
The War Against Testing

David W. Murray

This selection first appeared in *Commentary's* September 1998 edition. David W. Murray is director of the Statistical Assessment Service in Washington, D.C.

It is safe to say that Thomas Jefferson never took a standardized test, and would probably consider them hopelessly inadequate as measures of what an educated person should know. Yet Jefferson, in his way, was the inspiration behind our present vast apparatus for assessing academic aptitude and achievement. Looking toward America's future, he imagined an educational system that would seek young people from "every condition of life," students of "virtue and talents" who would someday form a "natural aristocracy" to replace the old-fashioned kind based on wealth and family background.

The U.S. of course has never fully achieved this ideal. But particularly in the period after World War II, as ever larger numbers of Americans entered colleges and universities, Jefferson's educational vision did begin to appear closer than ever to being realized. To an extent unimaginable a few generations earlier, access to American universities, and especially the elite ones, became based on considerations of merit. The chief instrument of this transformation was the standardized test—mass-administered, machine-scored, and utterly indifferent to every characteristic of a student save his ability to get the answers right.

And yet, for all its obvious benefits in helping to identify Jefferson's "natural aristocracy," and for all its widespread acceptance—this year, the Educational Testing Service (ETS), the organization that does the bulk of such evaluation, will administer its tests to some nine million students—the enterprise of testing has never been free from criticism. Today, in fact, its critics are more numerous and more vociferous than ever.

Indignant over the recent drop in minority enrollment at some state universities as a result of bans on affirmative action, the foes of standardized assessment argue with bitterness that America's vaunted meritocracy has never served all its citizens equally well. As they see it, moreover, the real issue is not the abilities of the test-takers, minority or

otherwise. Rather, it is the tests themselves, and the unreasonable emphasis placed on them by the gatekeepers of American higher education.

The oldest and most familiar accusation against standardized tests is that they are discriminatory. As the advocacy group FairTest puts it, a seemingly objective act, namely, “filling in little bubbles” with a No. 2 pencil, conceals a process that is “racially, culturally, and sexually biased.”

The prime evidence for this charge is the test results themselves. For many years now, the median score for blacks on the Scholastic Assessment Test (SAT) has fallen 200 points short of that for whites (on a scale of 400 to 1600, divided equally between math and verbal skills). Less dramatic, but no less upsetting to groups like the Center for Women’s Policy Studies, has been the persistent 35-point gender gap in scores on the math section of the SAT.

The SAT produces such disparate results, say critics, because its very substance favors certain kinds of students over others. Thus, fully comprehending a reading selection might depend on background knowledge naturally available to an upper-middle-class white student (by virtue, say, of foreign travel or exposure to the performing arts) but just as naturally unavailable to a lower-class black student from the ghetto. The education writer Peter Sacks calls this the “Volvo effect,” and has offered for proof an ETS study according to which, within certain income brackets, the difference between the test scores of white and black students disappears.

At the same time, women are said to be put at a disadvantage by the multiple-choice format itself. Singled out for blame are math questions that emphasize abstract reasoning and verbal exercises based on selecting antonyms, both of which supposedly favor masculine modes of thought. “[F]emales process and express knowledge differently, and more subtly,” explains FairTest’s Robert Schaeffer. “They look for nuances, shades of gray, different angles.”

In fact, so biased are the tests, according to their opponents, that they fail to perform even the limited function claimed for them: forecasting future grades. The SAT, says Peter Sacks, consistently “underpredicts” the college marks of both women and minorities, which hardly inspires confidence in its ability to measure the skills it purports to identify. As for the Graduate Record Exam (GRE), required by most academic graduate programs, a recent study of 5,000 students found that their scores told us almost nothing, beyond what we might already know from their grades, about how they would perform in graduate school.

Another line of attack against the tests grants their accuracy in measuring certain academic skills but challenges the notion that these are the skills most worth having. High test scores, opponents insist, reveal little more than a talent for—taking tests. According to a 1994 study by the National Association of School Psychologists, students who do well on the SAT tend to think by “rote” and to favor a “surface approach” to schoolwork. Low scorers, by contrast, are more likely to delve into material, valuing “learning for its own sake.”

It is likewise contended that no mere standardized test can capture the qualities that translate into real-world achievement. Thus, when it emerged last year that American children ranked dead last among the major industrial nations in the Third International Mathematics and Science Study, the Harvard education expert Howard Gardner declared himself unconcerned. The tests, after all, “don’t measure whether students can think,” just their exposure to “the lowest common denominator of facts and skills.” Besides, Gardner observed, at a time when America enjoys unrivaled prosperity, what could be more obvious than that “high scores on these tests . . . aren’t crucial to our economic success”?

In a similar vein, the social commentator Nicholas Lemann has called for a reassessment of what we mean by meritocracy. Our current view of it, he argued recently in the *New York Times*, is “badly warped.” If universities are to regain the “moral and public dimensions” that once connected them to the wider society, instead of being mere instruments for “distributing money and prestige,” they should begin to select not those students who excel on standardized tests but those with the skills necessary to lead “a good, decent life.”

This varied chorus of critics has already won some significant concessions from the current testing regime. For one thing, ETS, faced with both adverse publicity and threats of legal action by activists and the U.S. Department of Education, has tried to remedy differences in group performance. On the Preliminary Scholastic Assessment Test (PSAT), which is used for choosing National Merit Scholars, a new method of scoring was recently introduced in the hope that more women might garner the prestigious award. The old formula, which assigned equal weight to the math and verbal sections of the test, was replaced by an index in which the verbal score, usually the higher one for female test-takers, was doubled. The point, as a prominent testing official put it, was “to help girls catch up.”

More widely publicized was the massive “recentering” of SAT scores that went into effect with the 1996 results. Though the declared aim of ETS was a technical one—to create a better distribution of scores clustered around the test’s numerical midpoint—the practical effect was a windfall for students in almost every range. A test-taker who previously would have received an excellent score of 730 out of 800 on the verbal section, for example, is now granted a “perfect” 800, while the average scores for groups like blacks and Hispanics have received a considerable boost.

But since neither “recentering” nor any other such device has succeeded in eliminating disparities in scores, opponents of tests have had to look elsewhere. At universities themselves, affirmative action has long been the tool of choice for remedying the alleged biases of tests. With racial preferences now under siege, economic disadvantage is being talked about as a new compensating factor that may help shore up the numbers of minority students. The law school at the University of California at Berkeley, for instance, has introduced a selection system that will consider a “coefficient of social disadvantage” in ranking applicants.

Some schools go farther, hoping simply to do away with standardized tests altogether. There are, they insist, other, less problematic indicators of student merit. High school grades are a starting point, but no less important are essays, interviews, and work portfolios that offer a window into personal traits no standardized test can reveal.

Bates College in Maine, like several other small liberal arts schools, has already stopped requiring applicants to take the SAT. According to the college’s vice president, William Hiss, standardized scores are far less meaningful than “evidence of real intelligence, real drive, real creative abilities, real cultural sensitivities.” These qualities, moreover, are said to be especially prominent in the applications of minority students, whose numbers at the school have indeed shot upward since the change in policy.

Taken as a whole, the campaign currently being mounted against standardized testing constitutes a formidable challenge to what was once seen as the fairest means of identifying and ranking scholastic merit. Since that campaign shows every sign of intensifying in the years ahead, it may be relevant to point out that every major premise on which it rests is false.

In the first place, the SAT and GRE are hardly the meaningless academic snapshots described by their critics. Results from these tests have

been shown to correspond with those on a whole range of other measures and outcomes, including IQ tests, the National Assessment of Educational Progress, and the National Educational Longitudinal Study. Though each of these uses a different format and has a somewhat different aim, a high degree of correlation obtains among all of them.

This holds true for racial and ethnic groups as well. Far from being idiosyncratic, the scoring patterns of whites, blacks, Hispanics, and Asians on the SAT and GRE are replicated on other tests as well. It was in light of just such facts that the National Academy of Science concluded in the 1980's that the most commonly used standardized tests display no evidence whatsoever of cultural bias.

Nor do the tests fail to predict how minority students will ultimately perform in the classroom. If, indeed, the purported bias in the tests were real, such students would earn better grades in college than what is suggested by their SAT scores; but that is not the case. As Keith Widaman, a psychologist at the University of California, showed in a recent study, the SAT actually *overestimates* the first-year grades of blacks and Hispanics in the UC system.

Foes of testing are a bit closer to the mark when they claim that women end up doing better in college than their scores would indicate. But the "underprediction" is very slight—a tenth of a grade point on the four-point scale—and only applies to less demanding schools. For more selective institutions, the SAT predicts the grades of both sexes quite accurately.

As for the claim that test scores depend heavily on income, the facts again tell us otherwise. Though one can always point to exceptions, students who are not of the same race but whose families earn alike tend, on average, to perform very differently. A California study found, for example, that even among families with annual incomes over \$70,000, blacks still fell short in median SAT scores, trailing Hispanics by 79 points, whites by 148 points, and Asians by 193 points.

This suggests that universities turning to economic disadvantage as a surrogate for racial preferences will be disappointed with the results. And this has already proved to be the case. When the University of Texas medical school mounted such an effort, it found that most of its minority applicants did not qualify for admission, coming as they did from fairly comfortable circumstances but still failing to match the academic credentials of less-well-off whites and others. In fact, as a University of California task force concluded last year, so-called economic affirmative action, by opening the door to poor but relatively

high-scoring whites and Asians, might actually *hurt* the prospects of middle-class blacks and Hispanics.

What about relying less on tests and more on other measuring rods like high school grades? Unfortunately, as everyone knows, high schools across the country vary considerably, not only in their resources but in the demands they make of students. An A- from suburban Virginia's elite Thomas Jefferson High School of Science and Technology cannot be ranked with an A- from a school in rural Idaho or inner-city Newark, especially at a time of rampant grade inflation aimed at bolstering "self-esteem." It was precisely to address this problem that a single nationwide test was introduced in the first place.

Nor is it even clear that relying more exclusively on grades would bump up the enrollment numbers of blacks and Hispanics, as many seem to think. While it is true that more minority students would thereby become eligible for admission, so would other students whose grade-point averages (GPAs) outstripped their test scores. A state commission in California, considering the adoption of such a scheme, discovered that in order to pick students from this larger pool for the limited number of places in the state university system, the schools would have to raise their GPA cut-off point. As a result, the percentage of eligible Hispanics would have remained the same, and black eligibility actually would have dropped.

In Texas, vast disparities in preparation have already damped enthusiasm for a much-publicized "top-10-percent" plan under which the highest-ranking tenth of graduates from any Texas high school win automatic admission to the state campus of their choice, regardless of their test scores. Passed in the wake of the 1996 *Hopwood* case (1996), which scuttled the state university's affirmative-action program, the plan has forced many high schools to discourage their students from getting in over their heads when choosing a college. As one guidance counselor quoted in the *Chronicle of Higher Education* warned her top seniors, "You may be sitting in a classroom where the majority of students have demonstrated . . . higher-order thinking skills that are beyond what you have. You'll have to struggle."

Grades aside, what of the various less measurable signs of student potential? Should not a sterling character or artistic sensitivity count for something? What of special obstacles overcome?

Certainly, such things should count, and always have counted—more so today than ever, to judge by the sorts of questions most schools cur-

rently ask of applicants. But gaining a fuller picture of a particular student's promise is a difficult business, especially in an admissions process that very often involves sorting through thousands of individuals. Moreover, it can only go so far before it ceases to have anything to do with education. What a student is like outside the classroom is surely significant, but until we are prepared to say outright that the heart of the matter is something other than fitness for academic work, a crucial gauge of whether a student is going to be able to pass a biology final or write a political science research paper will remain that old, much-maligned SAT score.

There are, to be fair, social commentators who acknowledge this ineluctable fact, and who therefore urge us to direct all our remedial efforts toward improving the test scores of American blacks.* But for the true opponents of testing, such efforts—the work of generations—are clearly beside the point. Basically, what these critics are hoping to do is to achieve the ends of affirmative action by other, more politic means.

Hence the search for supposedly more “nuanced” measures of scholastic merit like “creativity” and “leadership,” tacitly understood as stand-ins for skin color. But there is no reason to think that minority students possess these qualities in greater abundance than do their peers. The attempt to substitute them for test scores will thus only perpetuate the corrupt logic of affirmative action by piling deception upon deception.

Whatever the euphemism used to describe it, only counting by race and gender can produce the result that will satisfy the most determined critics of standardized testing. If they have their way, and such testing wholly or partly disappears, we will have forfeited our best and most objective means of knowing how our schools are doing, as well as any clear set of standards by which students themselves can judge their own educational meritocracy will have come to an end. How this will benefit the poor and disadvantaged among us, or help them get ahead, is anybody's guess.

* See “America's Next Achievement Test” by Christopher Jencks and Meredith Phillips in the September-October 1998 *American Prospect*.