 33 has a valuation of \$35,000. What is the necessary levy to carry on a school seven months at \$50 per month, and have \$104 for incidentals?
2. What is the cost of a square farm at \$15 per acre, the distance around which is 640 rods?

U.S. HISTORY

1. Give the epochs into which U.S. History is divided.
2. Describe three of the most prominent battles of the Rebellion.

ORTHOGRAPHY

1. What are the following, and give examples of each: Trigraph, subvocals, diphthong, cognate letters, linguals?
2. Give four substitutes for caret “u”.

GEOGRAPHY

1. Name and describe the following: Monrovia, Odessa, Denver, Manitoba, Heela, Yukon, St. Helena, Juan Fernandez, Aspinwall, and Orinoco.
2. Name all the republics of Europe and give the capital of each.

Does the test seem a bit challenging, perhaps even unfair, or did you breeze through? Does poor performance mean our schools are failing and need to be replaced? Are you asking what we too infrequently ask: Who thought such questions are important? Well, not us. We borrowed these test questions from Saline County School District, Kansas. Students needed to pass this test in order to graduate from eighth grade...in 1895. The Saline test reminds us that what is essential knowledge at one time may eventually be irrelevant or erroneous. What is considered appropriate for twelve-year-olds in one context can stump professors and principals in another. Those in power determine what is worth knowing, and time often proves them wrong.

Problem 2: *Teaching to the Test: Is that the Point?*

Sorting out what is worth knowing is not the only tricky proposition; distinguishing test scores from learning is another. Evidence is mounting that much of what is going on in today’s classrooms is not so much learning as test-prep, and the school curriculum is being reduced accordingly.² In a nationwide poll of more than 1,000 public school teachers, *Education Week* found that two-thirds felt their states had become too focused on state tests.³ Entire subject areas — music, art, social studies and foreign languages — are de-emphasized because they are not tested. As one teacher put it, “At our school, third and fourth grade teachers are told not to teach social studies and science until March.”⁴ In fact, 85 percent of teachers report that their school gives less attention to subjects that are not on the state test, and 79% acknowledge that they spend time instructing students in test-taking skills.⁵ In Texas, James V. Hoffman and his colleagues found that reading teachers and supervisors “often” or “always” teach test taking skills, have students practice with tests from prior years, use commercial test preparation materials, give general tips on how to take tests, and demonstrate how to mark an answer sheet correctly.⁶

Does direct teaching of test-like questions help students master the subjects? Although test preparation boosts scores, it does not necessarily produce real understanding. Consider these findings:

- A study of 18 states with high-stakes testing compared trends in state test scores with long-term trends on other standardized tests.⁷ In more than half of these states, performance went *down* on the ACT, SAT, and the math test of the National Assessment of Educational Progress (NAEP). Specific test preparation and whether disabled or limited English speaking students were allowed to take the test determined test performance.
- Three-quarters of fourth-grade teachers and the majority of principals surveyed in Washington State believed that better test preparation, rather than increased student knowledge, was responsible for most test score gains.⁸

- In Kentucky’s state assessment, scores went up on test items that were reused, then dropped when new items were introduced. This discrepancy between new and reused items was larger in schools that had greater overall test score gains, a relationship that suggests students were coached on reused items.⁹

Problem 3: Standardized Tests Results Do Not Equal Educational Accountability

Standardized tests can be quite appealing: they are relatively inexpensive to administer, can be mandated with comparative ease and rapidly implemented, and can deliver measurable, visible results. But this new federal testing policy is not based on accumulated educational wisdom and research, or on clear design principles that have some basis in school practice. In the early stages of the testing movement, assessment had an expansive view of performance that included, in addition to tests, portfolios and formal exhibitions of students' work, student-initiated projects, and teachers' evaluations of their students. All this seems to have been forgotten, disdained, or eliminated as not cost effective, and what remains is a testing tyranny.

More than half of the states now have laws that use state tests to determine whether students graduate or are promoted to the next grade.¹⁰ Yet, retention in grade is the single strongest predictor of which students will drop out—stronger even than parental income or mother’s education level. The National Research Council found that low-performing students who are held back do less well academically, are much worse off socially, and are far likelier to drop out than equally weak students who are promoted.¹¹ The stakes are indeed high.

Problem 4: When Tests Fail

Tests themselves are often flawed, and high-stakes errors become high-stakes disasters. When Martin Swaden’s daughter failed the state math test by a single answer, Swaden requested to see the exam so that he could help his daughter correct her errors and pass the test next time around. It took a threatened lawsuit before he was able to meet with a state official to examine the answers. Together they made an amazing discovery: six of the sixty-eight answers were keyed incorrectly, not only for his daughter, but for all the students in Minnesota. Jobs had been lost, summers ruined, the joy of graduation turned to humiliation for those students who were misidentified as having failed. Suits followed and seven million dollars in damages were eventually paid, but the testing company argued that it was not liable for “emotional damages.”¹²

Unfortunately, such stories continue to mount as the crush of millions of new tests overwhelms the handful of testing companies. In Massachusetts, one senior spotted an alternative answer to a math question, and the scores of 449 students were suddenly propelled over the passing mark. A Massachusetts teacher noticed a question with two correct answers, and when the scores were adjusted, 666 more students passed. A flawed answer key incorrectly lowered multiple-choice scores for 12,000 Arizona students, erred in adding up scores of essay tests for students in Michigan and forced the re-scoring of 204,000 essay tests in Washington. Another error resulted in nearly 9,000 students in New York City being mistakenly assigned to summer school, and \$2 million in achievement awards being denied to deserving students in Kentucky.¹³ The National Board on Educational Testing and Public Policy reported that 50 high-profile testing mistakes had occurred in 20 states from 1999 through 2002.¹⁴ We cannot help wondering how many lives are altered by the continual missteps of the testing industry.

Problem 5: *The Gap Problem—and We Don't Mean the Score*

Using the same tests for all students, those in tony schools and those in dilapidated ones, is grossly unfair. Creating identical expectations for all students places the poorer ones at a distinct disadvantage. Students of color, students with disabilities, English-language learners, and low-income students are failing state graduation tests at rates as high as 60 to 90 percent, with a bleak future facing them.¹⁵ In Louisiana, parents requested that the Office for Civil Rights investigate why nearly half the students in predominantly poor and minority districts failed the state test, even after taking it for a second time. In Georgia, two out of every three low-income students failed the math, English, and reading sections of the state's competency tests. No students from well-to-do counties failed any of the tests and more than half exceeded standards. Even moderate-income differences have been powerful. In Ohio, almost half of the students from families with incomes below \$20,000 failed the state exams, while nearly 80 percent of students from families earning more than \$30,000 passed those same exams.¹⁶

Problem 6: *Test Sterilization*

While *NCLB* marches under the banner of improving achievement for *all* students, the tests themselves ignore diversity. Massachusetts Comprehensive Assessment System (MCAS) includes a 10th-grade history and social sciences test which focuses on world history. Of the 57 items on the 2000 test, about 40 referred to Europe, from the Byzantine Empire to the Cold War. Five were questions about capitalism. Only 12 were about the rest of the world. Students were given detailed maps of Eurasia and asked if the dotted routes represent the Crusades or the spice trade. For Africa, students received a bare outline of the continent and asked to point out the desert - where there are no people to have a history. Asians were reduced to shoguns, samurais, and the split between China and Taiwan. Latin America was but a colonial appendage of Europe. The failure rate for African-American and Latino students in 2000 on the MCAS 10th-grade history test was 77 percent and 85 percent.¹⁷

In a feat of literary sleuth work, Jeanne Heifetz, a weaver from Brooklyn and the mother of a high school senior, inspected 10 New York Regents high school English exams from recent years and discovered that most passages had been sanitized of virtually any reference to race, religion, ethnicity, and sex. The work of Jewish author Isaac Bashevis Singer becomes devoid of any reference to Judaism, and edits to Annie Dillard's memoir of racial segregation eliminates racial passages. Students are left to write essays and answer questions based on doctored and banal passages.¹⁸

Problem 7: *Teacher Stress*

In the quest for "magic bullets" to improve test scores, some schools and districts are returning to "teacher-proof" solutions. Rather than focusing their efforts on strengthening teachers' professional knowledge, administrators are seeking short-cuts that place teachers in the role of compliant technician to produce better test results.¹⁹ Not surprisingly, in a national study, nearly seven in ten teachers reported feeling test-stress, and two out of three believed that preparing for the test takes time from teaching important but non-tested topics.²⁰ Veteran elementary teachers request transfers, saying that they cannot stand the pressure of administering high-stakes exams to young children. Teachers recognized for excellence have left public schools for private schools where test preparation does not rule the curriculum.²¹ When 80 Arizona teachers and teacher educators were asked to visually depict the impact of standardized tests,

their drawings indicated test-driven classrooms where boredom, fear, and isolation dominate. Teachers felt that they are shortchanging schoolchildren from a love for learning.²²

The Forgotten Standards

While *NCLB* purports to measure schools, teachers, and student achievement, it may best be measuring measures of wealth. Those who start behind score behind, and these scoring gaps teach us once again about America's underclass. Although the rhetoric of the standards movement is that a rising tide raises all ships, in fact, without the adequate resources, some ships do not rise. The forgotten standards, "opportunity-to-learn" standards, were designed to insure a level playing field.

Opportunity-to-learn standards ask: Are students given an opportunity to relearn the standard if they fail the test? Are students given adequate resources to achieve the standards at their own pace? Is there adequate time to learn the standards, and adequate time to be tested on the standards? As test scores overshadow opportunity-to-learn standards, the real barriers to achievement remain firmly in place: racism, poverty, sexism, inadequate teacher salaries and training, language differences, inadequate facilities, and a disregard for individual differences.

World Class Standards

Let's imagine that the provisions in the *No Child Left Behind* legislation are fully implemented. What might American education look like? Schools would open and close according to how they ranked on their students' aggregated test scores. Less time would be spent on art and literature, and perhaps history, with more time spent drilling reading and math. Test errors would be downplayed, as would the troublesome correlation between test scores and social status, race, ethnicity, and gender. Any joy of learning would wither in the shadow of testing pressure. The testing industry would be to education what the arms industry is to defense. The creating, scoring, and reporting of test results would become one of the country's true growth industries, with spin-offs of test prep and tutoring. Perhaps the stock market would witness a boom as tax dollars are funneled to private corporations promising higher test scores.

A democratic or purposeful education should progress beyond a myopic focus on literacy and numeracy testing. Even a casual glance at pressing global challenges suggests the need for a strategic educational realignment. Our world is marked by misunderstanding and hatred, cultural and religious strife, national hubris and xenophobia, massive poverty and an AIDS epidemic. All these issues scream for the attention of not only political leaders, but educators and students as well. Our schools can do much more than teach test-taking skills; they can promote creativity, caring and more meaningful learning, what we term "enduring lessons."

Enduring Lesson #1 – Understand Our Roots

There is a fundamental, even driving need, to more deeply understand ourselves both as individuals, and our place in the wider community. Too often we learn only a few salient aspects of our backgrounds, personal and family history, cultural and religious beliefs, gender differences and challenges, and a single view of our national heritage. These are incomplete lessons. We lack a healthy appreciation of key forces that shape and limit our perceptions. Through skilled instruction, teachers and students can gain insights into our motivations, our strengths and our weaknesses, the way we behave and think. A realistic and healthy sense of connection with multiple communities is the product of the first enduring lesson: "know yourself."

Enduring Lesson #2 – Encourage Individual Talents and Contributions

High test scores predict high test scores, but not much else: not problem-solving skills, not good work habits, not honesty, not dependability, not loyalty, nor any cherished virtue. Schools should prepare students to live purposeful and satisfying lives, and to develop their unique interests, abilities, skills and talents. By measuring all students against the same yardsticks of literacy and numeracy, individual creativity and differences are lost or denigrated. Contrary to the current testing wave, we advocate less standardization and more individualization.

Enduring Lesson #3 – Celebrate Others

Once we are comfortable and secure with ourselves, we can reach out and seek insights from other peoples and cultures. While some have called this *teaching tolerance*, we do not believe that diversity should be *tolerated*. Today, we tolerate things from sleazy business practices to shoddy home repairs. Simply “tolerating” diversity falls woefully short. Diversity should be celebrated. Cultural, racial and ethnic, gender, and religious differences offer us wondrous insights into the human experience. We can learn much from each other, and our challenge is to learn from our differences, not fear them.

Enduring Lessons # 4 – Promote Purposeful Lives

At the Antioch College commencement in 1859, Horace Mann advised the graduates, “Be ashamed to die until you have won some victory for humanity.” This era of test mania and materialism seems light years from that sentiment. We must recommit ourselves to teaching students how to be honest and caring, treating their families, peers, and especially strangers with love, compassion, and forgiveness. The way we learn to live our lives as adults, and not our test scores, is the true measure of our schooling.

¹ Adapted from “Standards Back Then” in Myra Sadker and David Sadker. *Teachers, Schools and Society*, 6th ed. (New York: McGraw Hill, 2003), pp. 268-269.

² Jay Heubert, “First Do No Harm” *Educational Leadership*, vol. 60, n. 4, 2002/2003, pp. 26-30. Linda McNeil, “Contradictions of School Reform: Educational Costs of Standardized Testing” *Phi Delta Kappan*, vol. 81, n. 10, 2000, pp. 729-734. John Merrow, “Undermining Standards,” *Phi Delta Kappan*, vol. 82, n.9, 2001, pp.653-659.

³“Quality Counts 2001: A Better Balance.” *Education Week*, vol. 20 n.16. (Bethesda, MD: Editorial Projects, 2001).

⁴ James Hoffman, Czop, L. and Paris, S “High-stakes testing in reading: Today in Texas, Tomorrow?, *The Reading Teacher*, v. 54, no. 5, 2001, pp. 482-492.

⁵ “Quality Counts,” 2001.

⁶ Hoffman, et al, 2001.

⁷ Audrey Amrein and David Berliner, “High-stakes testing, uncertainty, and student learning,” *Education Policy Analysis Archives*, v. 10, no. 18, March 2002, <http://epaa.asu.edu/epaa/v10n18/>.

⁸ Brian Stecher (2000). *The Effects of the Washington State Education Reform on Schools and Classrooms*. (Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing).

⁹Daniel Koretz and Sheila Barron. (1998). The validity of gains in scores on the Kentucky Instructional Results Information System (KIRIS). (Santa Monica, CA: RAND).

¹⁰ Amrein and Berliner, 2002.

¹¹ Heubert 2002/2003; McNeil, 2000; Neill, 2003.

¹² Norman Draper, “Testing Firm to Pay Students \$7 Million for Grade Goof.” *Star Tribune Online*, November 26, 2002.

¹³ Diana Henriques, “Rising Demand for Testing Push Limits of Its Accuracy.” *New York Times online*, September 3, 2003.

¹⁴ Monty Neill, 2003. “The Dangers of Testing,” *Educational Leadership*, February 2002, pp. 42-46.

¹⁵ Heubert, 2002/2003.

¹⁶ Amrein and Berliner, 2002; Heubert, 2002/2003; Neill, 2003.

¹⁷ Derrick Jackson, “At Best, Silly, At Worst, Racist.” In Kathy Swopes & Barbara Miner (eds.) *Failing Our Kids: Why the Testing Craze Won't Fix Our Schools*. Milwaukee, WI: Rethinking Schools, 2000, pp.18-19.

¹⁸ N.R.Kleinfield, “The Elderly Man and the Sea? Test Sanitizes Literary Texts” *New York Times on the Web* June 2, 2002.

¹⁹ David Hoff, "Standards at Crossroads after Decade," *Education Week*, vol.19, n. 3, September 22, 1999, pp. 1, 9.

²⁰ “QualityCounts,” 2001.

²¹ John Goodlad. “Kudzu, Rabbits, and School Reform,” *Phi Delta Kappan*, vol. 84, n.1, September 2002, pp.16-23.

²² Tirupalavanam Ganesh, “Educators’ Images of High-Stakes Testing: An Exploratory Analysis of the Value of Visual Methods,” Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 2002.