



THE EDUCATIONAL
Forum

High-Stakes Hustle: Public Schools and the New Billion Dollar Accountability

by Lawrence A. Baines and Gregory Kent Stanley

Abstract

High-stakes testing costs up to \$50 billion per annum, has no impact on student achievement, and has changed the focus of American public schools. This article analyzes the benefits and costs of the accountability movement, as well as discusses its roots in the eugenics movements of the early 20th century.

Once upon a time, the dirty little secret of standardized testing was that the tests might be culturally biased. These days, the biggest drawback may be the exorbitant costs to develop, maintain, administer, prepare, and publish them.

The rise of expansive bureaucracies within state governments whose sole purpose is to keep up with student testing has forever altered the budgets of public schools. In 1967, 80 percent of a school's budget was devoted to regular instruction. By the late 1990s, the percentage of funds devoted to regular instruction had dwindled to about 50 percent (Rothstein and Miles 1996). To meet the prohibitive costs of testing, cash-strapped schools, particularly those in urban and rural areas, have paid for testing from funds originally designated for hiring teachers, fixing leaking roofs, and buying new books.

One unanticipated problem with high-stakes testing is that administrative costs have been not one-time debits, but continuous expenses. To keep up with the paperwork, state governments have created bureaucracies within bureaucracies—Offices of Educational Accountability, Offices of Test Administration and Reporting, Offices of Standards and Measurement—while simultaneously reducing other staff, including teachers, to stay within budget.

The annual cost of high-stakes testing rivals the gross national products of some small countries, somewhere between \$20 and \$50 billion, or 5.5–14 percent of every dollar spent for public schools (Center for Education Policy 2003). So what has the annual

expenditure of \$20–50 billion bought for America’s children? From all available evidence, these billions of dollars have improved almost nothing.

Ostensibly, one reason for crafting expensive testing systems was to assure that public schools adhered to high standards. Yet, the “subjects tested” and the difficulty of exams vary by state. Furthermore, the aforementioned \$20–50 billion does not take into account the costs of remediation for children who fail to make the grade. For example, if only 5 percent of New York City’s children fail the state exam—an impossibly optimistic number—holding these students back a year would cost an additional half-billion dollars (Education Priorities Panel 1999). Raising the bottom 50 percent of low-performing districts to the average level of performance would cost an additional \$25 billion—just in New York City. For the past few years, many urban districts (Los Angeles most notoriously) have attempted to implement a “no social promotion” policy only to discover that they lacked the funds and the physical space to retain failing students.

Mathis (2003) estimated that the new accountability requirements of *No Child Left Behind* would force state governments to increase their expenditures on education by 30 percent. In responding to NCLB, some wealthy schools have created new positions—deans of finance or assistant principals of accountability—whose full-time job is to keep the school in compliance with the numerous tenets of the law. These administrators, though necessary, were hired at a cost equal to the salaries of three to four new teachers per year (Robelen 2003).

The Benefits of High-Stakes Accountability

Because test performance has become the sole criterion for success, other measures such as student motivation, teacher satisfaction, and preparation for college are rarely mentioned in discussions of high-stakes accountability. Data indicated that student motivation is low and getting worse (National Center for Education Statistics 2003), teacher attrition is high and getting higher (Rose and Gallup 2003), and more than half of today’s high school graduates will not be ready for college-level math or science (Giegerich 2003; Greene and Winters 2002; National Center for Education Statistics 1996).

The California Department of Education (2003) issued a news release on August 15, 2003, detailing the improvements in student performance in all subjects and at all grade levels for five consecutive years on STAR (Standardized Testing and Reporting), its state exam. California also administers the National Assessment of Educational Progress (NAEP) test, a nationally normed achievement exam given since 1969. The performance of California students on the NAEP over the same time period, however, tells a very different story. According to the national test, student achievement in California was mixed during the past five years—scores were up in mathematics at grade 4, down at grade 8; up in reading at grade 4, down at grade 8; and down in science in all grades, but slightly up in writing. While STAR showed consistent gains in achievement, the NAEP tests showed no significant change in achievement among California students for a decade (NAEP 2003). California is not alone in hyping selective test results as indicators of enhanced achievement. Florida Governor Jeb Bush touted students’ accomplishments in a May 15, 2003, news release (Governor’s Office 2003):

Baines and Stanley

Governor Jeb Bush today announced that more students scored on grade level in 2003 than ever before on the Florida Comprehensive Assessment Test (FCAT). "The greatest improvements over last year were found in reading at the elementary grades. . . . We have seen continued rising student achievement across the state, regardless of ethnic background. This year brings the biggest improvement yet in student performance," said Governor Bush. "These results prove that our common sense approach to education reform through accountability works."

While it is true that the scores of Florida fourth-graders improved a few points on the FCAT from 1994 to 2002, scores of Florida students on the NAEP were still well below the national average. According to the state of Florida, 60 percent of fourth graders were proficient in reading in 2002; but according to the NAEP (2003), only 27 percent were proficient. A few months after Jeb Bush's celebratory note, Florida was notified by the federal government that 90 percent of its students would have failed to make *adequate yearly progress* (AYP) as set forth in the No Child Left Behind Act (Newkirk 2004).

One reason for impressive scores on state exams is that the precise contents of the exams typically are published years in advance. A second reason state exams can show increases in achievement is that students eventually become familiar with the content and format of the exam. Since the advent of high-stakes testing, students have started the year with pep rallies designed to drum up enthusiasm for taking the test, taken courses carefully aligned to the state-sanctioned curriculum, been required to answer sample test questions as part of a daily routine, and spent hundreds of hours in courses whose sole purpose is to teach about "test-taking tricks." The vast amounts of time, energy, and money spent on test-taking strategies have resulted in some increases in scores on standardized exams, but have offered little educational value.

From her vantage point as a high school English teacher in Florida, Nancy [Williams \(2003\)](#) noted the deleterious effects the shift from a low- to a high-stakes exam can have on teachers and students. Low-stakes tests involve only a half-day of administration and require no rehearsal or preparation on the students' part. Conversely, high-stakes testing mandates months of heavily monitored, test-preparation sessions followed by several weeks of intensive seatwork and, finally, the administration of the multi-part exam over a period of several days. Not only do high-stakes exams negatively impact instructional time, Williams (2003, 83) also found that they fostered "resentment and decreased motivation."

Of course, some high poverty schools inevitably slide to the state's "low-performing" status. When schools are designated as being "on probation" or "unacceptable" because of scores on the state exam, pressure increases and the curriculum narrows even further. If a state tests only reading and mathematics in elementary grades, then a low-performing school usually decides to abandon music, art, physical education, science, and history in favor of a curriculum consisting of only the subjects to be tested (Johnson and Johnson 2002). After a poor performance on the state exam for the third consecutive year, an elementary school in Lawrence, Massachusetts, even shoved aside core subjects so that they could implement 15 solid weeks of two-hour drills on test-taking tricks expressly designed to raise scores on the state exam ([Vogler and Kennedy 2003](#)).

In this manner, tests have not only changed the function of schools, but they also have become the focal point for schooling. This scripted approach to learning and emphasis on test-taking strategies makes schools more like private, for-profit, test-prep centers in the mold of Kaplan or Sylvan and less like schools—at least American schools of the past.

Valid Questions About Validity

Another area of concern about state exams is their validity. Recently, I, along with a group of teachers from around the state, was invited by a state department of education to offer input on a new testing program. Of the 80 multiple-choice questions that would decide whether or not a student would pass 11th grade U.S. history, five did not offer a correct answer choice.

One of the questions was: “The Spanish and English colonies in the New World during the 15th and 16th century differed from each other in which of the following ways?” The correct answer was supposed to be, “While the Spanish settlements were temporary, the English attempted permanent colonies.” The teachers noted that the trouble with the question was that Jamestown was not established until 1607, and it was more of a get-rich-quick scheme than a permanent settlement. The first permanent settlement was at Plymouth in 1620—during the 17th century.

The state testing director finally conceded the point, but claimed that the question was not a test of factual recall, but one of “spatial comparison.” The spatial comparison claim was made again when teachers noted that one of the map questions had several cities located in the wrong states. Difficulties with the validity of test items have arisen in Minnesota, New York, Texas, Georgia, Massachusetts, and several other states ([Rhoades and Madaus 2003](#)).

The state representative also had a hard time convincing teachers why all students must take the same test. One teacher asked, “Why should a learning disabled child who believes that his hair is always on fire be held to the same standard as my gifted, straight-A, student body president?” A second teacher asked why the tests would be given in only English and commented, “My district is more than half Hispanic, mostly Guatemalans whose parents come to work in the fields during the harvest. Most do not speak or read English. How will they pass this exam?”

The answer to both questions was the same: “Under the No Child Left Behind Act, there are no mitigating circumstances. Every student must take the same exam and receive the same instruction. There are no more excuses.”

High-Stakes Anxiety

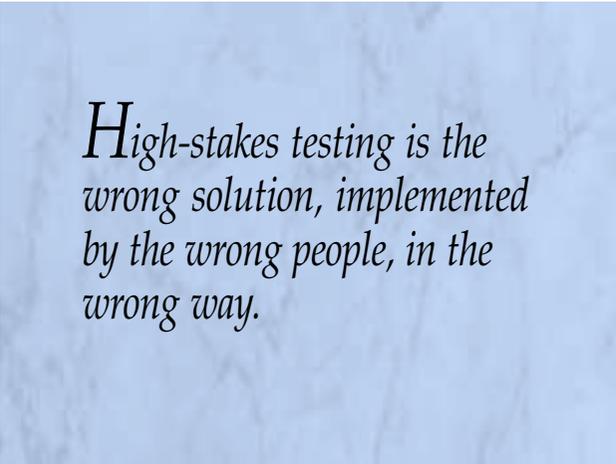
It is difficult to overestimate the burden on a school to increase test scores. Parents moving into a new community may choose residential locations based on test scores of the neighborhood school, now readily available on the Internet and through most real estate offices. Administrators may receive hefty bonuses for boosting scores, and teachers can lose their jobs or a significant portion of their paychecks if students fail to achieve AYP.

Baines and Stanley

Some state governments offer superintendents monetary incentives worth more than a beginning teacher's salary for better scores. Paradoxically, administrators may be able to pocket these inducements at the same time that their schools go without essential supplies, teachers get laid off, and class sizes double.

To achieve AYP, some schools already have resorted to number fudging—funneling borderline students into GED programs, steering non-achievers into special education IEP diploma tracks, selectively admitting students to the Scholastic Aptitude Test (SAT), and planning special field trips for low-achieving students so that they will be out of town on test days. Recently, the Department of Education in Massachusetts (Haney, Madaus, and Wheelock 2003, 4) claimed that 90 percent of students passed the Massa-

chusetts Comprehensive Assessment System (MCAS) by not counting 17,000 students who would have significantly reduced the overall pass rate to around 70 percent. The reported passing rate for specific ethnic groups was even more distorted. For example, the passing rate for Latinos was reported at 70 percent, while in actuality it was 45 percent (Haney, Madaus, and Wheelock 2003, 4).



High-stakes testing is the wrong solution, implemented by the wrong people, in the wrong way.

Such “data massaging” is not done to better serve the unique learning needs of students, but to make the numbers look better. In the new, billion-dollar

accountability system, the voice of teachers—the only people who have daily contact with students—have little say. The system is designed to hold teachers accountable, yet teachers have no vote in determining the composition of classes, the curriculum, or the assessment.

Foster (2003, 174) tracked the development of standardized exams to the “factory model” of education where “students were products and teaching was the machinery to make the products.” Popham (2003, 50) asserted that standardized tests “constitute a serious violation of any sort of truth-in-advertising precept. Standards-based tests don’t measure what they pretend to measure. . . . In no case do these tests provide data that teachers can easily use to appraise their own instructional effectiveness.”

Indeed, no professionals are held accountable in the same simplistic manner as teachers. Lawyers are not held accountable when their clients are sent to prison. Doctors say with resigned regularity that the operation was a success but the patient died anyway. If a patient smoked three packs of cigarettes per day and worked in an asbestos-filled environment, no one would blame the doctor if he couldn’t miraculously cure a case of lung cancer. Yet, such bogus accountability is imposed on teachers with regularity. If an emotionally disturbed, learning disabled child lives with a homeless crack addict and ends up missing 40 percent of the school year, the teacher

still is culpable for that student's performance on the standardized exam. With the new accountability system, having mainstreamed, learning-disabled, and emotionally disturbed students in a classroom can be detrimental for teachers who must post impressive gains in achievement.

Testing allows public officials to pretend that nothing external to the classroom influences student behavior, enabling what Claybaugh (2003, 60) called "the fraud of limitless teacher accountability" to run rampant.

The Beauty of Numbers

The most obvious benefit of high-stakes testing is that it enables a numeric value to be attached to every school and every student in the country. The reverence for numbers is an obsession that dates from the days of Isaac Newton, who promised that everything in the universe operated under quantifiable laws and demonstrable order. Unfortunately, like lines from Shakespeare, numbers can be molded to suit a particular purpose. It is common practice for pseudoscientists and politicians to promulgate absurd theories behind a flash of statistics and solid scientific data.

For example, Samuel Morton, one of the most famous men of science in antebellum America, believed that races could be ranked in order based upon easily quantifiable terms—in this case, head size. Craniometry offered a quick, simple, and undeniable answer to the problem of measuring human intelligence and soon swept through America and Europe. The only problem with craniometry was that it had no basis in scientific fact. To sell his theory, Morton cooked the data, discarded evidence that failed to support his findings, and exaggerated results that seemed to fit his theory. "To put it bluntly," [Gould \(1981, 57\)](#) wrote, "Morton's summaries are a patchwork of fudging and finagling in the clear interest of controlling *a priori* conclusions."

Craniometry lost its luster when IQ tests became the rage. The lure of IQ scores was that they required no physical measurement. Instead, a number representing general intelligence added a great deal of portability to the notion of measuring intelligence. A number could be calculated without even meeting the individual in question, and then numbers could be compared and ranked. Unfortunately, Binet's tests were utilized by eugenicists who, like Morton, saw a quick and easy way to measure human intelligence. Eugenicists, such as Goddard and Yerkes, believed that a better world could be achieved immediately without waiting around for evolution or the arduous process of social change. Their solution was simple: identify the inferior people and keep them from breeding. Goddard advocated confinement centers for the unfit and a ban on undesirable immigration.

© 2004 by James A. Banks

Yerkes fanned the flames of anti-immigration fervor by citing the low scores of Black Americans and newly arrived Southern Europeans on Army Intelligence Tests during World War I. Through such comparison, Yerkes sought to reveal what everyone already knew: Black Americans and Southern European immigrants were morons who threatened the future of the nation. According to Yerkes (1921), test scores provided sufficient proof of genetic inferiority.

Baines and Stanley

At the time, no one thought to question how a score on an exam that queried knowledge of American history and baseball would correlate with an individual's performance as a soldier on the battlefield or as a citizen of the United States. Instead, panic hit the nation and books poured off printing presses warning real Americans of the danger of the quantifiable menace (Wiggam 1923). Even the federal government bowed before the numbers, and in the 1920s Congress legislated a near-ban on inferior immigrants. It did so even though its own data gathered by the Dillingham Commission refuted the numbers gathered through Army Intelligence Tests ([Higham 1955](#)).

During the past 20 years, proponents of accountability decided that they could no longer trust teachers to teach. Consequently, control over learning and assessment was wrested from teachers and handed over to the states. Hundreds of billion dollars later, America has 50 sets of state standards, 50 new burgeoning bureaucracies, and no evidence whatsoever that anything has improved. Eventually, the absurdity of spending \$20–50 billion per annum on a failed educational strategy while chemistry classes go without chemicals, literature classes go without books, and computer classes go without computers may generate some interest.

Tests have not only changed the function of schools, but they also have become the focal point for schooling.

Fifty billion dollars would help pay for almost half of the cost to repair the crumbling infrastructure of public schools (U.S. Department of Labor 1995). Fifty billion dollars would pay the salaries and benefits of an additional one million teachers. Fifty billion dollars would furnish each student in American public schools with a laptop computer and unlimited Internet access. Fifty billion dollars would provide every child in North America with three meals a day.

Instead, America's \$50 billion is going toward an accountability system which yields data of dubious validity concerning student achievement on exams created by state bureaucracies. Even the most radical proponent of the accountability movement must concede that our \$50 billion is not buying much.

Eliminating high-stakes accountability would free up months of instructional time, reduce paperwork, reduce the size of state bureaucracies, and might convince students that learning can occur without coercion and without extensive training on the proper way to fill in a bubble on a Scantron test.

High-stakes testing is the wrong solution, implemented by the wrong people, in the wrong way. Like pseudoscientists who jumped on craniometry and IQ tests to distort their diagnostic purpose, proponents of the new, billion-dollar accountability, most of whom have never worked in the classroom, have used testing to seize control of public

schools. Soon, to graduate from high school, 70 percent of all American students will have to pass exit exams, prepared by state governments at additional taxpayer expense (Center for Education Policy 2003). A close scrutiny of the scores of students in certain classrooms may reveal that these public schools are full of morons who threaten the future of the nation. It doesn't take Yerkes to figure out what might happen next.

References

- California Department of Education. 2003. State schools chief O'Connell applauds 2003 STAR results showing fifth straight year of gains in student achievement. Los Angeles: CDE. Available at: www.cde.ca.gov/nr/ne/yr03/yr03rel43.asp.
- Center for Education Policy. 2003. *State high school exit exams: Put to the test*. Washington, DC: CEP.
- Claybaugh, G. K. 2003. What doesn't work. *Educational Horizons* 81(2): 58–60.
- Education Priorities Panel. 1999. Education: Higher standards, the need for more time and money. *EPP Monitor/Newspaper* 3(2). New York: Education Priorities Panel. Available at: www.edpriorities.org/Pubs/PubsArchive/pubs_99Win.Educ.html.
- Foster, H. 2003. The crazy curriculum: Teaching that matters. *Educational Horizons* 81(4): 174–77.
- Giegerich, S. 2003. ACT, SAT results show discrepancy: One test shows high scores; takers of other did poorly. *The Detroit News*, September 8, A3.
- Gould, S. J. 1981. *The mismeasure of man*. New York: Norton.
- Governor's Office. 2003. Governor Jeb Bush announces biggest improvement ever on FCAT. Tallahassee: State of Florida. Available at: http://sun6.dms.state.fl.us/eog_new/eog/library/releases/2003/May/fcat-05-15-03.html.
- Greene, J. P., and M. A. Winters. 2002. Public school graduation rates in the United States. Washington, DC: Center for Civic Innovation at the Manhattan Institute. Available at: www.manhattan-institute.org/html/cr_31.htm.
- Haney, W., G. Madaus, and A. Wheelock. 2003. DOE report inflates MCAS 'pass' rates for class of 2003. Newton, MA: National Board on Educational Testing and Public Policy. Available at: www.bc.edu/research/nbetpp/reports.html#monographs.
- Higham, J. 1955. *Strangers in the land: Patterns of American nativism, 1860–1925*. New Brunswick, NJ: Rutgers University Press.
- Johnson, D. D., and B. Johnson. 2002. *High stakes: Children, testing, and failure in American schools*. New York: Rowman & Littlefield.
- Mathis, W. J. 2003. No child left behind: Costs and benefits. *Phi Delta Kappan* 84(9): 679–86.
- National Assessment of Educational Progress. 2003. Reading: The nation's report card. Washington, DC: National Center for Education Statistics. Available at: <http://nces.ed.gov/nationsreportcard/reading/results2003>.
- National Center for Education Statistics. 1996. *Youth indicators 1996*. Washington DC: Institute of Education Services.
- National Center for Education Statistics. 2003. *The condition of education*. Washington, DC: Institute of Education Services.
- Newkirk, T. 2004. False positives. *Education Week*, March 3. Bethesda, MD: Editorial Projects in Education. Available at: www.edweek.org/ew/ewstory.cfm?slug=25Newkirk.h23.
- Popham, J. W. 2003. The seductive allure of data. *Educational Leadership* 60(5): 48–51.
- Rhoades, K., and G. Madaus. 2003. Errors in standardized tests: A systemic problem. Newton, MA: National Board on Educational Testing and Public Policy. Available at: www.bc.edu/research/nbetpp/reports.html#monographs.
- Robelen, E. W. 2003. More school districts feeling the effects of 'No Child' law. *Education Week*, 4 February. Bethesda, MD: Editorial Projects in Education. Available at: www.edweek.org/ew/ewstory.cfm?slug=21CEP.h23.
- Rose, L. C., and A. M. Gallup. 2003. The 35th annual Phi Delta Kappa/Gallup poll of the public's attitudes toward the public schools. Bloomington, IN: Phi Delta Kappa. Available at: www.pdkintl.org/kappan/k0309pol.htm.
- Rothstein R., and K. H. Miles. 1996. *Where's the money gone?* Washington, DC: Economic Policy Institute.
- U.S. Department of Labor. 1995. *School facilities: America's schools not designed or equipped for 21st century*. Washington, DC: U.S. General Accounting Office.
- Vogler, K. E., and R. J. Kennedy. 2003. A view from the bottom: What happens when your school system ranks last? *Phi Delta Kappan* 84(6): 446–48.
- Wiggam, A. E. 1923. *The new decalogue of science*. Indianapolis: Bobbs-Merrill.
- Williams, N. 2003. Thinking outside the bubble. *Educational Leadership* 61(3): 82–83.
- Yerkes, R. M., ed. 1921. *Memoirs of the National Academy of Sciences. Vol. 15: Psychological examining in the U.S. Army*. Washington, DC: GPO.

Lawrence A. Baines teaches at the University of Toledo. His most recent books are *How to Get a Life: Empowering Wisdom from Thinkers and Writers (2004)* and *Teaching Adolescents to Write: The Unsubtle Art of Naked Teaching (2003)*.

Gregory Kent Stanley teaches in the International Baccalaureate Honor's program at Calhoun High School in Calhoun, Georgia.