



Cutler

IN 1999, AFTER TAKING THE GRADUATE MANAGEMENT Admission Test (GMAT), the standardized exam required of applicants to business schools, Mark Breimhorst sued the test's maker, the powerful Educational Testing Service (ETS). Breimhorst was born without hands and thus had been given more time to complete the admissions exam. His lawsuit contested ETS's practice of informing schools when students take one of its tests under specialized conditions, effectively placing an asterisk or, in testing parlance, a "flag" next to their scores. For unexplained reasons, instead of weathering a trial, ETS settled the case and agreed to stop flagging GMAT scores. This had far-reaching implications, since ETS runs several other major testing programs, including the oracle of college admissions, the SAT. As part of the settlement, Disability Rights Advocates, the civil-rights organization that had represented Breimhorst, and the College Board, the consortium of colleges that owns the SAT, agreed to jointly appoint a panel of "experts" to study whether scores on the SAT should continue to be flagged when students take the test with extended time.

Disabling the SAT

How the College Board is undermining its premier test by MIRIAM KURTZIG FREEDMAN

Disability Rights Advocates argued that the flagging of SAT scores violated the rights of students with disabilities by identifying them as disabled, stigmatizing them, and leading many to avoid using accommodations to which they are entitled. Accommodations like extended time, they believe, are necessary to equalize the testing experience for disabled and nondisabled students and thus make the scores of disabled students more valid. In 2002, to avert further litigation, the College Board acted on the expert panel's recommendation to end the practice of flagging scores. (Two weeks later, the other major college-admissions testing program, the ACT, adopted the same policy.) As of October 1, 2003, the Board will no longer note "Nonstandard Administration" on the scores of any students who take the SAT with extended time.

This decision will have an immediate impact on students with disabilities who take the SAT with accommodations, presently 2 percent of the two million students who take the SAT each year. Extended time, the only accommodation flagged by the SAT, is by far the most widely used

ILLUSTRATION BY DAVE CUTLER

accommodation, received by nearly all students who obtain accommodations. It is the only accommodation that nearly every test-taker would take advantage of if possible (unlike, say, the exams with larger print that are given to visually impaired students). Since 1976, the population of students designated as having a “specific learning disability” has grown 300 percent; learning-disabled students now compose 50 percent of the special-education population. In turn, since 1987, the number of students taking the SAT with accommodations has grown by more than 300 percent, compared with an 18 percent increase in the test-taking population as a whole (according to an analysis of annual reports from ETS. See sidebar by Samuel Abrams on page 40). Once the alleged “stigma” of flagging is removed, this trend can be expected to continue—and perhaps increase significantly.

The College Board’s unfortunate and puzzling action—an action that no court was likely to order if the case had gone to trial—deserves far more scrutiny than it has received so far. While private and without legal standing, this settlement involves giants in the testing industry and may have a chilling effect on validity and technical standards in the nation’s K–12 testing programs.

Standards for Testing

To understand how the Board’s decision undermines the

Since 1987, the number of students taking the SAT with accommodations has grown by more than 300 percent, compared with an 18 percent increase in the test-taking population as a whole.

SAT, it is important to recognize what a “flag” used to do. The SAT is valuable for two main reasons: 1) It provides colleges with a common standard against which to evaluate students who attend high schools with varying grading policies and levels of rigor, and 2) it partially predicts students’ grades during their freshman year of college, a measure of how prepared they are for higher education. The College Board website tells aspiring matriculants, “Your scores show colleges how ready you are to handle the work at their institutions and how your verbal and math skills compare with those of other applicants.” The flagging of SAT scores protected the test’s usefulness as a common standard of measurement by informing readers, such as college-admissions officers, when the test had been taken under unusual conditions, such as receiving time and a half to finish the standard three-hour exam. Without flagging, an admissions officer has no way of knowing whether the SAT scores of two candidates can be compared with one another, since there is no way of knowing whether the two candidates took the test under the same conditions. As of October 2003, admissions officers’ ability to assess the

meaning of test results and to make reasonable decisions for all students will be compromised. According to press accounts, 79 percent of college-admissions officers opposed the College Board’s decision.

Guidelines for administering norm-referenced tests such as the SAT are laid out in professional technical standards that testing companies follow. The leading authority in this regard

Alternative Solutions

The College Board had three better options than the one it chose. All would protect both the SAT’s validity and students’ rights to confidentiality regarding their disabilities.

by MIRIAM KURTZIG FREEDMAN

UNTIMED SATS FOR ALL. If, as the College Board asserts, time doesn’t affect validity, administer the SAT untimed for everyone! If the timing doesn’t really matter, don’t waste any more ink or enrich any more lawyers on this issue. Why should the Board continue to torture students, includ-

ing some with disabilities, by putting them up against the clock for no valid reason? As Brent Bridgeman, ETS’s principal research scientist, and his colleagues wrote in 2003, “A less speeded [mathematics] test is desirable now that scores of disabled students who are granted extra time will no

longer be flagged.” Surely, there are many students (with or without diagnosed disabilities) who don’t test well for whatever reason. Many may have anxiety about the ticking clock, others may benefit from a less stressful testing experience, and numerous students (especially those of low socioeconomic status) may have undiagnosed disabilities. Permitting all students to take the SAT without time constraints preserves the required standardization of test-taking conditions. Yet, while this option may have a certain appeal, the SAT

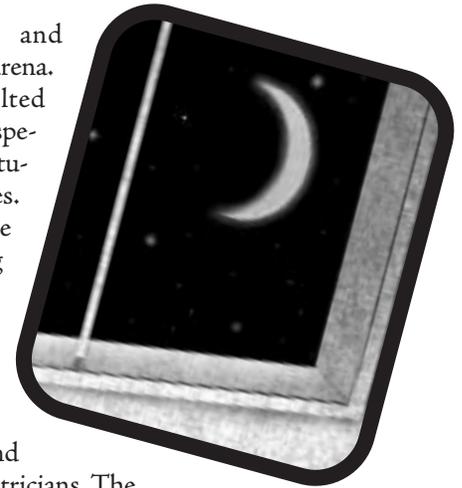
is the *Standards for Educational and Psychological Testing* developed jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. On the issue of flagging, the most relevant standard reads:

When there is credible evidence of score comparability across regular and modified administrations, no flag should be attached to the score. When such evidence is lacking, specific information about the nature of the modification should be provided, if permitted by law, to assist test users properly to interpret and act on test scores.

In other words, the standards require test administrators to note when a test has been taken under modified conditions unless there is “credible evidence” that the scores of students who took the test under standard and modified conditions are comparable—that is, that the scores carry the same meaning and weight. Historically, to comply with this requirement, the College Board and other testing companies have flagged results that were obtained under modified conditions such as extended time. This practice has long been considered legal under both case law and more than 25 years of guidance and rulings from the federal Office for Civil Rights. While the laws require reasonable accommodations for disabled individuals, they do not require fundamental alterations or the lowering of standards.

From the start, the makeup of the seven-person panel established by the College Board and Disability Rights Associates to review the Board’s policy on flagging should have sent a warning signal to anyone who is concerned about valid testing standards. Besides the chairperson, who was not expected to vote except to break a tie, the panel included two testing experts (psychometricians), one college admissions officer, and three persons with experience and training in the

special-education and learning-disabilities arena. The panel was tilted toward those with a special commitment to students with disabilities. The ultimate 4–2 vote in favor of ending flagging revealed the importance of its composition, as the two “nay” votes were from the admissions officer and one of the psychometricians. The differences between the two sides were so great that the minority wrote the scholarly equivalent of a judicial dissent.



The Panel’s Findings

The chief issue confronting the panel was whether the scores of students taking the test under standard conditions and with accommodations are “comparable”—in other words, whether they have the same weight and meaning and predict freshman GPAs with the same degree of accuracy. The SAT’s credibility as an admissions tool derives from its ability to predict, on average, how students will perform in their first year of college. The question is whether taking the test with extra time simply enables learning-disabled students to demonstrate their true level of knowledge and skills, which would presumably be reflected in their college GPAs, or whether it fundamentally alters the test. If extended time instead lends an artificial boost to their scores, one would expect their SAT scores to “overpredict” their GPAs, meaning that they would actually perform worse in college

would be a wholly different test, and that difference must be carefully considered.

LET THE STUDENTS DECIDE. If time does affect validity and standardized norms, as we have been led to believe since the precursor of the SAT began in 1926, then the Board can avoid the allegation of discrimination by allowing all students to choose whether they want extended time. Ask each test-taker: “How do you wish to take the SAT? You may either a) take the test within time con-

straints (and not have results flagged), or b) take it with extended time (and have the result flagged).” Presumably, poor test-takers and those who work slowly for whatever reason — as well as some students with diagnosed disabilities — will find these options appealing. This solution would undercut the theory of the threatened lawsuit, since a flag would no longer identify students as disabled.

DEFEND THE SAT. The College Board could void the settlement. If actually sued,

it could defend the SAT in court. Based on a long line of court decisions and guidance handed down by the federal Office for Civil Rights, which administers both nondiscrimination statutes — the Americans with Disabilities Act and Section 504 — a court would most likely defer to educational experts, uphold standards supported by evidence of the SAT’s validity, reliability, and technical underpinnings, and find flagging not to be unlawful discrimination.

The Demand for Special Accommodations

by SAMUEL J. ABRAMS

Presently, 2 percent of all students taking the SAT receive special accommodations because a professional has found them to be disabled in some way. Up from just 0.5 percent in 1987 (see Figure 1), this reflects an increase of more than 300 percent in the number of students taking the SAT with accommodations, compared with overall growth in the test-taking population of only 18 percent.

Ordinarily, when the number of disadvantaged students taking the SAT increases, average scores decline, as those with more extreme disadvantages are drawn into the test-taking pool. Indeed, this is one explanation people gave for the well-known drop in SAT test scores during the 1960s and 1970s.

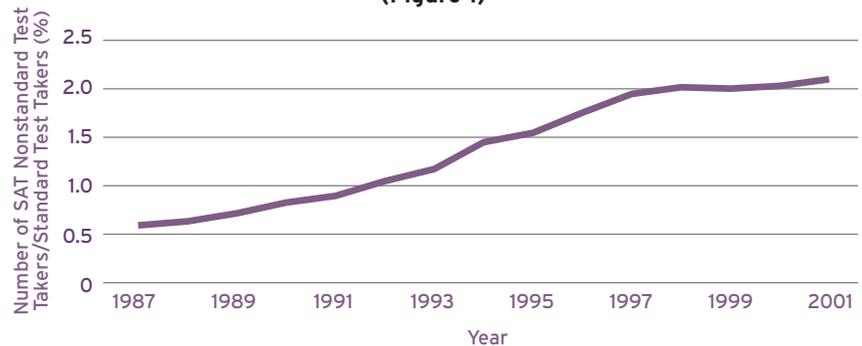
But exactly the opposite has occurred since 1987 in the case of those granted special accommodations on the SAT. As the numbers increased, so did test scores. Verbal scores on nonstandard administrations of the SAT are up by about 0.2 standard deviation since 1987, a small increase (see Figure 2). But math scores are up by 0.4 standard deviations, a large gain (see Figure 3). For those without special accommodations, one finds no gains in verbal scores and just a 0.1 standard deviation gain in math scores over this period of time.

What does this mean? One cannot be sure, but it appears as if those being granted special accommodation today are, on average, less in need of it than those in 1987. If this has been occurring with flagged scores, what will happen when the scores become unflagged? Will the growth in test-taking with accommodations take another leap forward?

—Samuel J. Abrams is a research fellow in the Center for Basic Research in the Social Sciences at Harvard University.

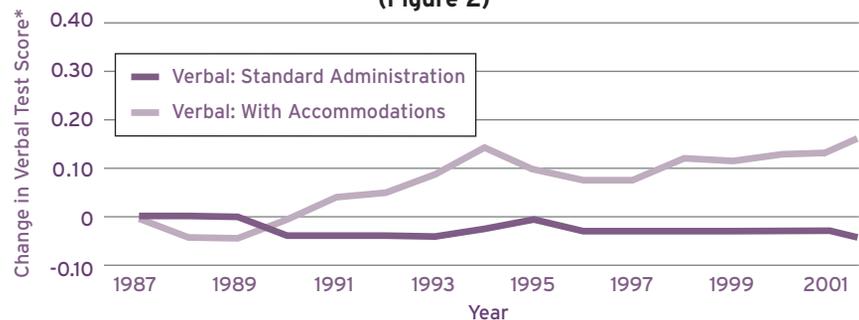
Growth in Percentage of Students Given Special Accommodations When Taking the SAT, 1987-2001

(Figure 1)



Gain in SAT Verbal Scores Among Those Granted Testing Accommodations, 1987-2002

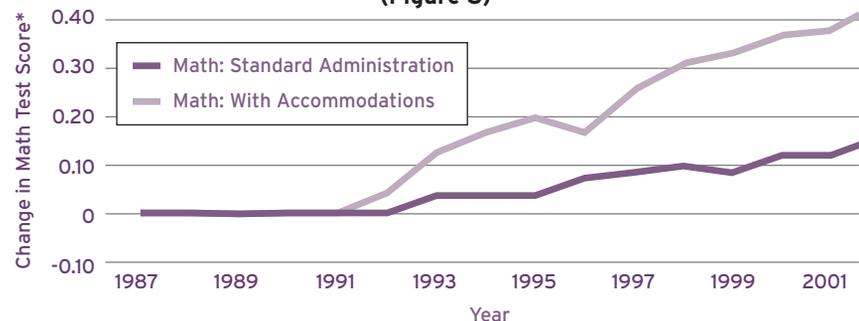
(Figure 2)



* In standard deviations. Standard deviation calculated for all test-takers.

Gain in SAT Math Scores Among Those Granted Testing Accommodations, 1987-2002

(Figure 3)



* In standard deviations. Standard deviation calculated for all test-takers.

than their SAT scores predicted.

In recommending an end to flagging, the panel majority concluded that “there are situations when it is necessary to treat people differently in order to treat them equally.” This conclusion emanated not from any independent assessment of the evidence but largely from the College Board and ETS’s own in-house research. The panel’s majority interpreted this research as showing that disabled students benefit more from taking the SAT with extended time than nondisabled students do.

In a study released by the College Board in 2003 (the panel’s majority report cited a 2001 draft version of this study), Brent Bridgeman, ETS’s principal researcher, and his colleagues analyzed the degree to which the SAT is “speeded”—or how much the SAT’s standard three-hour time limit affects students’ scores. To conduct the study they deleted some questions from the SAT’s experimental section, which appears in every test-taker’s booklet but does not count toward their scores (some students receive experimental math sections, others verbal, and students are not told which section of the test is experimental). The effect was to give a randomly selected group of nondisabled students extra time, about the equivalent of time and a half. The researchers then extrapolated from the results on the experimental section in order to estimate the impact of extra time on students’ overall scores.

They concluded that students who were allowed additional time on the entire exam would have scored no more than 10 points higher in verbal and 20 points higher in math, though there was no advantage whatsoever among students whose overall score in math was lower than 400. Higher-scoring students improved the most. This study did not examine the effects of extended time for disabled students. In addition, students did not know in advance that they would have extra time on the test, as disabled students do when they are granted an accommodation beforehand. Students might have performed differently had they been able to adjust their preparation or test-taking strategies to account for having more time.

A separate study, released by the College Board in 1998, investigated the effect of granting extra time to disabled students on a retest. Students who took the SAT once under standard conditions and then requested an accommodation scored 45 and 38 points higher in verbal and math, respectively, when they retook the test with extended time. Some of these gains are undoubtedly due to simply

The SAT taken by disabled students with extended time is a different test. That’s why their scores are flagged.

retaking the SAT, since nondisabled students who took the test twice under standard conditions improved their scores by 13 points in verbal and 12 in math. And since the learning-disabled students were not a randomly selected group but a group of students who requested accommodations for their retest, one has no idea whether the findings can be generalized to the disabled population as a whole. Yet the panel majority concluded, “This finding is consistent with the argument that students with disabilities need more time to demonstrate their knowledge, skills, and abilities than the nondisabled students, and suggests that the scores of these

students taken under the condition of extended time are more representative of their true performance than are the scores they would obtain from a standard administration.”

The bottom line is that the panel majority had no research directly comparing changes in the performance of nondisabled and disabled students when both are given extended time. Moreover, the existing research, given the limitations outlined above, hardly establishes that many nondisabled students would not benefit from having extra time. This is not to criticize the research, since the researchers themselves acknowledge these limitations. The problem is with the panel majority’s drawing of firm conclusions based on inconclusive evidence. Since the argument against flagging appears to be “more dependent on showing that extra time is of minimal benefit for the nondisabled population,” as Bridgeman and his colleagues write, the panel ought to have seen the findings from this research as equivocal, at best. The two psychometricians on the panel certainly recognized this. In a joint statement accompanying their individual reports to the panel, when they asked themselves the key question—whether scores from the SAT taken with and without extended time are comparable—one answered “no” while the other said “not sure.”

The debate should have ended right there. Neither testing expert could conclude that giving students extended time yielded scores that were comparable or predicted freshman GPAs with an equal degree of accuracy. Indeed, Robert Brennan of the University of Iowa (who directs the Iowa testing programs), the psychometrician who said “no” and voted with the minority, wrote, “Crucial evidence from prediction studies does not support a conclusion that scores on College Board standardized tests administered with extended time to disabled students are comparable to scores on the same tests administered to nondisabled students without



extended time.” The professional testing standard printed above says that when “credible evidence” that scores are comparable is “lacking,” scores should be flagged. The two psychometricians agreed that they lacked credible evidence of comparability, yet the College Board ended flagging anyway.

Why did one of the psychometricians, Stephen Sireci of the University of Massachusetts Amherst, vote to end flagging even though he was “not sure” that scores from tests taken with and without accommodations are comparable? Among his arguments, Sireci found that the College Board’s research on how well SATs taken with and without extended time predict college grades was “inconclusive” and criticized the professional testing standards themselves, including their failure to define what would constitute “credible evidence.” As the majority’s final report put it, there is no “commonly accepted criterion for determining” when “differences in predictive validity signify noncomparable scores.” The majority seemed to require those who believe flagging should be retained to show that scores from standard and nonstandard administrations of the test are *not* comparable, when in fact the standards require the opposite: that readers of score reports be notified when students are tested under nonstandard conditions if there is no “credible evidence” that scores from nonstandard and standard administrations of the test *are* comparable.

Undoubtedly, this issue requires more research. But without “credible evidence” that scores from tests taken with extended time are comparable to those taken under standard conditions, the majority had no basis within current technical standards for its recommendation that consumers of SAT information not be informed of the testing conditions under which a particular SAT score was achieved.

The majority used another puzzling argument to dismiss the evidence that scores on tests given with extended time overpredict a student’s freshman grades. They cited studies showing that the SAT also overpredicts the grades of male African-American students by about the same amount, yet their scores are not flagged. Furthermore, tests taken with extended time do not overpredict the GPAs of disabled *female* students, who account for about 40 percent of the students who receive extra time on the test. The majority concluded that the flagging policy effectively singled out one group, learning-disabled students, and



treated them differently even though the predictive power of their scores was no less than that of other groups.

This argument ignores the pivotal fact: *flagging is about the test, not the students.* The SAT taken by disabled students with extended time is a different test. That’s why their scores are flagged. Because the other groups take the SAT in the standard way, their scores should not be flagged. The majority missed the mark entirely. If the SAT does not reliably predict the freshman GPAs of a number of groups of students, then fix the SAT or develop a different test. But don’t muddle the issue by treating a test administered with extended time no differently from the standard test.

Facts of Life

The majority wrote that since most learning disabilities are rooted in dyslexia—an inability to read and process information quickly—and as the SAT is not supposed to be a test of speed (or so the College Board says), a dyslexic student’s need for more time should be allowed without consequence.

Their argument may have a certain logic in the abstract, but it ignores the realities of both the special-education system and, indeed, life itself. The panel majority assumed that the vast majority of students identified as learning disabled truly suffer from dyslexia. But there is no bright line separating learning-disabled students from students who learn more slowly than others or who were merely poorly taught. The College Board relies on K–12 schools, both public and private, to determine students’ eligibility for accommodations, and it is widely acknowledged that, from district to district or state to state, there is no consistency regarding the identification or assessment of students with learning disabilities. All kinds of students wind up being designated as learning disabled for a variety of reasons. Notably, in the 2003 College Board

study, Bridgeman and his associates acknowledged this reality, writing that if all students were given more time on the math portion of the SAT, “the pressure for students to get a sometimes questionable diagnosis in order to qualify for extra time would be substantially reduced.”

The rates at which students receive testing accommodations also vary dramatically by zip code, with well-to-do, empowered parents being able to pressure the system into giving their

While we all support the right of students with disabilities to opportunities in all aspects of education, compromising the SAT does not protect those rights.

children extra support. Is it fair for children of the wealthy to receive accommodations without consequences while poor children with undiagnosed learning disabilities languish under the rigors of a timed SAT? The panel majority encouraged the College Board to reach out proactively to disadvantaged students in order to inform them of their right to request accommodations, but the decision to end flagging certainly wasn't made contingent on its happening.

The Board's decision to end flagging is likely to exacerbate these problems. Now that there is no consequence for taking the SAT with extra time, so-called diagnosis shopping will undoubtedly become even more common among the well heeled, who can afford the private psychologists and pricey lawyers. And what's to stop them? School districts certainly don't have any incentive to limit the number of students who take the SAT with extended time, since higher scores look good to parents, taxpayers, and real estate agents. Who will be the gatekeepers?

Moreover, speed, whatever the College Board's assertions, is an important factor in the SAT. In fact, students report that the hardest thing about the SAT is the speed at which they need to work in order to answer questions accurately and still try to finish. ETS's own research shows that students perform better when given extra time. Time matters in the world as well. Processing speed is a valid indicator of certain skills. By removing the "Nonstandard Administration" flag, the College Board deprives college admissions officers of a valid criterion for considering whether a student is likely to succeed at the university. It is not unfair or illegal to test with time constraints—if that is within the purpose, norming, and standardization of the test. And, of course, if the College Board truly believes that time doesn't matter on the SAT, then why does it continue to time it for most students?

The members of the College Board's panel who voted to end flagging clearly didn't agree with this reasoning. In fact, in discounting the claim that extended time has been shown to corrupt the SAT's ability to predict freshman grades, the majority's report argues that we can't even measure the SAT's predictive validity unless students with disabilities are given accommodations in their college courses as well.

All in all, the College Board appears to have lost its perspective with this decision. In announcing the settlement, Board president Gaston Caperton said that "the rights of disabled persons should prevail over other considerations." A spokesman for the ACT echoed this sentiment, saying, "We are all in the business to make the tests as fair to everybody as we can." These statements misapply the concept of "fairness." Of course the SAT should be fair. It should be fair to all students. That means that all students are entitled to take the same test, with or without accommodations, flagged appropriately, to assure the requisite comparability of scores. Fairness is not about favoring the "rights" of some students over those of oth-

ers. The Board's role is not to engage in misplaced social maneuvering to lessen perceived (or real) stigma. Its role is to develop a test with a specific purpose, one that measures what it is intended to measure. While we all support the right of students with disabilities to opportunities in all aspects of education, compromising the SAT does not protect those rights.

The Board also ignored the fact, highlighted in the minority report, that users of the SAT, colleges and universities, bear the ultimate responsibility to interpret test results appropriately. The Board's responsibility is to provide them with all the relevant information they need to weigh and interpret the scores. By elevating the right of disabled persons over "other considerations," the Board ignored industry standards, compromised its flagship product, and misapplied federal laws, including the American with Disabilities Act and Section 504 of the Rehabilitation Act of 1973.

While these laws require nondiscriminatory policies, including providing "reasonable accommodations" for students with disabilities to demonstrate their knowledge on tests, they do not require the use of accommodations that fundamentally alter what the test measures. The SAT, a timed test, has been fundamentally altered by this settlement because the scores *are not* or *may not be* comparable. The Supreme Court's decision in *Southeastern Community College v. Davis*, decided almost 25 years ago, included the following: "Section 504 imposes no requirement upon an educational institution to lower or to effect substantial modifications of standards to accommodate a handicapped person." It is still good law.

One can only guess why the College Board blinked in the face of threatened litigation and did not take one of the reasonable alternatives enumerated in the sidebar on page 38, any of which would have promoted fairness while also protecting the SAT's integrity. Perhaps the recent attacks on the SAT, especially University of California president Richard Atkinson's 2001 proposal to eliminate the SAT as a requirement for application, have made the Board wary of controversy. In any case, this settlement's deviation from accepted technical testing standards further compromises faith in the validity of the SAT. Since the SAT is a giant in the testing industry, this settlement should be of great concern to those involved with the nation's academic standards and testing programs. With the decision to end flagging, most students will take the test in three hours, some in four and a half hours, others in five hours (for a shorter version of the test), without any reporting of these differences. Does that pass the test of common sense? Yet the lack of controversy surrounding this issue has been striking. A discussion is long overdue.

—Miriam Kurtzig Freedman is an attorney at Stoneman, Chandler & Miller LLP, a law firm in Boston, Massachusetts. She represents school districts on issues of testing, standards, and students with disabilities, and writes, speaks, and consults nationally on these issues.