

Special Section Implementing Standards In Schools

Standards Are Not Enough: Essential Transformations for School Success

Douglas B. Reeves

This article discusses five transformations necessary to bring standards from theory into reality. These changes, though difficult to elicit, will have a positive and productive effect when made collaboratively and thoughtfully. The concept of power standards, a subset of standards that involves thoughtful focus, is proposed to ensure successful educational practice and improved test performance.

Almost every public school district in the nation has academic standards. Although standards alone are clearly an insufficient instrument for the improvement of student achievement, the essence of standards—the clear articulation of what students should know and be able to do—forms the basis for the essential transformations necessary for school success. In this article I discuss five essential transformations necessary to bring standards from theory to reality:

1. A change in instructional strategies from the extremes of mindless test preparation and unsupervised anarchy to a consistent requirement for thinking, reasoning, and communication proficiency for all students.
2. A change from the guesswork and alchemy surrounding letter grades to the use of comprehensive diagnostic assessments before and during high school.
3. A change in curriculum from frantic coverage of every standard to the use of “power standards.”
4. A change from assigning teachers to buildings and courses on the basis of seniority to a more equitable assignment policy that is blind to the economic status and skin color of students.
5. A change from high-stakes tests as the sole criterion for critical decisions to the use of a body of evidence for high-stakes decisions.

These changes strike at the heart of some of the most hallowed traditions of secondary education. But I will be clear: The choice is not whether these changes will occur, but whether they will be made collaboratively and thoughtfully or through a haphazard combination of politics, state intervention, and litigation. For those who favor the former process of change, the following ideas may have some value (see figure 1).

Transformation 1. From Test Prep to Thinking, Reasoning, and Writing

Many individuals believe that the appetite of politicians for higher test scores can be satisfied only when teachers engage in mindless test preparation at the expense of thinking, reasoning, and writing. Consider the words of standards critic Alfie Kohn (1999):

Not long ago, I am told, a widely respected middle school teacher in Wisconsin, famous for helping students design their own innovative learning projects, stood up at a community meeting and announced that he “used to be” a good teacher. These days, he explained, he just handed out textbooks and quizzed his students on what they had memorized. The reason was very simple. He and his colleagues were increasingly being held accountable for raising test scores. The kind of wide-ranging and enthusiastic exploration of ideas that once characterized his classroom could no longer survive when the emphasis was on preparing students to take a standardized examination. Because the purveyors of Tougher Standards had won, the students had lost. (73)

Many educators share experiences similar to those addressed by Kohn, and this anecdotal evidence suggests that some administrators and teachers embrace a regime of test preparation through the repetitious use of mock test items. This leads to the widespread conclusion that the very existence of standards and associated tests leads to bad decisions by administrators about teaching strategies. At one end of the rhetorical continuum are the fearful administrators demanding a curriculum laden with test preparation activities. At the other end are those who contend that no supervision is appropriate and echo the mantra, “Just leave me alone and let me teach!” Consider the words of Susan Ohanian (1999):

What I do know is that across the land teachers are bowing to the pressure of corporate-political-infotainment pronouncements: They are eliminating recess and putting away the building blocks, the tempera paints, and the picture books that don’t introduce phonemes in the sequence chosen by the publishing conglomerate so venerated in Texas and California. They are bringing out the skill drill worksheets that will get every kid in America learning the schwa on schedule. (19)

These comments beg the question, Does test preparation (along with the subsequent abandonment of thinking, reasoning, and communication skills) improve test scores? If the answer to this question is yes, then there ought to be data to support such a notion. Instead, the key to better multiple choice scores lies in a focused, multidisciplinary requirement for students to think, reason, and write in a clear, accurate, and persuasive manner.

If educators could step back from the rhetorical wars surrounding high-stakes testing, it should become obvious that thinking rather than memorization is the key to better test results. An examination of case-study data at the classroom, school, district, and state level indicates that when students write more frequently and when they score higher on writing performance assessments, their scores on multiple-choice tests improve. The association between writing and multiple-choice scores ranges from .7 to .9 (see figure 2). Other case studies in Riverview Gardens, Mo., Milwaukee, Wisc., and Orange County, Calif., support this quantitative evidence. These case studies in a variety of school systems including those with high levels of poor and minority students confirm a strong relationship between an increased frequency of writing performance assessments

and improved test scores. More writing in the classroom may, indeed, take time away from test preparation, but it does not hinder test scores.

Does an emphasis on more writing and its attendant time demands diminish other subjects in the curriculum? The evidence from Riverview Gardens, Mo., school district suggests the opposite. The Write Focus program comprises collaboratively scored writing assessments in every class, including physical education, music, and all other disciplines, on a monthly basis. Using a common rubric or scoring guide, the district tripled the number of students who were proficient or better in writing. According to an editorial in the St. Louis Post-Dispatch (2000), this “diversion” of time and effort from other subjects had an interesting consequence: dramatic increases in state test scores in science and social studies. Similar associations between strong writing programs and cross-academic improvements can be found in Orange County, Calif. (Groves 2000); Milwaukee, Wisc.; and Indianapolis, Ind. In Wayne Township in Indiana, for example, there is a common writing scoring guide for grades 7 through 12 so that students from science to sociology and mathematics to music receive consistent messages on the importance of expressing their reasoning and thinking in appropriate written form. These model districts make it clear that reasoning, thinking, and communications—not mindless test drills—are the keys to improved student achievement.

The administrators who demand test preparation are wrong; the standards critics who assert that only terrible teaching practices will lead to higher multiple-choice test scores are also wrong. The positive impact of writing, thinking, and reasoning on test scores is no surprise to secondary school educators who have successfully prepared students for Advanced Placement (AP) and International Baccalaureate (IB) exams for years. The rigor of those classes, with their emphasis on effective communication and thinking and a fundamental understanding that one cannot possibly memorize every conceivable test item, is good educational practice for all students, not merely those with the designation of “advanced.”

Thus the first essential transformation for successful schools is to challenge both extremes in this acrimonious and unenlightening debate. Micromanagement of teachers with mindless test prep is not effective and demoralizes the best teachers. However, abandoning students to the whim of teachers who do not like standards but are also unwilling to engage in systematic thinking, reasoning, and writing is neither effective nor fair to students. The plain fact is that competent teachers require frequent thinking, reasoning, and writing from students. The most effective writing programs include collaborative scoring by teachers who use a consistent scoring guide across all disciplines. Administrators who facilitate this collaboration and require more writing rather than more test preparation are likely to be rewarded in their efforts with more motivated teachers and more successful students.

Critics of standards and testing will respond that this evidence is irrelevant if administrators do not believe it; fearful administrators will reject thinking, writing, and reasoning in favor of test drill, and thus the problem remains. Their point is well taken; it is not leadership demands per se that are wrong but leadership demands inconsistent with

the evidence that are unproductive. Schools that demand more writing will produce more students with improved skills in thinking, reasoning, and communication. These schools will also benefit from higher test scores.

Transformation 2. From Mystery Grades to Assessments that Are Accurate, Fair, and Educative

Too often, educational tests, grades, and report cards are treated by teachers and parents as autopsies when they should be viewed as physicals. That is, assessments are often perceived as only summative in nature when they should be formative—helping teachers “diagnose” student difficulties and gain insight into how to improve students’ achievement. Wiggins (1988) deems these incisive evaluations “educative assessments”: they provide feedback that promotes learning rather than letters that signify a mere evaluation. Educators should consider four critical practices to move toward more educative assessment.

First, educators must abandon the average, or arithmetic mean, as the predominant measurement of student achievement. Often, proficient students receive low grades because of failures early in the semester, whereas students in the same class who lack proficiency in the subject but offer consistent performance and good attitudes receive the same grade. The public would never accept such inaccuracy in something it values, such as sporting contests. In that domain, the public insists that the scores reflect the proficiency of the players, and rule changes are sometimes implemented to elevate the relative importance of skill over chance. Yet the public routinely tolerates inconsistent and inaccurate grading systems in those areas of less concern, such as the academic performance of students. Although it is true that students acquire and process information at different rates, it is not the pace of learning that standards-based systems should evaluate. Perhaps a state or district will promulgate the standard that one must “learn algebra quickly.” Until that time, however, accurate assessment will focus on the fact that a student is proficient in a subject and not on the pace at which proficiency is acquired.

The practical impact of abandoning the average goes beyond the method of calculating grades. The acknowledgement of variable paces of learning also requires acceptance of variable schedules. In a standards-based school, one student may require four semesters to become proficient in algebra whereas another student requires one semester. If the academic standards refer to achievement rather than pace, then it is the schedule that must be the variable, not the expectations of the teacher or the achievement of the student. This renders obsolete the common requirement that students have three years of math or two years of science or four years of English. In fact, some students need 16 semesters—two periods each day for four years—of English language arts to achieve the state literacy standards. Other students might need 12 semesters of mathematics to acquire proficiency in algebra and geometry. A growing number of secondary schools are making such adjustments in their schedules, making time the variable and maintaining high expectations for all students.

The second change in assessment practice educators must make is a rejection of is the notion that every review of student work results in an evaluation. The assumption

underlying this time-honored practice is that unless work is graded, the students will be unmotivated. This assumption is a splendid one from the point of view of teachers and administrators who, college degrees in hand, think that a D (or, given grade inflation, a B minus) is a terrible grade. Students disagree; they routinely choose a lower grade over the bleaker prospect of more work. In effective schools discussed previously, one of the most consistent practices of successful teachers is the provision of multiple opportunities to learn (Reeves 2000). This does not represent sympathy or low standards, but instead signals a fundamental change in the reaction to unacceptable student work. The consequence for a student who fails to meet a standard is not a low grade but rather the opportunity—indeed, the requirement—to resubmit his or her work. The positive results of this policy include increased respect for teacher feedback and improved learning by students. Such results are only possible, however, when the assessment is not a one-shot affair but rather a process in which students have the opportunity to take teacher feedback, reflect on it, and use it to improve performance. When a final exam is administered on Wednesday and grades are due on Friday, the unspoken message is that teacher feedback is worthless and student efforts to consider teacher feedback and use it to improve academic performance are fruitless.

Third, educators must relate numerical and letter grades with accuracy. Thanks to the predominant use of computer-based grade reports in many secondary schools, teachers are required to use points, typically on a scale of 0 to 100, for student assignments. Teachers then create a system of linking letter grades to these numerical scales, such as A = 93 to 100; B = 85 to 92; C = 77 to 84; D = 69 to 76; and F = less than 69. Although the specifics vary among secondary schools, there is a striking consistency: the interval between grades A through D is markedly different from the interval between grades D and F. In the preceding example, the interval between grades A through D is 7 points, yet the potential interval from grades D to F is 69 points. The result is one with which all students who have ever missed an assignment are familiar: the 0 grade has a disproportionate impact on the average grade. If the teacher had assigned values of 4, 3, 2, 1, and 0 for grades of A, B, C, D, and F, then the impact of a single failure would be mathematically equivalent to a one-grade difference. But when a failure receives a grade of 0 in a 100-point system, the impact of the failure is many times greater than the difference of one grade.

When this inconsistency is pointed out to teachers and administrators, they frequently offer excuses about the manner in which the computer is programmed and the inability of teachers to do much about it. The solution is clear and straightforward: If educators must use a numerical scale, then the lowest possible number on the scale should be the one grade value lower than a D. Using the preceding scale as an example, if the lowest D is a 69 and the interval between grades is 7 points, then the value of a failure should be 69 minus 7, or 62. It certainly should not be 0. This has nothing to do with sympathy or justification for students who fail to meet the standards; it is simply a matter of mathematical accuracy in grading. Without such a commitment to accuracy, a grading system lacks credibility.

Fourth, educators must assume the responsibility to describe completely the grades they assign. This requires the use of a Standards Achievement Report (Reeves 1997) similar to that routinely used in elementary schools. At the secondary level, teachers often raise the objection that an elaboration on the letter grade takes too much time, given a student population of 150 or more for each teacher. In fact, the grade books of secondary educators are full of details, representing thousands of individual marks and many hours of work by the teachers. Unfortunately, these details remain unknown to all but a few students and parents. Those few who discover the labors of the teachers are, unfortunately, the students who challenge a grade and discover that the teacher has documented, with care and precision, the justification for the evaluation. Thus the explanations for letter grades already exist in exquisite detail. The simple change from a one-line-per-student grade book to a one-page-per-student Standards Achievement Report would require no additional effort by the teacher but would provide vastly improved information for students and parents. Figure 3 provides an example of such a report. This example demonstrates that not every assignment encompasses every standard, but every assignment addresses several of the most important standards for this class. Moreover, the teacher has clearly identified that there are both academic and behavioral standards for students. This particular report shown is the profile of a student who is proficient or exemplary in every academic standard, but who displays a rather consistent inability to turn work in on time. The feedback provided by this report is far more revealing than the letter grade.

Transformation 3. From Coverage to Power Standards

The notion of coverage of standards can be a troubling one for teachers. In the casual conversation of the faculty lounge, a teacher may comment that, “Hey, if those kids don’t get the volume of a prism, it’s not my fault. I can document that I covered it on March 23rd at 10:40 a.m. It’s in the lesson plan!” The imperative of coverage of standards is largely the result of the displacement of rigor with girth by the various state academic standard-setting authorities. As Marzano, Kendall, and Cicchinelli (1998) note, the reach of state standards vastly exceeds their grasp. This leaves schools with two choices: frantic coverage or thoughtful focus. Both choices are fraught with risk, with the former risking learning and the latter risking omission of some standards. The evidence supports those who take the second risk.

The entire standards movement is put at risk by those who have replaced focus and meaning with quantity and coverage. One must not, however, confuse effectiveness with popularity. When one suggests that the memorization of the dates of the Crimean War may be less meaningful than a deep understanding of national economics and international conflict, one is besieged by those who equate intimacy with the Crimean War with cultural literacy. When one suggests that students might well focus on measurement, calculation, and problem solving rather than the lower level distinctions between the trapezoid and the rhombus, one is reviled by the partisans of the rhombus. Even the stegosaurus has its lobbyists who are armed for battle when one suggests that students are better served by immersion in the scientific method than by one more diorama in which prehistoric creatures are displayed with craftsmanship that outstrips paleontology.

The solution to the triumph of superficiality is the concept of power standards, that small subset of standards that must pass three crucial thresholds. First, do the knowledge and skills implied by this standard endure? That is, will they last beyond a few test questions in a single year? The scientific method endures, whereas the stegosaurus, alas, does not. Second, does the standard have leverage? That is, is it applicable across a wide spectrum of other standards? For example, the facility of a student to create and draw inferences from tables, charts, and graphs has extraordinary leverage, as these skills are found not only in the mathematics standards but also in science, social studies, and language arts. Similarly, the ability to write a descriptive, expository, and persuasive essay crosses the boundaries of many disciplines. Third, is the standard required for the next level of instruction? This threshold would require a focus in middle school on calculation, problem solving, measurement, and a few other areas. Most high school algebra teachers would be delighted if students came to them proficient in these few power standards, even if their young charges are rhombus impaired.

The best way to identify power standards is to ask faculty members this question: If you were giving advice to a teacher in the next lower grade, what knowledge and skills would that teacher need to impart to students in order that they enter your class next year with confidence and success? This question has been asked of hundreds of teachers at all levels of instruction, and few have replied that the teacher in the next lower grade must cover the standards. Rather, educators universally provide responses with two characteristics: brevity and balance. The lists are invariably short, rarely comprising more than 10 items. Moreover, the lists are balanced, including not only discipline-specific information but also a healthy dose of literacy, organization, time management, and appropriate behavior. It is the wisdom of these teachers, not the mindless pursuit of coverage, that leads to success in future classes and on high-stakes tests. Another question may help to disarm those who remain advocates of coverage because the “rhombus might be on the test”: When you took your last high-stakes test—any in the alphabet soup of tests in your past including the GRE, ACT, SAT, and PSAT—did you prepare for the exam by attempting to memorize the item base of the Educational Testing Service? Serious reflection suggests that a focus on a few skills was the key to success at the time when the coverage advocates were students themselves. Power standards, not frantic coverage, form the answer to both successful educational practice and improved test performance.

Transformation 4. From Seniority to Equity in Teacher Assignment

Without question, the most controversial change suggested here is a challenge to the tradition in which the seniority system is used to allocate the greatest teacher experience, certification, and quality to schools with the wealthiest students. The data are unambiguous on this point; Haycock (1998) referred to a synthesis of the evidence when she noted that the impact of subject-matter certification is far more significant than demographic variables on student performance, yet around the nation the allocation of teachers with subject-matter certification is markedly inequitable. Poor and minority students are far less likely to have certified teachers than their economically advantaged peers. Within districts or even individual school buildings, the inequities are pervasive:

teachers with greater experience, degrees, and certifications are assigned to students in advanced classes, whereas the poorest and least able students frequently find themselves in classes led by long-term substitute teachers or those new to the profession. To be sure, some new or uncertified teachers are extraordinarily competent, whereas some veteran, certified teachers are extraordinarily incompetent. Further, these generalizations should not diminish the heroic efforts of the many teachers with multiple certifications, advanced degrees, and years of experience who labor in the most disadvantaged and challenging schools. On the whole, however, the certification, degrees, and experience of teachers are highly associated with student success, and teachers with those characteristics in abundance are overwhelmingly assigned to students who are economically advantaged (Rothstein 2000).

There are some notable exceptions. Some schools in California, Texas, and Illinois take pride in their policy of equitable assignment of teachers. Within those schools, every teacher has the same opportunity to work with the most and least able of the students. These schools, however, represent the exceptions. The typical reaction to the suggestion that school systems replace seniority with equity is the question, “Do you mean that a 30-year veteran with a master’s degree should have to teach basic math while a new teacher should teach AP Calculus?” The answer is, as a matter of simple justice, yes. The argument is not that all veteran teachers be reassigned to poor schools, but only that educators provide equal educational opportunity for all children. This should not be an excessively demanding standard for a nation and a profession claiming a commitment to equity.

The predominant litigation in school finance over the past few decades has focused on the equitable allocation of financial resources. Although this is important, financial resources are not sufficient. The most important resource any educational system has to allocate is the expertise of its teaching professionals. The assignment of teachers to different schools and courses, decidedly more sensitive and less convenient than the reallocation of financial resources, has the greatest opportunity for a profound impact on student equity.

Transformation 5. From High Stakes to Meaningful Evidence

By 2003, 26 states will use some version of a high-stakes test as a criterion for a high school diploma (Coleman 2000). Despite the apparent popularity of high-stakes tests with educational policymakers, there remain two significant problems: students may be proficient in state standards and fail these tests, and students may succeed on the tests without being proficient. Indeed, the two errors feed on one another, as fear of the first error (and the attendant threat of litigation) leads to a greater likelihood of the second error.

The quest for the holy grail of testing—the perfectly (or even reasonably) reliable and accurate large-scale test of student competence that in a few hours can sum up 13 years of public schooling—is destined to fail. This pessimism is not based entirely on the sorry record of large-scale standardized tests (Lemann 1999; Popham 2000), though the history is hardly distinguished. Although validity, reliability, alignment with the curriculum, and

the opportunity for students to succeed are the criteria that would make a high-stakes test defensible, these criteria cannot be present simultaneously.

For a test to be valid, it must test what it is intended to test. That means, for example, that tests of writing test the ability of the student to write, rather than the student's knowledge of a specific content area suggested by a writing prompt. The larger the body of writing examined, the more likely that one obtains a valid estimate of the ability of that student to express a coherent thought in writing. The smaller the body of writing examined (in the extreme case, when the ability of a student to write is estimated on the basis of on a single writing sample), the more likely that the inference drawn about that student is representative of the content rather than the written expression. In brief, smaller samples of writing are more likely to yield invalid results. In the context of mathematics, science, or social studies, the requirement for validity conveys the notion that assessments are testing what the student has been led to believe is a proficient understanding of these disciplines. Many state standards, however, require deep understanding rather than mere factual recall. Deep understanding, whether of the scientific method or of the analysis of historical trends or of the challenges of Internet-based research, is not revealed in isolated tests. The very words of the standards require evidence that extends far beyond the score on a single test.

For a test to be reliable, it must evaluate student performance consistently. In the case of multiple-choice tests, this implies a consistency among items. If a student answers 9 out of 10 arithmetic problems correctly and misses the tenth problem, then either the last problem was quite difficult or, if other similarly able students got the tenth problem correct, then that test item was unreliable. It is as if the same student mounted a scale 10 times, and for the first 9 readings the student's weight was pronounced healthy and on the tenth reading the student's weight was labeled unhealthy. Because the same characteristic was being measured and the student had presumably not changed much between the ninth and tenth reading of the scale, it is the scale that is the problem, not the student.

The essence of the problem is that the requirement for validity implies multiple tests with some variations in administration and with independent judgments rendered regarding the ability of the student. The requirement for reliability requires a single instrument, consistently administered. Think about it. Most standardized tests fulfill only the latter requirement: single instruments, consistently administered. That is why college admissions tests such as the SAT and ACT are considered reliable. Few observers, however, continue to cling to the fiction that those tests are valid—that is, that they accurately predict how students will perform in college (Rooney and Schaeffer 1998).

A solution to the validity-reliability conundrum relies on the abandonment of the illusion that a single testing instrument such as the individual high-stakes test is the appropriate or accurate (or reliable or valid) measurement of student achievement. Student mastery is best evaluated when it is based on a body of evidence collected throughout an academic career, rather than on a single score. The consequence of the failure of a student to provide evidence of success should not be a lifetime of discredit and an eternity of litigation with the school system. Rather, the consequence of insufficient evidence of

student proficiency should be another opportunity to provide that evidence. Indeed, it is the feedback on the first attempt that informs the second attempt. It is the transparency, not the secrecy, of an assessment that makes it useful to students and believable by the public.

The use of a body of evidence rather than a single assessment score to ascertain student mastery is certainly more cumbersome and labor-intensive. Despite the logistical challenges, many school systems are succeeding. High schools from the Rocky Mountain School of Expeditionary Learning in Denver, Colo., to schools in Middletown, R.I., and North Clackamas, Ore., to more than a dozen schools in Wisconsin all require evidence that is based on multiple assessments with multiple opportunities for student success as part of their high school graduation criteria. This is the only meaningful way to evaluate students, particularly when the consequence of inadequate assessment is the denial of a high school diploma.

The use of a body of evidence to analyze student achievement does not negate all uses of standardized tests. Although it is politically popular to excoriate these tests, it is more useful for school administrators to understand the limits and strengths of such assessments. The critical issue is not the tests themselves, but how the information from those tests is used. On the one hand, a sound assessment used unwisely to make inappropriate decisions leads to the classic case of a valid and reliable test used to make invalid and unreliable policy. On the other hand, tests can provide vital information for school leaders and policymakers, and the information yielded from those tests should not be casually dismissed merely because the results are unpopular or embarrassing.

The Endurance of Standards and Civil Discourse

Will standards become another passing fad in American education, or will the standards movement endure over time? If standards are to endure, then educators must get beyond test prep and embrace thinking, reasoning, and communication. They must make grades accurate and meaningful. They must replace frantic coverage with power standards. They must commit to the principle of equity in teacher assignments. Finally, they must move from single high-stakes tests to a body of evidence as the measurement of the degree to which students meet standards.

The transition of academic standards from passing fad to permanent educational fixture will be achieved with enormous difficulty. The controversy surrounding standards will continue for some time, but surely both defenders and critics of standards can set an example for students by elevating the discourse and conducting the debate with greater civility and respect.

References

Coleman, A. L. 2000. None of the above. *Education Week* 19 (43): 42, 45.

Groves, M. 2000. California schools show big gains in test scores. *Los Angeles Times*, 5 October.

Kohn, A. 1999. *The schools our children deserve: Moving beyond traditional classrooms and “tougher standards.”* Boston: Houghton Mifflin.

Haycock, K. 1998. Good teaching matters: How well-qualified teachers can close the gap. *Thinking K-16* 3 (Summer): 1-16.

Lemann, N. 1999. *The big test: The secret history of the American meritocracy.* New York: Farrar, Straus & Giroux.

Marzano, R. J., J. S. Kendall, and L. F. Cicchinelli. 1998. What Americans believe students should know: A survey of U.S. adults. Aurora, Colo.: Mid-continent Regional Education Laboratory.

McNeil, L. M. 2000. Creating new inequalities: Contradictions of reform. *Phi Delta Kappan* 81 (10):728-34.

Ohanian, S. 1999. *One size fits few: The folly of educational standards.* New York: Heinemann.

Popham, W. J. 2000. *Testing! Testing!:* What every parent should know about school tests. Boston: Allyn and Bacon.

Reeves, D. B. 1997. *Making standards work: How to implement standards-based assessments in the classroom, school, and district.* Denver, Colo.: Advanced Learning Press.

———. 2000. *Accountability in action: A blueprint for learning organizations.* Denver, Colo.: Advanced Learning Press.

Rooney, C., and B. Schaeffer. 1998. Test scores do not equal merit: Enhancing equity and excellence in college admissions by de-emphasizing SAT and SAT results. Cambridge, Mass.: FairTest.

Rothstein, R. 2000. Getting good teachers for poor schools. *The New York Times*, 20 September, national edition, sec. A, p. 21.

St. Louis Post-Dispatch. 2000. Editorial, 14 September.

Wiggins, G. 1988. *Educative assessment.* San Francisco: Jossey-Bass.