

## 13

## *The Uses and Misuses of Tests*

(1984)

THE debate about standardized testing has been one of the most rancorous educational issues of the past decade. Since the case against standardized testing has received a great deal of attention in the popular and scholarly media, the nature of the indictment is by now familiar. Articulate critics have charged that such tests measure only a narrow spectrum of abilities; that the tests by their very nature discourage creative and imaginative thinking; that the results of the tests have far too significant an effect on the life chances of young people; that the emphasis in a multiple-choice test is wrongly on "the right answer" and on simplicity instead of thoughtful judgments; that the tests favor the advantaged over the disadvantaged while claiming to be neutral; and that the tests are inherently biased against those who are unfamiliar with the language

### *The Uses and Misuses of Tests*

and concepts of the majority culture. In short, say the critics, the tests corrupt education, subjugate millions of students to their mechanistic requirements, and limit access to educational opportunity.

In examining the uses and misuses of testing, it is necessary to reflect on this upsurge of hostility to the testing process and to ask why it has occurred now.

My own view is that the tests have become increasingly controversial because they have become increasingly indispensable. Objections to standardized testing have accompanied the period in which the tests have become a fixture not only in educational decision making but in entry to the labor market. One of the sources of this increased criticism of the tests is egalitarianism, for the egalitarian complaint is that the tests discriminate among test takers and favor those with the best education and the most verbal ability. But the force that makes standardized testing an omnipresent feature of our society is also egalitarianism, because testing continues to be the most objective mechanism available to allocate benefits. In education, tests have grown more important to the extent that other measures have been discarded or discredited. Although it is easy to forget the past, we should recall that the tests helped to replace an era in which many institutions of higher education made their selections with due regard to the student's race, religion, class, and family connections. For many years, the objectivity of the tests was believed to be the best guarantee that selections would be made on the basis of ability, rather than status.

The tests have assumed an exceptional importance in college admissions, because other measures have been rendered useless. Personal recommendations today carry far less weight than they once did, because letter writers can no longer rely on the confidentiality of their statements. High school grades are a questionable standard, not only

because of the variability from one school to another but because of the prevalence of grade inflation. If almost everyone applying for admission to a select college presents an A record, then the grade point average becomes meaningless in the admissions process. In the current situation, the students who selected demanding courses and the schools that resisted grade inflation are handicapped when colleges attach importance to the grade point average. Personal interviews are helpful, but they are limited in value by the interviewer's prejudices and the student's ability to present himself. When all of these factors are considered, the tests—despite all of their flaws—are left as the fairest measure of a student's academic ability.

Thus the contemporary paradox. The more egalitarian our society becomes, the more important are standardized tests. Yet the more important the tests are, the more they are subject to egalitarian criticism for assuming too much power in determining future life chances. So long as there are educational institutions where there are more applicants than places, there must be an objective way to decide who gets in. This being so, the egalitarian critique of testing founders precisely because no other objective means has been discovered to take the place of ability testing.

Unless some more objective means is devised, testing will continue to be pervasive, perhaps even more than it is now. This is not necessarily a development to be welcomed, since it goes hand-in-hand with the growing bureaucratization of American education. However, it is important to note that the influence of standardized testing in college admissions is limited by demographic factors. Although critics frequently complain about the unconstrained power of the testers, a recent survey by the College Board showed that fewer than 10 percent of all institutions of higher education are highly selective. Most colleges and universities accept all prospective students who apply or

require only that they meet minimal standards. For the overwhelming majority of students, the tests are used for placement, not for exclusion from educational opportunity.

While they are certainly not perfect instruments of assessment or prediction, tests have appropriate uses for students, teachers, and educational institutions. Students who take the Preliminary Scholastic Aptitude Test (PSAT) or the SAT, for example, get a measure of their strengths and weaknesses relative to other students. Correctly read, not as a life sentence but as a one-shot assessment of verbal and mathematical abilities, the test score can direct the student toward appropriate study to improve areas of academic weakness. For teachers and schools, the tests are useful as rough indicators of how well students are learning the specific skills that are tested. The test scores can help the school in diagnosing educational problems and in prescribing appropriate remedies.

The chief virtue of the standardized test is that it may serve as an early warning system. If a student scores a 350 on the SAT, counselors and teachers should be alerted to find out why and to do something about it. If a school administrator sees a steady downward trend in the scores for a school or a district, it should also be considered a warning of possible problems in the teaching of academic skills.

The best example of how the tests function as an early warning system occurred during the past several years. In 1975, the College Board acknowledged that SAT scores had steadily and sharply declined since 1963-4. More than any other single factor, the phenomenon of falling test scores stimulated a national debate about education policies. As a result, the public and policymakers became concerned about the decline of academic standards and of literacy.

Initially, some in the educational field tried to explain away the score decline, either by questioning the validity

of the SAT or by pointing to the increased numbers of minority students in the pool of test takers. These attempts to allay public concern were soon rebutted, however, as additional research provided evidence that other standardized tests of verbal skills showed the same pattern of falling scores over the same period. In particular, Annegret Harnischfeger and David E. Wiley's article, "Achievement Test Score Decline: Do We Need to Worry?" documented a parallel drop in scores in a wide variety of tests, beginning in about the fifth grade.<sup>1</sup>

The second claim—that the score decline was caused by the inclusion of large numbers of poor and minority students in the test cohort—was effectively dismissed by the blue-ribbon panel appointed by the College Board and chaired by Willard Wirtz. The Wirtz panel found that the new students had contributed to the decline until about 1970; after that date, the composition of the test-taking population had stabilized, yet the SAT averages continued to fall and to fall even faster than before 1970.

The report of the Wirtz panel identified a number of in-school practices that probably contributed to the score decline. It observed that absenteeism, grade inflation, and social promotion had become widespread, while the assignment of homework had shrunk. One of its internal studies, prepared by Harvard reading expert Jeanne Chall, found that the verbal content of widely used high school textbooks had been reduced by as much as two grade levels. Although the panel was careful not to pin the blame for the score decline on any particular factor, it did note that there was "almost certainly some causal relationship between the shift in the high schools from courses in the traditional disciplines to newer electives." It further pointed out that its "firmest conclusion is that the critical factors in the relationship between curricular change and the SAT scores are (1) that less thoughtful and critical reading is

now being demanded and done and (2) that careful writing has apparently about gone out of style."<sup>2</sup>

The SAT score decline sounded a national warning bell that something might be terribly wrong in the schools. The reaction was not long in coming, and it was not always wisely considered. In almost every discipline, teachers reported the pressures of a "back-to-basics" movement that demanded greater attention to basic skills and disparaged innovative practices. Within five years after the news of the score decline broke, nearly forty state legislatures had adopted minimum competency tests in an effort to restore value to the high school diploma; such tests of minimal skills did little to raise overall educational quality. In response to these developments, more than two dozen commissions, task forces, and study panels were established to examine the problems of American education, with special focus on the high schools.

The spring of 1983 saw the release of reports from four of these groups, and several more followed in the fall of the same year. For the first time in a generation, the public became deeply concerned about the problems of American education. Hardly a day went by without an article in the news about merit pay, teacher education, curricular change, tightened standards for high school graduation or college admission, or some other educational subject that a year earlier would have not made it into the papers, let alone onto the agenda of the state legislature.

This time of ferment and reform was directly stimulated by the impact of the SAT score decline. No other single indicator had the power to alert the public to a national erosion of educational quality, nor the power to elicit research focusing on problems of educational quality. Though one would wish it were possible to generate interest in educational reform without developing so drastic a symptom, nonetheless the SAT score drop dramatically

raised the level of public attention to education.

These then are the uses of well-made standardized tests: as an assessment tool to help individual students identify their strengths and weaknesses, as a diagnostic and prescriptive technique to improve individualized learning programs, as a yardstick to help competitive colleges select their students, as a barometer to gauge the learning of academic skills, and as an early warning system to measure national trends in learning these skills.

But the tests are not an unmixed blessing. Many of the criticisms that have been made of them are on the mark. The tests can easily be misused and become an end in themselves, rather than a means. It is true that standardized tests measure only a narrow spectrum of abilities and that they cannot measure many valuable ways of thinking. The tests have validity only because the narrow spectrum of abilities that they do measure tends to be central to the learning process in college. The odds favor the future academic success of the student who scores 700 over the student who scores 400, yet the odds are not always right. We all know students who don't test well, who freeze up in the test situation, or who have gifts that the tests don't measure. Sensible admissions officers know this and are on the lookout for youngsters who have the imagination, creativity, or drive that doesn't register on the SAT.

The critics also have a point when they speak of the simplistic thinking that multiple-choice questions promote. While it is true that many questions asked on the SAT and on achievement tests have only one correct answer among those presented, the very emphasis on the right answer may itself be educationally counterproductive. As an historian, I am aware that the more I know, the less I am sure of. I am troubled when one of my children is asked to give the three reasons for the outbreak of some war or the four causes of some movement. When the

event or movement in question is still being debated by historians, as most everything is, then I am especially annoyed by the idea that test makers and teachers should treat them as settled issues. As a parent, I want my children to see history, politics, literature, and art in relation to one another, and not as compartmentalized events that can be defined in short answers or in multiple-choice questions. Furthermore, I want them to learn that most questions cannot be answered with a "yes" or a "no," that most judgments must be hedged by qualifications, and that questions about literature and history usually require complicated answers that must be explained, justified, and defined. In a better world, educational testers would value the slow, thoughtful response over the fast, reflexive answer.

Overreliance on standardized testing may be dangerous to the health of education. It is certainly dangerous to the integrity of the high school curriculum. The introduction of the SAT, which (in its verbal component) is curriculum free, left many high schools without a good argument for requiring students to take history, literature, science, or anything not specifically demanded by the college of their choice. The old College Boards were based on a very specific curriculum and on specific works of literature and periods of history; the elite secondary schools agreed on what was important to teach, and their students were well prepared for the examinations, which relied heavily on essay answers. It was a move toward democratic admissions when the SAT was adopted, because the SAT tested scholastic aptitude and made no assumptions about what curriculum the student had studied. As a result, public school students all over the country were able to compete fairly for places in the prestigious colleges. Unlike the authors of the College Entrance Examinations, the makers of the SAT do not care whether the student has ever read

Jane Austen or Charles Dickens or any particular work.

Now, it is not the fault of the Educational Testing Service that students may arrive at college with high test scores and appallingly little substantive knowledge of history or literature. But the curriculum-free SAT has presented no impediment to high schools that thoughtlessly decimated their own curricular requirements. Because the SAT is curriculum free, students who are good test takers are justified in thinking that they can do very well in the admissions process even if their preparation for college has been haphazard. Again, I want to stress that the SAT did not cause the curricular chaos that has come to be the bane of American high schools. But any admissions officer who relies on SAT scores without scrutinizing the content of the student's high school coursework is gravely misusing the test.

Standardized tests are misused when teachers, textbook publishers, curriculum planners, and administrators permit ordinary classroom practice to be dominated by the fill-in-the-blanks mentality, to the virtual exclusion of writing. Researchers have reported a sharp increase in the time spent in elementary schools and even in high schools on workbooks and busywork. The study of textbooks by Jeanne Chall for the Wirtz panel documented a marked increase in emphasis on "objective answers." Chall found that "generally, the assignments in the Reading, History and Literature textbooks [ask] only for underlining, circling and filling in of single words." When these busywork activities are substituted for student writing, they are anti-intellectual and subversive of good learning. Filling in the blanks is not equivalent educationally to the intellectual tasks involved in writing an essay, in which the student must think through what he wants to say, must organize his thoughts, must choose his words with care, and must present his ideas with precision.<sup>3</sup>

The harm in minimizing the practice of writing in the

classroom is not merely to the student; teachers are also injured. Workbook activity requires minimal skill and thought by teachers; they become technicians, checking for the correct answer, a rather low-grade form of labor. When they teach writing, their own intelligence and judgment and skill are brought into play. In order to teach writing, they must make decisions; they must provide guidance; they must set standards of accomplishment. In short, they must wear the mantle of professionalism. The shift in the classroom from teacher control to materials control no doubt contributes to what some observers have called the "deskilling" or the "technicization" of teaching, a process that converts teachers from professionals to civil servants.

In sum, there can be no doubt that the tests have their uses as well as their misuses. The standardized test should be seen as a measuring device, an assessment tool, never as an end in itself. The skills that it measures are important, but it does not measure every important skill. The information that it gives us about the state of a student's learning is never definitive, but tentative and subject to future change. Above all, we should not permit the standardized test to become the be-all and end-all of educational endeavor; we send our children to school not in order to do well on tests but in order to become educated people, knowledgeable about the past and the present, and prepared to continue learning in the future. Tests help us check up on how well children are learning, and this is their major value. Their uses are clear and limited. The mastery of tests should not be permitted to fill in the blank of what should be our educational philosophy.

Those who believe in the value of tests have a particular responsibility to guard against their misuse in the classroom, the press, admissions offices, and the workplace.