# ISSUE 9

## Do High-Stakes Assessments Improve Learning?

YES: Nina Hurwitz and Sol Hurwitz, from "Tests That Count," *American School Board Journal* (January 2000)

NO: Martin G. Brooks and Jacqueline Grennon Brooks, from "The Courage to Be Constructivist," *Educational Leadership* (November 1999)

ISSUE SUMMARY

YES: High school teacher Nina Hurwitz and education consultant Sol Hurwitz assemble evidence from states that are leading the movement to set high standards of educational performance and cautiously conclude that it could stimulate long-overdue renewal.

NO: High school superintendent Martin G. Brooks and associate professor of education Jacqueline Grennon Brooks contend that the push for standardized state assessments constricts student learning and prevents implementation of constructivism.

Jin the 1980s a number of national reports found America's public schools to be seriously lacking in the production of students who were qualified to compete successfully in the emerging global economy. Among these reports were the National Commission on Excellence in Education's *A Nation at Risk*, the Education Commission of the States' *Action for Excellence*, the Twentieth Century Fund's *Making the Grade*, and the National Science Foundation's *Educating Americans for the Twenty-First Century*. In response to these calls for higher standards, the Bush administration adopted the following national goals in its "America 2000" plan: all children starting school prepared to learn, at least 90 percent of students graduating from high school, all students being able to cope with challenging subject matter (particularly math and science), all adults being literate and responsible citizens, and all graduates being able to compete in a global economy.

To move toward these goals the Republican administration emphasized more choice and competition, more influence from business leaders, and the development of nationwide curriculum standards and testing programs. The



Clinton administration adopted the goals, renaming them "Goals 2000: Educate America," but downplayed the role of the private sector and placed responsibility for assessing student progress on the individual states. The George W. Bush administration, while backing away from privatization and voucher issues, has strengthened the call for accountability and rigorous testing.

To date, many states have risen to the challenge, imposing statewide standardized tests of subject matter and mental skills. In some cases, states have set up procedures for taking over the administration of chronically underperforming local schools. In others, schools that dramatically increase student performance are rewarded in some tangible way.

Peter Sacks, in *Standardized Minds: The High Price of America's Testing Culture and What We Can Do to Change It* (1999), contends that "the case against standardized testing is as intellectually and ethically rigorous as any argument about social policy ... and yet such testing continues to dominate the education system, carving further inroads into the employment arena as well." He further warns that "when thinking becomes standardized people are easily objectified, their skills and talents translated into the language and mechanisms of commercial enterprise." This sentiment is shared by Alfie Kohn, who, in "Unlearning How We Learn," *Principal* (March 2000), says that "raising standards has come to mean little more than higher scores on poorly-designed standardized tests," leading to abandonment of the best kind of teaching and learning.

This central concern about the direct impact of test mania on the nature of the learning process has been widely voiced by educators who advocate instructional approaches based on the theory of constructivism. Drawn from the thinking of John Dewey, Jean Piaget, Lev Vygotsky, Howard Gardner, and others, constructivism views learning as an active, group-oriented process in which students "construct" an understanding of knowledge utilized in problem-solving situations. Such sense-making activities can be time-consuming and therefore can get in the way of teachers and schools whose primary focus is on test performance. Applications of the theory are explained in Mark Windschitl's "The Challenges of Sustaining a Constructivist Classroom Culture," *Phi Delta Kappan* (June 1999).

Those who see the standards and testing movement as the clearest path to school improvement include Joan L. Herman, "The State of Performance Assessments," *The School Administrator (December* 1998); Jerry Jesness, "Why Johnny Can't Fail: How the 'Floating Standard' Has Destroyed Public Education," *Reason* (July 1999); and Mike Schmoker, "The Results We Want," *Educational Leadership* (February 2000). These and other advocates feel that a standardized testing program ensures acquisition of basic skills, holds schools accountable for results, and identifies problem areas.

In the following selections, Nina Hurwitz and Sol Hurwitz examine experiences with high-stakes testing in Texas, Chicago, and New York in order to identify crucial elements of successful implementation, while Martin G. Brooks and Jacqueline Grennon Brooks argue that the central aim of improving student learning on a long-term basis is not well served by high-stakes accountability pressures.

**Nina Hurwitz**
**and Sol Hurwitz**


# Tests That Count

They are tests that count, high-stakes tests that can deny promotion or graduation to students with failing scores. Schools with too many low-performing students can be exposed to the glare of publicity, placed on probation, or closed. A widening coalition of governors, business leaders, parents, and teachers—appalled that youngsters can advance through school, receive a diploma, and seek further education or a job without mastering basic skills—is promoting the use of these tests as a means of boosting standards and improving accountability in public education.

The movement is gaining national momentum. Forty-nine states have adopted performance standards for elementary and secondary education; 26 have high school exit exams in place or in process; 19 publicly identify failing schools. President Clinton is in the vanguard, calling for higher standards and a crackdown on social promotion. Last fall he urged the nation's governors, "Look dead in the eye some child who has been held back [and say], 'We'll be hurting you worse if we tell you you're learning something when you're not.' "

## High Standards, High Stakes

High-stakes testing is forcing the debate over a fundamental question in American education—whether it is possible to achieve both excellence and equity. On one side are those who claim that tests with consequences are the only sure route to higher standards and stricter accountability. On the other are those who contend that high-stakes tests are a command-and-control instrument for "standardizing" education and punishing disadvantaged and minority children. But a more pragmatic middle position is evolving based on the experience of front-line practitioners: High-stakes testing can work with clear but limited goals, flexibility in meeting those goals, and the will to address head-on the problems of students at risk of failure.

Texas, Chicago, and New York City and state, discussed below, are being carefully watched by educators and decision makers nationwide for both positive and negative lessons. The states are driving the high-stakes movement: Kentucky, Maryland, Massachusetts, North Carolina, and Virginia are running noteworthy programs as well.

From Nina Hurwita and Sol Hurwitz, "Tests That Count," *American School Board Journal* (January 2000). Copyright © 2000 by The National School Boards Association. Reprinted by permission of *American School Board Journal.*





Even as states and school districts attempt to raise standards and impose high-stakes tests, they are confronted with excessive numbers of their urban, minority, and disadvantaged students who are failing these tests. In urban districts, large-scale failure is inevitable, says *Education Week's* Ron Wolk, given the shoddy education these students are receiving. "For tens of thousands of urban youngsters, it's a kind of double jeopardy," Wolk declares. "The system failed to educate them adequately, and now it punishes them for not being educated." Schools and school districts might face punishment as well: Low scores could result in the reorganization of schools or a shift of resources to charter schools or private-school vouchers.

Parent advocacy and civil rights groups are challenging the tests on racial and equity grounds. The penalties, they claim, fall disproportionately on minority and at-risk students, who have been shortchanged in their education. Meanwhile, teachers and researchers are beginning to question the tests' educational validity: Do they, in fact, improve learning?

Educators are unanimous that high-stakes tests should be aligned with curriculum and instruction—they should measure what students have been taught and are expected to know—and that teachers should be involved in the process. But only gradually are states and school districts committing sufficient resources and time to achieve proper alignment with full teacher participation. The time lag, educators argue, makes it risky to impose consequences prematurely.

Disagreement between states and urban school districts over which test to use means students in the same grade might have to take two tests—in the same subject. Learning suffers, educators say, when teachers spend time preparing students for too many tests. "The first thing to go in a school or district where these tests matter," says education expert Alfie Kohn, "is a more vibrant, integrated, active, and effective kind of instruction." [See "Raising the Scores, Ruining the Schools," *ASBJ,* October 1999.] A fifth-grade teacher in Virginia concurs: "Sometimes, when I wish I could stay longer on a subject, I have to move on to prepare my kids for the tests."

A 1999 study titled *High Stakes* by the prestigious National Research Council sharply criticized the practice of relying solely on tests to determine promotion or graduation. Such decisions, the council argued, "should be buttressed by other relevant information about the student's knowledge and skills, such as grades, teacher recommendations, and extenuating circumstances." Many educators question the value of holding kids back *period-but* certainly not without a highly structured, and often costly, intervention and remediation strategy.

The growing public demand for standards with accountability has made high-stakes testing a tempting political issue. The public is fed up with low standards and courses that lack content—they want American students to be able to compete favorably with kids in other countries. Test scores provide an aura of businesslike accountability for superintendents, principals, and teachers and a stimulus for students. Initially, at least, testing seems easy and inexpensive compared with more deep-seated reforms such as hiring and training competent teachers, reducing class size, or repairing crumbling school buildings. But achieving accountability is neither simple nor cheap.

The states and school districts that have had the most success with high-stakes testing share several common characteristics. They have maintained bipartisan political support and the backing of a broad coalition of interest groups, including the business community, over a sustained period. High-stakes tests have not only raised standards but have stimulated systemwide reform. Most important, there has been a heavy investment in addressing the academic performance of the weakest students.

## Turnaround in Texas

Texas is a dramatic case in point. Once considered one of the nation's educational backwaters, the Texas public school system, according to the *New York Times,* is now viewed by educators as "a model of equity, progress and accountability." The state's education reforms have spanned the administrations of former Democratic governor Ann W. Richards and current Republican governor and presidential hopeful George W. Bush. In a system of 3.7 million students that is half African American and Hispanic, the scores of African-American and Hispanic students on national assessments in reading and mathematics in 1996 and 1998 outranked those of most other states, and scores on state assessments for all students have risen for the fourth straight year.

A unique feature of the Texas system is the Texas Assessment of Academic Skills (TAAS), the state's high-stakes exam program, introduced in 1990. The tests combine clearly stated educational standards with a detailed reporting of results by ethnicity and class. Scores are sorted according to white, African-American, Hispanic, and economically disadvantaged groups. Along with attendance and dropout rates, TAAS scores are used to identify a school as failing if any one of its four demographic cohorts falls below standard. "Disaggregation of scores has focused the schools' attention on kids that were once ignored," according to University of Texas professor Uri Treisman, who is director of the Charles A. Dana Center in Austin. Texas is gradually raising the passing bar to 50 percent for each cohort from the original 20 percent.

Until recently, the high stakes associated with the TAAS have consisted almost entirely of public disclosure of school-by-school test results, a process Gov. Bush calls "shining a spotlight of shame on failure." The ratings, published on the web, identify schools as exemplary, recognized, acceptable, or low-performing and provide strong incentives to improve for adults and students alike. For example, superintendents, principals, and teachers find it hard to get jobs if they come from failing schools. Although low-performing schools are bolstered by additional financial support, they are rarely closed. "There are no great ideas on what to do with really problematical schools," says Treisman.

Elementary and middle-school students are tested in grades three through eight on various combinations of reading, writing, and mathematics, with science and social studies added in the eighth grade. In response to political pressure, Texas will move to end social promotion by 2003. Hoping to avoid widespread retention, the state has instituted the Student Success Initiative, an early-intervention program, starting with the current year's kindergarten class. A skeptic on retention, Treisman cautions that research on the dropout problem indicates that "being overage in your class has the single highest correlation for dropping out and is twice as high as for any other factor, including race." Also, as pressure mounts to pass the TAAS, state officials have become increasingly concerned over outbreaks of alleged cheating.



There is wide agreement that Texas high schools have not shown as much improvement as elementary schools. However, the state plans to beef up the content of the lOth-grade exit exam and move it to the llth grade and to allow substitution of end-of-course exams in core subjects. The present 10th-grade exam is the subject of a lawsuit by the Mexican-American Legal Defense Fund in the U.S. District Court in San Antonio, which claims the test discriminates against minority students. Gov. Bush counters this claim: "Some say it is a racist test," he told the *New York Times* [recently]. "I strongly say it is racist not to test because by not testing we don't know, and by not knowing we are just moving children through the system." The outcome is sure to have an impact on other states, researchers agree.

Remarkably, the state's educational resurgence has occurred while expenditures remained below the national average: In 1998-99, Texas spent $5,488 per student compared with the national average of $6,407. Striking, too, is the autonomy that Texas gives its principals and teachers as long as test results remain positive. Bilingual education, for example, is a local option. However, scores for Spanish-speaking and special education students must now be included in overall ratings to ensure more accurate results.

## Success in Chicago

Just as Texas has drawn raves for educational attainment at the state level, Chicago, with an enrollment of 431,000 students, has become the promised land for city reformers. Mayors, superintendents, and educators have flocked there to study the remarkable turnaround orchestrated by chief executive officer Paul Vallas, formerly budget director under Mayor Richard M. Daley. With no previous experience in education, Vallas has performed what many consider an educational miracle in a school district that U.S. Secretary of Education William J. Bennett in 1987 called "the worst in the country." Vallas achieved credibility largely through the selective but determined application of highstakes testing. Scores on Chicago's performance benchmark, the Iowa Test of Basic Skills in Reading and Mathematics, have risen for the fourth straight year.

When Vallas took charge in 1995, an earlier reform effort, which stressed decision making by local school councils, had virtually hit bottom. With high truancy, low standards, and rampant grade inflation, Vallas declared, "there was wide agreement that the earlier reform initiative had failed." Vallas exploited public dissatisfaction with the previous reform, while drawing grassroots support from a network of parents, community groups, foundations, and universities to fashion a new strategy.

Fundamental to his success was the solid backing of Mayor Daley and Gery Chico, president of the Chicago School Reform Board (successor to the former elected school board), whose members were all mayoral appointees.

Vallas forged a close working relationship with Tom Reece, who heads both the Chicago Teachers Union and the Illinois Federation of Teachers, and together they have succeeded in avoiding strikes and confrontations by building communication and trust. In addition, he won points with the public for his skills as a financial manager by stamping out waste, ending deficits, and securing state funds for building and renovating schools.

Three years ago, Chicago gained national attention as the first big-city school system to end social promotion. Students who don't meet minimum standards on the Iowa Tests are at risk of retention, but the passing bar was set low at first to avoid massive failure and is only gradually being raised.

Chicago's promotion gates kick in for students in grades three, six, and eight. For those who fail, the city's mandatory Summer Bridge Program, staffed with experienced teachers, provides a scripted curriculum from the central administration with hour-by-hour guidelines. University of Chicago professor Melissa Roderick, a member of the Consortium on Chicago School Research, believes the program goads parents, students, and teachers to work harder to avoid retention. "Students love Summer Bridge," says Roderick, because they know it helps them. Of the estimated 25,000 students who attended summer school last year, two-thirds moved to the next grade. Chicagoans call the policy "retention plus" because it comes front-loaded with ample resources for intervention and remediation. "Retention is a last resort," Vallas maintains.

Chicago also provides tutoring during school, and in an after-school Lighthouse Program (with supper included), for students who fail. Cozette Buckney, the system's chief education officer, shares the prevailing view of education experts that students should not merely repeat the same curriculum once they are held back. "You must teach them differently, use different materials-give them a different experience," she says.

Finally, if students have not passed the eighth-grade test by age 15, they move to "academic prep centers" that offer small transitional classes and intensive test preparation, where expenditures per pupil are one-and-a-half times those for high schools. Most students move on to high school after one year, although some teachers believe the centers accentuate problems of self-esteem and increase the tendency to drop out.

So dazzling is Chicago's success in the lower grades that outside observers have hardly noticed that real achievement stops at the high school door. "In the high schools, we have been at a loss," Buckney admits. Standards remain low, and there is widespread disengagement of students, a weak curriculum, and meager support services. Half of the city's ninth-graders fail two or more courses.

The Chicago Academic Standards Exams (CASE), which are end-of-semester high school tests in core subjects, are currently being upgraded. Teachers now receive detailed content guides from the administration but complain about the rigidity that the guides impose on their teaching. Although most teachers allow the exams to count for only 15 percent of the semester grade, a biology teacher contends that "the tests shape what I teach, what order I teach in, and how long I spend on each subject."



High school teachers have been more resistant to control from the central office than elementary school teachers, according to Vallas. "They view themselves as college professors—they're more set in their ways," he complains. Last year George Schmidt, an activist teacher, published parts of the CASE tests in protest, and students at top-rated Whitney Young High School boycotted the tests. Vallas dismisses such opposition, saying, "There is enough to be irritating but not enough to delay reform."

Recent efforts to close down and reconstitute Chicago's worst high schools have proved unsuccessful, and discharging low-performing administrators and teachers has been difficult. "The burden of proof [on the school administration] in removing failing teachers is pretty strong," Vallas admits.

Vallas is seeking to garner support for Chicago's high-stakes tests by allowing waivers and retesting. He is also identifying at-risk students early and conducting special programs for pregnant teens and teen mothers. "Good attendance, behavior, and grades" can help students get promoted, he says. The time will come, he predicts, when tests will become a diminishing factor in promotion decisions.

## Disappointment in New York

Chicago's success in strengthening standards and ending social promotion in the early grades contrasts sharply with New York City's recent dismal experience with high-stakes testing. New York's gigantic scale—it is the nation's largest school system with 1.1 million students—and the fractured relationship between the schools chancellor and the mayor have vastly complicated attempts to impose high stakes on state and city tests. Unlike Chicago, where Paul Vallas and Mayor Daley work in blissful harmony, New York City's Schools Chancellor Rudy Crew has recently been at loggerheads with Mayor Rudolph W. Giuliani.

High-stakes testing [has been brought] to the boiling point. After months of cramming and intense pressure on students, teachers, and parents, the news came in May [1999] that 67 percent of New York City students had failed the state's new and more demanding fourth-grade language arts test. The test was given over three days and included passages to be read for comprehension and answered in essay form. Stunned by the low test scores, the mayor proposed the removal of principals from the bottom third of all city elementary schools and called for a major management shake-up. State Education Commissioner Richard P. Mills recommended summer school for all students who failed the test.

Then came the disappointing results on year-end city tests, administered and graded by CTB/McGraw-Hill, which showed only 44.6 percent of students reading at grade level, a five-point decline from the previous year. Mathematics scores were even lower, falling 10 percent. After this second dose of bad news, the mayor prescribed even stronger medicine. Impressed by Chicago's example, he called for abolishing the semi-independent board of education and placing the schools under his own control.

After constant badgering from the mayor, Crew responded in mid-June with a hastily arranged mandatory summer-school program starting in July

for 37,000 third-, sixth-, and eighth-graders who had scored at or below the 15th percentile on the city's standardized reading test and the 10th percentile on the mathematics test, both taken in the spring. Students failing the tests a second time would be held back.

The summer-school program was plagued with problems. For six weeks the schools were forced to cope with thousands of youngsters who needed to pass the city tests to avoid retention. Many teachers were handicapped by a lack of student records and by inadequate course materials, and school buildings were stifling from a record-setting heat wave. Although paid at a lower rate than their school-year salaries, many teachers had to buy their own materials and bring fans from home. In some instances, students who were supposed to take only the mathematics test were drilled mainly for the reading test.

Chancellor Crew's pride in announcing that 64 percent of the students passed the tests was soon dampened when scoring errors by CTB/McGraw-Hill revealed that more than 8,600 students were sent to summer school by mistake. Lack of accurate summer-school attendance figures cast further doubt on the number of students who would be retained. The Puerto Rican Legal Defense Fund and Advocates for Children, a nonprofit legal services organization, have threatened court challenges, citing late warning of the new requirement to attend summer school and the use of a single test to determine promotion.

New York City's problems are likely to be exacerbated by Commissioner Mills' determined belief in high-stakes testing as a means of raising standards for the state's high schools. In a program that is unique in the nation, Mills and the State Board of Regents are requiring that by 2003, all students will have to receive a passing grade of at least 65 percent on the state's tough new Regents examinations in five subjects—English,mathematics, science, global studies, and American history—before they can graduate. Currently less than a quarter of New York City students qualify for a Regents diploma.

Turning aside his critics, Mills contended in a [recent] *New York Times* interview that without high standards, "You simply decide in advance that some students don't have access to the good life. They don't have access to jobs, they don't have access to enriched curriculum and everything that goes with it."

Mills understood the need for a structured program of remediation and support for those who fail the Regents tests. To this end, he proposed a $900 million program targeted toward poor districts, but Gov. George E. Pataki's budget came nowhere near to providing that amount.... [S]tate lawmakers—under pressure from parents, teachers, and school administrators who feared widespread failure on the Regents tests—argued for scaling down requirements. In October a consortium of parents and educators at 35 New York City alternative high schools asked legislators to compel the Regents to exempt its students from the new English Regents exams. With public opposition rising, it is doubtful the Regents will have enough public support to sustain such an extensive testing program with such high stakes over the long term.

As the results of [the] more rigorous six-hour, two-day English Regents tests were being released, accusations of deceptive scoring on the essay questions began to surface. According to a Harlem high-school teacher, "I never would have given points in a regular class for the kind of answers we were getting on those essays."

Many teachers had never seen the state's new standards; nor had they been trained to teach courses to the level of the Regents' demands. "No business or military organization would do that kind of campaign without adequate training," Thomas Sobol, former state education commissioner, asserted at a meeting in Purchase, N.Y.... For New York City's students, the stakes are overwhelming and probably unrealistic.



## Making High-Stakes Testing Work

High-stakes tests are transforming the education landscape, and lawmakers and educators are learning to navigate in uncharted terrain. Conditions and requirements vary state-by-state, and progress in meeting the new standards requires patience. But some early lessons can be drawn from states and school districts that are beginning to achieve success:

Make sure that learning—not testing-4s the goal "Are we teaching for testing or teaching for knowledge?" a senior administrator asks. Tests can be important in identifying weaknesses. But too much testing in too many subjects overwhelms teachers, drains resources from enriched educational programs, stifles creativity, and increases cheating.

Give disadvantaged students special assistance High-stakes tests can be a powerful tool for raising standards for at-risk students, but only if resources are reallocated to schools that serve them. And the testing program must be held accountable for ensuring that the tests are reliable, fair, and free of cultural bias.

Set failure rates at a realistic level Most schools lack the resources and capacity to absorb masses of failing students in after-school and summer-school remediation programs and to conduct programs for students who repeatedly fail. But setting failure rates too low damages credibility in the system's standards. The right balance will vary according to circumstances, but finding it is crucial.

Invest in a wide range of educational reforms—not just tests Tests don't work in a vacuum but in an environment that supports systemwide reform. Tests should be part of a program that encourages early childhood education; the recruitment, training, and development of capable teachers; smaller class size; and safer buildings.

Make retention a last resort Most studies show that retention does more harm than good. Frequent failure erodes self-confidence, and students who are retained have a higher probability of dropping out. If retention helps at all, it does so only when students are supported by innovative learning strategies.

Decisions to deny promotion should not be based on a single test and should involve the teacher.

Use publicity to force improvement School rankings draw attention to the weakest schools and can be used to drive decisions regarding school reform, reorganization, or closure. School officials have an obligation to interpret test results to the public consistently and accurately and to be forthright about problems in the system.

Focus on urban high schools Tests can be effective in raising standards but only if problems of school climate are addressed. Expect high school exit exams to be challenged in the courts by minority groups. Excessive testing narrows curriculum choice, and the need for remediation may lead to de facto tracking and high dropout rates.

Prepare for the long haul It is the rare state or school district that gets high-stakes testing right the first time. Success takes time and requires experimentation. Be ready to adapt, adjust, and compromise in order to achieve long-term success.

On balance, high-stakes tests that are well-designed and carefully administered appear to be working, at least in the lower grades. But if their benefits are oversold and their dangers ignored, disenchantment could lead to diminished support for public education. If, on the other hand, they call attention to failure and encourage strategies to ensure success, they could stimulate a long-overdue educational renewal for the nation's neediest students.

NO

**Martin G. Brooks and Jacqueline Grennon Brooks**

# The Courage to Be Constructivist

J^or years, the term *constructivism* appeared only in journals read primarily by philosophers, epistemologists, and psychologists. Nowadays, *constructivism* regularly appears in the teachers' manuals of textbook series, state education department curriculum frameworks, education reform literature, and education journals. Constructivism now has a face and a name in education.

A theory of learning that describes the central role that learners' ever-transforming mental schemes play in their cognitive growth, constructivism powerfully informs educational practice. Education, however, has deep roots in other theories of learning. This history constrains our capacity to embrace the central role of the learner in his or her own education.We must rethink the very foundations of schooling if we are to base our practice on our understandings of learners' needs.

One such foundational notion is that students will learn on demand. This bedrock belief is manifested in the traditional scope and sequence of a typical course of study and, more recently, in the new educational standards and assessments. This approach to schooling is grounded in the conviction that all students can and will learn the same material at the same time. For some students, this approach does indeed lead to the construction of knowledge. For others, however, it does not.

The people working directly with students are the ones who must adapt and adjust lessons on the basis of evolving needs. Constructivist educational practice cannot be realized without the classroom teacher's autonomous, ongoing, professional judgment. State education departments could and should support good educational practice. But too often they do not.

Their major flaw is their focus on high-stakes accountability systems and the ramifications of that focus on teachers and students. Rather than set standards for professional practice and the development of local capacity to enhance student learning, many state education departments have placed even greater weight on the same managerial equation that has failed repeatedly in the past: State Standards = State Tests; State Test Results = Student Achievement; Student Achievement = Rewards and Punishments.

We are not suggesting that educators should not be held accountable for their students' learning. We believe that they should. Unfortunately, we are not holding our profession accountable for learning, only for achievement on

From *Educational Leadership*, vol. 57, no.3, April 1991, pp. 18-24. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright © 2001 by ASCD. All rights reserved.

high-stakes tests. As we have learned from years of National Assessment of Educational Progress research, equating lasting student learning with test results is folly.

## The Emerging Research From Standards-Driven States

In recent years, many states have initiated comprehensive educational reform efforts. The systemic thinking that frames most standards-based reform efforts is delectably logical: Develop high standards for all students; align curriculum and instruction to these standards; construct assessments to measure whether all students are meeting the standards; equate test results with student learning; and reward schools whose students score well on the assessments and sanction schools whose students don't.

Predictably, this simple and linear approach to educational reform is sinking under the weight of its own flaws. It is too similar to earlier reform approaches, and it misses the point. Educational improvement is not accomplished through administrative or legislative mandate. It is accomplished through attention to the complicated, idiosyncratic, often paradoxical, and difficult to measure nature of learning.

A useful body of research is emerging from the states. With minor variations, the research indicates the following:

- Test scores are generally low on the first assessment relating to new standards. Virginia is an extreme example of this phenomenon: More than 95 percent of schools failed the state's first test. In New York, more than 50 percent of the state's 4th graders were deemed at risk of not graduating in 2007 after taking that state's new English language arts test in 1999.
- Failure, or the fear of failure, breeds success on subsequent tests. After the first administration of most state assessments, schools' scores rise because educators align curriculum closely with the assessments, and they focus classroom instruction directly on test-taking strategies.
- To increase the percentages of students passing the state assessments—and to keep schools off the states' lists of failing schools—local district spending on student remediation, student test-taking skills, and faculty preparation for the new assessments increases.
- Despite rising test scores in subsequent years, there is little or no evidence of increased student learning. A recent study by Kentucky's Office of Educational Accountability (Hambleton et al., 1995) suggests that test-score gains in that state are a function of students' increasing skills as test takers rather than evidence of increased learning.

## When Tests Constrict Learning

Learning is a complex process through which learners constantly change their internally constructed understandings of how their worlds function. New information either transforms their current beliefs—or doesn't. The efficacy of the learning environment is a function of many complex factors, including curriculum, instructional methodology, student motivation, and student developmental readiness. Trying to capture this complexity on paper-and-pencil assessments severely limits knowledge and expression.

Inevitably, schools reduce the curriculum to only that which is covered on tests, and this constriction limits student learning. So, too, does the undeviating, one-size-fits-all approach to teaching and assessment in many states that have crowned accountability king. Requiring all students to take the same courses and pass the same tests may hold political capital for legislators and state-level educational policymakers, but it contravenes what years of painstaking research tells us about student learning. In discussing the inordinate amount of time and energy devoted to preparing students to take and pass high-stakes tests, Angaran (1999) writes

> Ironically, all this activity prepares them for hours of passivity. This extended amount of seat time flies in the face of what we know about how children learn. Unfortunately, it does not seem to matter. It is, after all, the Information Age. The quest for more information drives us forward. (P. 72)

We are not saying that student success on state assessments and classroom practices designed to foster understanding are inherently contradictory. Teaching in ways that nurture students' quests to resolve cognitive conflict and conquer academic challenges fosters the creative problem solving that most states seek. However, classroom practices designed to prepare students for tests clearly do not foster deep learning that students apply to new situations. Instead, these practices train students to mimic learning on tests.

Many school districts question the philosophical underpinnings of the dominant test-teach-test model of education and are searching for broader ways for students to demonstrate their knowledge. However, the accountability component of the standards movement has caused many districts to abandon performance-based assessment practices and refocus instead on preparing students for paper-and-pencil tests. The consequences for districts and their students are too great if they don't.

## Constructivism in the Classroom

Learners control their learning. This simple truth lies at the heart of the constructivist approach to education.

As educators, we develop classroom practices and negotiate the curriculum to enhance the likelihood of student learning. But controlling what students learn is virtually impossible. The search for meaning takes a different route for each student. Even when educators structure classroom lessons and curriculums to ensure that all students learn the same concepts at the same time, each student still constructs his or her own unique meaning through his or her own cognitive processes. In other words, as educators we have great control over what we teach, but far less control over what students learn.

Shifting our priorities from ensuring that all students learn the same concepts to ensuring that we carefully analyze students' understandings to customize our teaching approaches is an essential step in educational reform that results in increased learning. Again, we must set standards for our own professional practice and free students from the anti-intellectual training that occurs under the banner of test preparation.

The search for understanding motivates students to learn. When students want to know more about an idea, a topic, or an entire discipline, they put more cognitive energy into classroom investigations and discussions and study more on their own.We have identified five central tenets of constructivism (Grennon Brooks & Brooks, 1993).

- First, constructivist teachers seek and value students' points of view. Knowing what students think about concepts helps teachers formulate classroom lessons and differentiate instruction on the basis of students' needs and interests.
- Second, constructivist teachers structure lessons to challenge students' suppositions. All students, whether they are 6 or 16 or 60, come to the classroom with life experiences that shape their views about how their worlds work. When educators permit students to construct knowledge that challenges their current suppositions, learning occurs. Only through asking students what they think they know and why they think they know it are we and they able to confront their suppositions.
- Third, constructivist teachers recognize that students must attach relevance to the curriculum. As students see relevance in their daily activities, their interest in learning grows.
- Fourth, constructivist teachers structure lessons around big ideas, not small bits of information. Exposing students to wholes first helps them determine the relevant parts as they refine their understandings of the wholes.
- Finally, constructivist teachers assess student learning in the context of daily classroom investigations, not as separate events. Students demonstrate their knowledge every day in a variety of ways. Defining understanding as only that which is capable of being measured by paper-and-pencil assessments administered under strict security perpetuates false and counterproductive myths about academia, intelligence, creativity, accountability, and knowledge.

## Opportunities for Constructing Meaning

Recently, we visited a classroom in which a teacher asked 7th graders to reflect on a poem. The teacher began the lesson by asking the students to interpret the first two lines. One student volunteered that the lines evoked an image of a dream. "No," he was told, "that's not what the author meant." Another student said that the poem reminded her of a voyage at sea. The teacher reminded the student that she was supposed to be thinking about the first two lines of the poem, not the whole poem, and then told her that the poem was not about the sea. Looking out at the class, the teacher asked, "Anyone else?" No other student raised a hand.



In another classroom, a teacher asked 9th graders to ponder the effect of temperature on muscle movement. Students had ice, buckets of water, gauges for measuring finger-grip strength, and other items to help them consider the relationship. The teacher asked a few framing questions, stated rules for handling materials safely, and then gave the students time to design their experiments. He posed different questions to different groups of students, depending on their activities and the conclusions that they seemed to be drawing. He continually asked students to elaborate or posed contradictions to their responses, even when they were correct.

As the end of the period neared, the students shared initial findings about their investigations and offered working hypotheses about the relationship between muscle movement and temperature. Several students asked to return later that day to continue working on their experiments.

Let's consider these two lessons. In one case, the lesson was not conducive to students' constructing deeper meaning. In the other case, it was. The 7th grade teacher communicated to her students that there is one interpretation of the poem's meaning, that she knew it, and that .only that interpretation was an acceptable response. The students' primary quest, then, was to figure out what the teacher thought of the poem.

The teacher spoke to her students in respectful tones, acknowledging each one by name and encouraging their responses. However, she politely and calmly rejected their ideas when they failed to conform to her views. She rejected one student's response as a misinterpretation. She dismissed another student's response because of a procedural error: The response focused on the whole poem, not on just the designated two lines.

After the teacher told these two students that they were wrong, none of the other students volunteered interpretations, even though the teacher encouraged more responses. The teacher then proceeded with the lesson by telling the students what the poet really meant. Because only two students offered comments during the lesson, the teacher told us that a separate test would inform her whether the other students understood the poem.

In the second lesson, the teacher withheld his thoughts intentionally to challenge students to develop their own hypotheses. Even when students' initial responses were correct, the teacher challenged their thinking, causing many students to question the correctness of their initial responses and to investigate the issue more deeply.

Very few students had awakened that morning thinking about the relationship between muscle movement and temperature. But, as the teacher helped students focus their emerging, somewhat disjointed musings into a structured investigation, their engagement grew. The teacher provoked the students to search for relevance in a relationship they hadn't yet considered by framing the investigation around one big concept, providing appropriate materials and general questions, and helping the students think through their own questions. Moreover, the teacher sought and valued his students' points

of view and used their comments to assess their learning. No separate testing event was required.

## What Constructivism Is and Isn't

As constructivism has gained support as an educational approach, two main criticisms have emerged. One critique of constructivism is that it is overly permissive. This critique suggests that constructivist teachers often abandon their curriculums to pursue the whims of their students. If, for example, most of the students in the aforementioned 9th grade science class wished to discuss the relationship between physical exercise and muscle movement rather than pursue the planned lesson, so be it. In math and science, critics are particularly concerned that teachers jettison basic information to permit students to think in overly broad mathematical and scientific terms.

The other critique of constructivist approaches to education is that they lack rigor. The concern here is that teachers cast aside the information, facts, and basic skills embedded in the curriculum—and necessary to pass high-stakes tests—in the pursuit of more capricious ideas. Critics would be concerned that in the 7th grade English lesson described previously, the importance of having students understand the one true main idea of the poem would fall prey to a discussion of their individual interpretations.

Both of these critiques are silly caricatures of what an evolving body of research tells us about learning. Battista (1999), speaking specifically of mathematics education, writes,

> Many ... conceive of constructivism as a pedagogical stance that entails a type of non-rigorous, intellectual anarchy that lets students pursue whatever interests them and invent and use any mathematical methods they wish, whether those methods are correct or not. Others take constructivism to be synonymous with "discovery learning" from the era of "new math," and still others see it as a way of teaching that focuses on using manipulatives or cooperative learning. None of these conceptions is correct. (P. 429)

Organizing a constructivist classroom is difficult work for the teacher and requires the rigorous intellectual commitment and perseverance of students. Constructivist teachers recognize that students bring their prior experiences with them to each school activity and that it is crucial to connect lessons to their students' experiential repertoires. Initial relevance and interest are largely a function of the learner's experiences, not of the teacher's planning. Therefore, it is educationally counterproductive to ignore students' suppositions and points of view. The 7th grade English lesson is largely nonintellectual. The 9th grade science lesson, modeled on how scientists make state-of-the-art science advancements, is much more intellectually rigorous.

Moreover, constructivist teachers keep relevant facts, information, and skills at the forefront of their lesson planning. They usually do this within the context of discussions about bigger ideas. For example, the dates, battles, and names associated with the U.S. Civil War have much more meaning for students when introduced within larger investigations of slavery, territorial expansion, and economics than when presented for memorization without a larger context.



State and local curriculums address *what* students learn. Constructivism, as an approach to education, addresses *how* students learn. The constructivist teacher, in mediating students' learning, blends the *what* with the *how*. As a 3rd grader in another classroom we visited wrote to his teacher, "You are like the North Star for the class. You don't tell us where to go, but you help us find our way." Constructivist classrooms demand far more from teachers and students than lockstep obeisance to prepackaged lessons.

## The Effects of High-Stakes Accountability

As we stated earlier, the standards movement has a grand flaw at the nexus of standards, accountability, and instructional practice. Instructional practices designed to help students construct meaning are being crowded out of the curriculum by practices designed to prepare students to score well on state assessments. The push for accountability is eclipsing the intent of standards and sound educational practice.

Let's look at the effects of high-stakes accountability systems. Originally, many states identified higher-order thinking as a goal of reform and promoted constructivist teaching practices to achieve this goal. In most states, however, policymakers dropped this goal or subsumed it into other goals because it was deemed too difficult to assess and quantify. Rich evidence relating to higher-order thinking is available daily in classrooms, but this evidence is not necessarily translatable to paper-and-pencil assessments. High-stakes accountability systems, therefore, tend to warp the original visions of reform.

Education is a holistic endeavor. Students' learning encompasses emerging understandings about themselves, their relationships, and their relative places in the world. In addition to academic achievement, students develop these understandings through nonacademic aspects of schooling, such as clubs, sports, community service, music, arts, and theater. However, only that which is academic and easily measurable gets assessed, and only that which is assessed is subject to rewards and punishments. Jones and Whitford (1997) point out that Kentucky's original educational renewal initiative included student selfsufficiency and responsible group membership as goals, but these goals were dropped because they were deemed too difficult to assess and not sufficiently academic.

Schools operating in high-stakes accountability systems typically move attention away from principles of learning, student-centered curriculum, and constructivist teaching practices. They focus instead on obtaining higher test scores, despite research showing that higher test scores are not necessarily indicative of increased student learning.

Historically, many educators have considered multiple-choice tests to be the most valid and reliable form of assessment—and also the narrowest form of assessment. Therefore, despite the initial commitment of many states to per-

formance assessment, which was to have been the cornerstone of state assessment efforts aligned with broader curriculum and constructivist instructional practices, multiple-choice questions have instead remained the coin of the realm. As Jones and Whitford (1997) write about Kentucky,

> The logic is clear. The more open and performance based an assessment is, the more variety in the responses; the more variety in the responses, the more judgment is needed in scoring; the more judgment in scoring, the lower the reliability.... At this point, multiple-choice items have been reintroduced, performance events discontinued. (P. 278)

Ironically, as state departments of education and local newspapers hold schools increasingly accountable for their test results, local school officials press state education departments for greater guidance about material to be included on the states' tests. This phenomenon emboldens state education departments to take an even greater role in curriculum development, as well as in other decisions typically handled at the local level, such as granting high school diplomas, determining professional development requirements for teachers, making special education placements, and intervening academically for at-risk students. According to Jones and Whitford (1997),

> [In Kentucky] there has been a rebound effect. Pressure generated by the state test for high stakes accountability has led school-based educators to pressure the state to be more explicit about content that will be tested. This in turn constrains local school decision making about curriculum. This dialectical process works to increase the state control of local curriculum. (P. 278)

## Toward Educational Reform

Serious educational reform targets cognitive changes in students' thinking. Perceived educational reform targets numerical changes in students' test scores. Our obsession with the perception of reform, what Ohanian (1999) calls "the mirage theory of education," is undermining the possibility of serious reform.

History tells us that it is likely that students' scores on state assessments will rise steadily over the next decade and that meaningful indexes of student learning generally will remain flat. It is also likely that teachers, especially those teaching in the grades in which high-stakes assessments are administered, will continue to narrow their curriculum to match what is covered on the assessments and to use instructional practices designed to place testing information directly in their students' heads.

We counsel advocacy for children. And vision. And courage.

Focus on student learning. When we design instructional practices to help students construct knowledge, students learn. This is our calling as educators.

Keep the curriculum conceptual. Narrowing curriculum to match what is covered on state assessments results in an overemphasis on the rote memorization of discrete bits of information and pushes aside big ideas and intellectual curiosity. Keep essential principles and recurring concepts at the center.

Assess student learning within the context of daily instruction. Use students' daily work, points of view, suppositions, projects, and demonstrations to assess what they know and don't know, and use these assessments to guide teaching.

Initiate discussions among administrators, teachers, parents, school boards, and students about the relationship among the state's standards, the state's assessments, and your district's mission. Ask questions about what the assessments actually assess, the instructional practices advocated by your district, and the ways to teach a conceptual curriculum while preparing students for the assessments. These are discussions worth having.

Understand the purposes of accountability. Who wants it, and why? Who is being held accountable, and for what? How are data being used or misused? What are the consequences of accountability for all students, especially for specific groups, such as special education students and English language learners?

Students must be permitted the freedom to think, to question, to reflect, and to interact with ideas, objects, and others—in other words, to construct meaning. In school, being wrong has always carried negative consequences for students. Sadly, in this climate of increasing accountability, being wrong carries even more severe consequences. But being wrong is often the first step on the path to greater understanding.

We observed a 5th grade teacher return a test from the previous day. Question 3 was, "There are 7 blue chips and 3 green chips in a bag. If you place your hand in the bag and pull out 1 chip, what is the probability that you will get a green chip?" One student wrote,"You probably won't get one." She was "right"-and also "wrong." She received no credit for the question.

## References


Angaran, J. (1999, March). Reflection in an age of assessment. *Educational Leadership, 56,* 71-72.

Battista, M. T. (1999, February). The mathematical miseducation of America's youth: Ignoring research and scientiflce study in education. *Phi Delta Kappan 80(6),* 424-433.

Grennon Brooks, J., & Brooks, M. G. (1993). *In search of understanding: The case for constructivist classrooms.* Alexandria, VA: ASCD.

Hambleton, R., Jaeger, R. M., Koretz, D., Linn, R. L, Millman, J., & Phillips, S. E. (1995). *Review of the measurement quality of the Kentucky instructional results information system 1991-1994.* (Report prepared for the Kentucky General Assembly.) Frankfort, KY: Office of Educational Accountability.

Jones, K., & Whitford, B. L. (1997, December). Kentucky's conflicting reform principles: High stakes accountability and student performance assessment *Phi Delta Kappan,* 78(4), 276-281.

Ohanian, S. (1999). *One size fits few.* Portsmouth, NH: Heinemann.

# POSTSCRIPT

## Do High-Stakes Assessments Improve Learning?

Whether standardized tests are a crucial tool in improving overall student performance or whether they rob teachers of the autonomy and creativity needed for lasting improvement of learning is among the most hotly debated topics on the current scene. Is Jerry Jesness correct in condemning "floating" standards that shield the status quo and guarantee the reign of mediocrity, or is Susan Ohanian right in demolishing high-stakes testing in her book *One Size Fits Few: The Folly of Educational Standards* (1999)?

Here are some further sources to tilt your thinking one way or the other: Mary E. Diez, "Assessment as a Lever in Education Reform," *National Forum* (Winter 1997); Elliot W. Eisner, "Standards for American Schools: Help or Hindrance?" *Phi Delta Kappan* (June 1995); Jack Kaufhold, "What's Wrong With Teaching for the Test?" *The School Administrator* (December 1998); Frederick M. Hess and Frederick Brigham, "None of the Above," *American School Board Journal* (January 2000); Jeff Berger, "Does Top-Down, Standards-Based Reform Work?" *NASSP Bulletin* (January 2000); Peter Schrag, "High Stakes Are for Tomatoes," *The Atlantic Monthly* (August 2000); Georgia Hedrick, "Real Teachers Don't Test," *Educational Horizons* (Winter 2002); Dale DeCesare, "How High Are the Stakes in High-Stakes Testing?" *Principal* Qanuary 2002); Mary Ann Raywid, "Accountability:What'sWorth Measuring?" *Phi Delta Kappan* (February 2002); Mark F. Goldberg, "The Test Mess," *Phi Delta Kappan* (January 2004); and Michael G. Gunzenhauser, "High-Stakes Testing and the Default Philosophy of Education," *Theory Into Practice* (Winter 2003).

Although most of the focus is now on state mandates, proposals for national testing are still under consideration. Long a practice in many foreign countries, high-stakes national examinations in valued subject matter areas are an explosive topic in the United States. A sampling of opinion can be found in "Yes to National Tests," by Diane Ravitch, *Forbes* (May 5, 1997); "Getting Testy," *The New Republic* (September 29, 1997); and "National Tests Are Unnecessary and Harmful," by Monty Neill, *Educational Leadership* (March 1998).

Multiple articles on the standards and testing movement can be located in the May 1999 issue of *Phi Delta Kappan,* the Winter 1999 issue of *Kappa Delta Pi Record,* the February 2000 issue *of Educational Leadership,* the January 2001 issue of *NASSP Bulletin,* the January 2001 issue of *Principal Leadership,* the December 2001 issue of *The School Administrator,* the February 2002 issue of *Phi Delta Kappan,* the Winter 2002 issue of *Kappa Delta Pi Record,* and the September 2003 issue *of American School Board Journal.*