

Are We Measuring Student Success with High-Stakes Testing?

by Kathleen Anderson Steeves,
Jessica Hodgson,
and Patricia Peterson

A researcher, whether scientific, political, or educational, will tell you that evaluative measures begin with a working theory of the expected outcome. If high-stakes standardized exams and tests of similar ilk are intended to measure student success, why then have we not seen more discussion and careful thought placed on understanding what the outcomes of such testing actually mean and say about our students? What, in effect, do such tests actually set out to measure in our children?

Should we not, as teachers, administrators, and community leaders, be more worried about the measures and results of these annual "intelligence" tests that are increasingly being used to label students? FairTest (2002) recently noted that the United States already tests more children more often than any other nation. As a FairTest (1998) survey concluded, "There are more regulations governing the food we feed our pets than the tests used to make decisions about our children."

Should we not, as educators, insist on the time to step back and answer some fundamental questions? What is a successful student? What constitutes successful student habits, knowledge, abilities, and work? How can high-stakes testing be used more appropriately as a powerful tool for teachers to assist student learning?

Should we not open the debate about whether high-stakes tests, as they currently exist in most states, are accurate measures of what we want our students to be able to do before grade promotion or graduation? Do we want to stay on a path of instruction in which decisions are made about student intelligence, teacher ability, and school value based on a one-time test of general knowledge that has not been clearly defined or examined? As a country adopting more standardized testing, should we not see if its outcomes meet its admittedly high goals?

As educators who see the real effects of the tests on a daily basis, we believe that redefining and reexamination is overdue. Only after reevaluation and setting the parameters for student success can we go on to say definitively that such tests adequately measure the intended variables. This reevaluation must be undertaken before current testing trends, which may label students unfairly and inaccurately, become further entrenched in our society as the all-encompassing measure of a good student.

HISTORICAL DEFINITIONS OF THE SUCCESSFUL STUDENT

Though the specific details of student success have varied throughout the history

of education—from the Puritan’s scripture-citing prodigy to the Industrial Revolution’s well-oiled cog of a learner—certain variables have remained intact (McNergney and Herbert 1995). As Goodlad (1984, 36) stated, there are a “broad array of educational goals in four [predominant] areas that have emerged in this country over more than 300 years.” Though not nationally recognized, these four areas have become educationally institutionalized. They represent a logical first step in the search for the “successful” student, including, as Goodlad (1984, 51) noted, “A. Academic Goals (including: 1. Mastery of Basic Skills and Fundamental Processes, and 2. Intellectual Development); B. Vocational Goals; C. Social, Civic, and Cultural Goals; and D. Personal Goals.”

According to Goodlad (1984), both parents and teachers have ranked academic goals of schooling as their top priority. These goals consisted of enabling students to: read, write, and solve basic arithmetic problems; acquire ideas from reading and listening; learn to communicate ideas through writing and speaking; utilize available sources of information; employ problem-solving skills, principles of logic, and different modes of inquiry; use and evaluate knowledge through critical and independent skills; accumulate a general fund of knowledge; and develop a positive attitude toward intellectual development.

Of the eight sub-skills constituting academic success, only one—enabling students to “accumulate a general fund of knowledge”—suggests that student success is connected to the mere “knowing”

Kathleen Anderson Steeves is Associate Professor of History/Social Studies Education at George Washington University in Washington, D.C. Her research interests include secondary history and social studies, education reform and standards, teacher leadership, and the history of education.

Jessica Hodgson teaches social studies at Chantilly High School in Fairfax County, Virginia. She is currently involved in research on using writing as a learning tool in the history classroom.

Patricia Peterson teaches 7th- and 8th-Grade English at Garnet-Patterson Middle School in Washington, D.C. Her focus is on literacy development through poetry, music, and service learning. Ms. Peterson recently established a student Poetry Club connecting students with community artists and writers.

(or testing) of general information. Instead, Goodlad's goals point to a student's ability to "do" as the keystone to academic success. Words like "acquire," "problem-solve," "utilize," "analyze," and "inquire" rest at the helm of a student's successful navigation and eventual completion of school.

Currently, standards in all disciplines are academically rigorous. In describing their own standards for student success, national learning organizations have defined success in terms of what the student can "do":

- Apply reflective thinking and decision making when analyzing current civic/social events (National Council for the Social Studies 1993).
- Develop the abilities that characterize science as inquiry (Center for Science, Mathematics, and Engineering Education 1996).
- Investigate, make sense of, and construct meanings from new situations; make and provide arguments for conjectures; and use a flexible set of strategies to solve problems from both within and outside mathematics (National Council of Teachers of Mathematics 1989).
- Apply a wide range of strategies to comprehend, interpret, evaluate, and appreciate texts, including drawing on prior experience, interactions with other readers and writers, knowledge of word meaning of other texts (National Council of Teachers of English 1996).
- Use English to communicate in social settings, to achieve academically in all content areas and in socially and culturally appropriate ways (Teachers of English to Speakers of Other Languages 1997).

Indeed, teacher conceptions of the "successful" student appear to be aligned with those of the community, parents, and employers, each defining success as a student's ability to solve problems, seek solutions, and create and understand

meaningful texts. A research study of employers in Oxon Hill, Maryland (Walker 1999), found that employers, like teachers, defined a successful student as one able to acquire and apply knowledge, such as: 1) locating and using meaningful information, 2) deciphering written and spoken language, 3) employing English in a context-appropriate manner, and 4) working with others in a professional manner.

Members of the National Education Goals Panel (1994) imagined successful students as those "practicing the scientific method, solving problems as a group, analyzing data, expressing their findings in writing, and defending their analysis in discussion." Educators such as Theodore R.Sizer and Ernest Boyer have envisioned similar learning environments and roles for the "successful" student. For Sizer (1992, 72), "The residue of serious learning is a mixture of awareness and logic. One exercises qualities of the mind with specifics, but the qualities themselves are the end to be pursued." In short, the goal of learning for Sizer is not the knowledge garnered but rather the processing of learning to garner it. As Boyer (in Fiske 1991, 65), former president of the Carnegie Foundation for the Advancement of Teaching, has argued, "The key to problem solving is not to figure out the answer to a question that someone else hands you, but to define the right problem. An educated person today is someone who knows the right questions to ask." When so many seem to agree on the current definition of a successful student, what's the problem?

HIGH-STAKES TESTING IN PRACTICE

Since U.S. educators first invented the modern standardized norm-referenced test, as Debra Meier (2000, 25) recently declared:

Our students have been taking more tests more often than any nation

on the face of the earth, and schools and districts have been going public with test scores starting almost from the moment children enter school. From the third- or fourth-grade level (long before any of our international counterparts bother to test children) we have test data for virtually all schools, by race, class, and gender. We know exactly how many kids did better or worse in every subcategory. We have test data for almost every grade thereafter in reading and math, and to some degree in all other subjects. This has been the case for nearly half a century.

Current trends in testing, though still tentative because there has not been enough time for in-depth study, include: use of multiple-measure assessments, emphasis on reading, and some sort of transitional accountability system as an interim step to mandatory state testing (Anderson, Fiester, Gonzales, and Pechman 1996). According to FairTest (2002), as of 1999, 17 states had graduation tests and five more were planning to implement such tests. Of the 15 states south of the Mason-Dixon line, 11 required students to pass graduation tests. In all, 39 states give either a criterion-referenced or norm-referenced test to determine their students' proficiency level. Almost all states are in the process of implementing proficiency tests of some kind, a process required now by the new federal education legislation. Illustrations from several jurisdictions provide a snapshot of some of the issues for teachers and students in this high-stakes testing trend.

In the District of Columbia, principals' and teachers' evaluations are initially to be based on student scores and improvement on the Stanford-9 (SAT-9) test battery, a multiple choice test adopted in 1996 to measure student achievement in reading and mathematics. A committee that in-

cluded several teachers argued that the SAT-9 best matched the district's public schools curriculum, which is now aligned with national standards. The SAT-9 is available with various sets of empirical, normative information; ironically, none of the individual tests apply specifically to urban school districts (Harcourt Brace Educational Measurement 1999).

In Maryland, the state's School Performance Assessment Program (MSPAP), established in 1990 as the result of the Governor's Commission on School Performance, tests students' mastery in math, reading, and writing. Tasks require students to respond to questions or directions that lead to a solution to a problem, a recommendation, a decision, an explanation, or a rationale for the students' response to the task or question (Maryland State Department of Education [MSDOE] 1998). Once hailed for offering one of the first "performance-based" school tests in the nation, few states have followed Maryland's lead with MSPAP (Argetsinger and Nakashima 1998), finding such tests to be too expensive and time-consuming. MSPAP critics have complained that the test is used only to rate schools and does not measure individual student performance. An analysis of the 2001 MSPAP scores for the eighth-grade exam showed a dramatic decline in scores that left educators without an explanation. As a result, schools were given an option as to whether to use it or not. Questions were raised about the test's validity and the reliance on one test as a measure of teacher and school success (Shulte 2002). Starting with the class of 2004, high school students will be required to pass the Maryland Functional tests in content courses as a requirement for graduation (MSDOE 1998).

Massachusetts has spent \$20 million on the development of the Massachusetts Comprehensive Assistance System, which

will be required of all tenth-graders beginning in 2003. The state test will include multiple-choice and open-response questions. A newly formed group, the Coalition for Authentic Reform in Education, is actively opposing the notion that one test should be the only measure of student achievement (Public Broadcasting Service [PBS] 2002).

In Virginia, public schools began using the Standards of Learning (SOL) tests to assess students and schools in 1998. Students are tested in English, mathematics, history/social studies, science, and computer/technology in grades 3, 5, and 8 and at the end of core courses in high school. Beginning with the class of 2004, passing the SOL tests will be a determinant of earning a high school diploma. In addition, by 2006, a school's accreditation will depend on a 70 percent student-passing percentage on the SOL tests (Virginia Department of Education [VDOE] 1999–2001). In the second year of testing, 93 percent of public schools failed to reach the approved benchmark. The state was considering ways to adjust its timetable to implement sanctions (Mathews 1999). At this point, about half of the schools have achieved a passing rate on the required core SOL exams—though ESL students' scores were not included (VDOE 2001). In an effort to comply with expected requirements of President Bush's (2001) "No Child Left Behind" federally backed education legislation, Virginia is moving to expand the SOL testing to grades 4, 6, and 7.

HIGH-STAKES TESTING MAY FAIL STUDENTS AND TEACHERS

If these high-stakes tests are to be the measure of students, as well as of schools, it is imperative that we assess the effect of such testing on student academic success. When our country looks blindly to one test to tell us all we hope to know about stu-

dents, teachers, and schools, then educators must demand a closer look at the actual meaning of these test results. Many educators know the flaws in the system firsthand. For example, when important consequences—such as student graduation and school accreditation—are linked to high-stakes testing, teachers are likely to respond by "teaching to the test" (Gordon and Reese 1997, 346). To ensure that students pass the assessment, teachers cover material that they believe will most likely be included on the test. However, teaching to these high-stakes tests has several negative consequences for both teachers and students.

When teaching to the test, the test does not become an assessment of a student's mastery of content; it is, instead, a powerful curricular tool. The teacher is forced to make instructional decisions that are not based on prior professional experience, what is of academic importance, or what is in the best interest of the student; instead, decisions are based on what is most likely to be included on a standardized test (Shepard 1991). In essence, the test becomes a teacher's filter for making instructional decisions. As the curriculum becomes more narrow, content and skills that are not on the standardized assessment are eliminated. In fact, teachers feel pressure to make sure classroom activities correspond to material on the assessment even though they may know other materials will better prepare students for success in the world.

Gordon and Reese (1997) conducted a study on the effects of high-stakes testing on teachers in Texas. In this case, the high-stakes assessment became the object of instruction rather than the outcome of previous instruction. The Texas Assessment of Academic Skills (TAAS) evaluates schools based on the results of the assessment and gives cash awards to schools with high test scores, while poorly performing schools

face “sanctions and intervention” (Gordon and Reese 1997, 348). These researchers found that one way Texas teachers have responded to high-stakes testing is to emphasize skills on the test. Teachers tended to spend more time on “drill and practice” and less time on “hands-on activities” found in the curriculum standards. Teachers in Texas complained that the TAAS made them accountable in terms of teaching TAAS-related content, but it did not make them accountable in terms of being effective teachers (Gordon and Reese 1997). Critics of the state’s testing procedures have noted that dropout rates have increased, as has cheating (PBS 2002).

Gary Natriello (in Benning 1999), a professor at Columbia University’s Teachers College, has argued that the new emphasis on testing may discourage teachers from entering the field because they really want to shape young minds rather than raise a test score by a point or two. Furthermore, schools—especially in urban settings—already have difficulty filling teaching positions. If the broad application of high-stakes test scores demoralizes the teachers currently working, and discourages potential teachers from applying, student achievement ultimately suffers.

Students are likewise affected by high-stakes testing. One of the most troubling effects of high-stakes testing is that it “can force students to leave school before they have to take the examination, or after failing it” (Madaus 1991, 229). Another effect is that the performance of principals and superintendents is often linked to the number of students who reach the bar set by the state. Unfortunately, this situation may mean it is

in an administrator’s best interest to eliminate systematically from the test those students who probably wouldn’t make the bar and to keep enrolled discipline-problem students who will get high scores on the test. In fact, in efforts to increase school performance on high-stakes testing, some schools have relied on increased special education placement and retention in grade. This system allows schools to control, to some extent, the student population whose test scores are reflected in school scores (Allington and McGill-Franzen 1992).

Whatever philosophical perspective one has toward high-stakes testing, it is clear that these tests have a significant effect on both teachers and students. In narrowing the curriculum, changing instructional strategies, and impacting the classroom and student demographics, high-stakes testing negatively alters the educational environment for teachers and students.

Overall, educators understand that the emphasis these tests place on lower-level skills is not in line with what discipline-based organizations are asking their teachers to do in meeting content standards. Given the increased focus on nationwide standards, it must be asked whether schools *are* supplying their students with the skills and knowledge necessary to become “successful.” There is no evidence that passing the current tests equates to student success. Actually, for many, it may even mean the opposite. These questions must then be asked: Are we producing students who can problem-solve, analyze, and ask the right questions? Are we evaluating them accordingly? Are these popular and

There is no evidence that passing the current tests equates to student success. Actually, for many, it may even mean the opposite.

increasingly endorsed “high-stakes” tests, which neatly tell us who has succeeded and who has failed, attuned to our definition of a successful student or to the philosophy of success that shapes current practice?

RETHINKING HOW WE MEASURE SUCCESS

With the current rush toward adoption of high-stakes testing forms that align with state standards, we must encourage important and necessary public discussion about the outcomes and what we are actually trying to assess in our students today (U.S. Department of Education 1998). As schools hurriedly move toward the narrow use of standardized tests to determine eligibility for graduation and school and teacher evaluation, we argue that it still has not been made clear what these tests really measure and if they actually come close to identifying our society’s agreed-upon definition of a successful student.

It is necessary to reopen the discussion on testing, not as a measure to be removed, but as something that must be reexamined for what test results actually do mean and how the billions in testing dollars could be spent best in assessing our students. We must begin a discussion about what a successful student is and whether a revamped test or a restructured system of accounting for the results could more accurately measure this success.

As part of the call to reenergize this important discussion, various issues must be considered. The discussion has begun in a number of states, but it must be expanded.

How do we increase the validity of

these tests? Most research employs a margin of error. Why is that not true in high-stakes tests? A student scoring one point below the cut-off score (defined in no logical way) gets no credit, while the student scoring at the cut-off point or one point above gets full credit and is promoted.

Which standards are we using to measure success? All content areas include important skills and knowledge, but what consistent measure of comparison is being used (state, county, district, or national standards)?

What are the long-term effects of high-stakes testing on students who do not graduate or drop out before they fail yet again? Do we have the collective will across the board in urban, suburban, and rural schools to use the test data to make a difference and assist students who need additional instructional time and resources to achieve the standards?

In the end, we must ask one another if we have developed a clear, agreed-upon definition of a successful student so that we may really put required resources into achieving success for all students. We must be certain that the promise to “leave no child behind” is not just overlooked by current community leaders and politicians who herald the tests unchanged. We must insist that high-stakes, standardized testing becomes less a political tool for finger pointing and assessing student and school failure and more of a reasoned way to help pinpoint why these students are failing and where resources are needed to produce the truly successful students we all want!

REFERENCES

- Allington, R. L., and A. McGill-Franzen. 1992. Does high-stakes testing improve school effectiveness? *ERS Spectrum* 10(2): 3–12.
- Anderson, L., L. Fiester, M. Gonzales, and E. Pechman, eds. 1996. *Improving America's schools: Newsletter on issues in school reform* [Spring]. Washington, D.C.: Improving America's Schools Association. Available at: <http://www.ed.gov/pubs/IASA/newsletters/standards>.
- Argetsinger, A., and E. Nakashima. 1998. In Md., the ‘bubble’ test has burst: Analytical exam gains favor over mul-

- multiple-choice format. *Washington Post*, 11 May, A1.
- Benning, V. 1999. Teachers wary of new exams: Poll finds support for D.C. program. *Washington Post*, 5 July, A1.
- Bush, G. W. 2001. No child left behind. Washington, D.C.: U.S. Department of Education. Available at: <http://www.ed.gov/offices/OESE/esca/nclb/titlepage.html>.
- Center for Science, Mathematics, and Engineering Education. 1996. *National science education standards*. Washington, D.C.: National Academy Press.
- FairTest: The National Center for Fair & Open Testing. 1998. FairTest applauds National Research Council report opposing high-stakes tests and endorsing tougher regulation of exam use. Press release, 3 Sept. Cambridge, Mass.: FairTest. Available at: <http://www.fairtest.org/pr/nrc-pr.htm>.
- FairTest: The National Center for Fair & Open Testing. 2002. Will more testing improve schools? Fact sheet. Cambridge, Mass.: FairTest. Available at: <http://www.fairtest.org/facts/Will%20More%20Testing%20Improve%20Schools.html>.
- Fiske, E. B. 1991. *Smart schools, smart kids: Why do some schools work?* New York: Simon & Schuster.
- Goodlad, J. I. 1984. *A place called school: Prospects for the future*. New York: McGraw-Hill.
- Gordon, S. P., and M. Reese. 1997. High-stakes testing: Worth the price? *Journal of School Leadership* 7(4): 345-68.
- Harcourt Brace Educational Measurement. 1999. Stanford-9 Achievement Test Series: It's like nothing you've ever seen before. San Antonio: HBEM.
- Madaus, G. F. 1991. The effects of important tests on students. *Phi Delta Kappan* 73(3): 226-31.
- Maryland State Department of Education. 1998. Maryland school performance report, 1998: State and school systems. Annapolis: MSDOE. ERIC ED 429 348.
- Mathews, J. 1999. Va. schools may get reprieve on standards. *Washington Post*, 16 September, B1.
- McNergney, R. F., and J. M. Herbert. 1995. *Foundations of education: The challenge of professional practice*. Boston: Allyn & Bacon.
- Meier, D. 2000. *Will standards save public education?* Boston: Beacon Press.
- National Council for the Social Studies. 1993. Social studies online: Standards and position statements. Washington, D.C.: NCSS.
- National Council of Teachers of English. 1996. *Standards for the English language arts*. Urbana, Ill.: NCTE and the International Reading Association.
- National Council of Teachers of Mathematics. 1989. *Curriculum standards for grades 9-12*. Reston, Va.: NCTM. Available at: <http://standards.nctm.org/Previous/CurrEvStds/currstand9-12.htm>.
- Public Broadcasting Service. 2002. Frontline: Testing our schools. Television program, 28 March.
- Shepard, L. A. 1991. Will national tests improve student learning? *Phi Delta Kappan* 73(3): 232-38.
- Shulte, B. 2002. County questions drop in MSPAP scores: Despite new initiatives, district's results decline. *Washington Post*, 31 January, T14.
- Sizer, T. R. 1992. *Horace's school: Redesigning the American high school*. Boston: Houghton Mifflin.
- Teachers of English to Speakers of Other Languages. 1997. ESL standards for pre-K-12 students. Alexandria, Va.: TESOL. ERIC ED 420 991.
- U.S. Department of Education. 1998. Turning around low-performing schools: A guide for state and local leaders. Washington, D.C.: USDE. ERIC ED 420 119.
- Virginia Department of Education. 1999-2001. Standards of learning discussion forum: SOLs unfair to students. Richmond: VDOE. Available at: <http://www.vdoe.vipnet.org/messages/23/36.html?996190534>.
- Virginia Department of Education. 2001. Spring 1999 Standards of Learning test results. Richmond: VDOE. Available at: <http://www.pen.k12.va.us/sol99>.
- Walker, L. 1999. A community case study. GWU class, Teacher Education, SPED #233, Professor West, Washington, D.C.



© Kappa Delta Pi