Court Ruling Rekindles Controversy Over SATs

Despite allegations, researchers have found no evidence of race or sex bias in college admissions tests

A NEW YORK FEDERAL JUDGE ruled in early February that the exclusive use of the Scholastic Aptitude Test (SAT) for awarding New York state scholarships discriminates against women. The decision has fueled a renewed debate over whether standardized tests are biased against women and minorities.

The decision by Judge John M. Walker came the month after a ruckus over a vote by the National Collegiate Athletics Association requiring students to get at least 700 on the SAT in order to win college athletic scholarships.

Walker issued a preliminary injunction based on his conclusion that the tests result in "disparate impact" on women and, therefore, sole reliance on the scores for awarding scholarships is discriminatory. He did not, however, contend that the test itself is biased.

Although males and females earn comparable grades in high school, females generally score about 60 points lower on the SAT's 1600 point scale. In 1987 this policy resulted in females winning 43% of New York's Regents scholarships and only 28% of the Empire State Scholarships.

In 1988, when high school grades were included in the eligibility determination, the percentages of female winners increased to 49% and 38%. But last fall, the state reverted to exclusive reliance on the SAT, arguing that the use of grade point averages does not take into account the disparity among schools, and encourages grade inflation.

The Educational Testing Service (ETS), which owns the SAT, has always recommended that its test be used in conjunction with high school grades in predicting ability to do college work. However, in an amicus curiae brief, it defended the state's right to use the test as it saw fit.

The ETS asserted that the validity of the SAT "is at least as strong for women as for men," and that the correlation of test scores with first-year college grades is, in fact, slightly higher with females.

In a brief prepared by the American Civil Liberties Union, the plaintiffs, a group including the National Organization for Women, claim the test is discriminatory—among other reasons—because it underpredicts women's college grades. Contrary to ETS's claims, they say female SAT scores do not correlate with first-year college grades as well as they do for males.

Judge Walker affirmed that the test is valid but that it also underpredicts female academic performance. He agreed with the plaintiffs that sex difference in test scores cannot be totally explained by "neutral" variables. Therefore, "disparate impact" exists. He affirmed a principle that has emerged from court cases on racial bias—that disparate impact is acceptable proof of discimination regardless of whether there is intent to disciminate.

Judge Walker affirmed that the test is valid but underpredicts female academic performance.

Although the judge's conclusion was that "the use of the SAT as the sole criterion" for selection is discriminatory, some test critics have interpreted the ruling as support for their belief that the test itself is biased.

The case has been hailed by the National Center for Fair and Open Testing (FairTest), a coalition of groups who oppose standardized tests, as "the first time the issue of bias has ever been decided by a court in regard to the SAT." Robert A. Schaeffer of FairTest predicted that the ruling would inspire other legal challenges to the test.

The question of "bias" is a tricky one. A test is biased if scores reflect factors extraneous to what it is designed to measure—family income, for example. But performance on tests may reflect the influence of such factors. Hence, if low income correlates with low scores, tests will be susceptible to charges of anti–low income bias. Scientifically speaking, bias against a particular group exists if the correlations between test scores and the "criterion"—in this case, freshman college grades—are different between two groups. There are a variety of ways of looking for bias. One is by ascertaining whether rankings of test items according to difficulty are different for different groups.

No sex bias has been detected in college admissions tests despite the fact that women's SAT (and ACT*) scores are lower than men's. In 1988 they averaged 45 points less on the SAT math score, and since 1972 they have fallen behind on the verbal score— by 13 points in 1988—as well.

Why do females have lower scores? The ETS has attributed the discrepancy to the fact that "the group of women who take the SAT have somewhat different educational and socioeconomic backgrounds than [the] men." (SAT-takers are 52% female.) Feminists blame social conditioning which discourages women from taking science and math.

Possible hormonal contributions to the discrepancy have been hypothesized by researchers at the Johns Hopkins University Study of Mathematically Precocious Youth, who have found consistent sex differences favoring boys in spatial and mathematical reasoning abilities from the age of 11 onward. Camilla Benbow, now at Iowa State University, believes this could help explain why girls perform well in math achievement tests, but do not do as well on independent reasoning tasks that are not directly based on learned material.

The female decline in verbal scores is something of a puzzle. Some test critics ascribe it to the existence of too many concepts based on sports and science. But according to a College Board spokesman, an analysis of 150 standardized tests has confirmed an across-the-board decline in verbal scores for females.

A subsidiary issue relates to the fact that female grades, in both high school and college, tend to be slightly higher than would be predicted by their SAT scores. Most people assume the reason is that females are more conscientious and obedient students. However, the ETS says "the available evidence indicates that the effect is due to differences in course selection" between the sexes "and in grading practices in the courses selected."

Psychologist Rogers Elliott of Dartmouth College has done research which he believes demonstrates that most of the apparent underprediction by the tests would disappear if college grades were corrected to account for the fact that females tend to take courses for which grading is less rigorous.

Elliott developed an index of differential grading rigor by ranking average grades given in various courses at Dartmouth. The hardest grading was economics, followed by

*The American College of Testing, which administers the other widely used college admissions test.

math and computer science, chemistry, biology, and physics. The easiest were drama, comparative literature, music, sociology, and Spanish. Anthropology stood in the middle. In two studies, one of the class of 1986 and the other of students enrolled in psychology courses, he found that adjustment of grades according to the index accounted for most of the underprediction for females. He also found that the adjustment improved the predictions for academic performance (made on the basis of SAT scores, high school grades, and achievement test scores) not only for freshman grades but for grades throughout the 4 years of college.

Although people have spent an awful lot of time trying to figure out why women are not as good as men in math and the physical sciences, the deficiency—which is only striking at the very high end of the distribution—is a minor one and arguably of questionable social significance. According to the College Board, there is no evidence to support allegations that women are being discouraged by their test scores from applying to selective colleges. Among the 77 that accept 50% or fewer of their applicants, 52% of the students are women.

The situation with regard to non-Asian minorities, especially blacks, is quite different. Although the black-white gap in scores has narrowed in the past decade, blacks still score a full standard deviation, or 200 points, below whites on the SAT.

The extreme sensitivity of the issue was once again shown by the extensive publicity accorded a recent decision by the National Collegiate Athletics Association. That group voted by a narrow margin to end exemptions to a rule that requires athletes to have a grade point average of 2 and either an SAT score of 700 or an ACT score of 15 (out of a possible 36). Of the estimated 600 athletes a year who would be affected, 90% are black, mostly from southern high schools. Forty-five percent of blacks who take the SAT (compared with 15% of whites) score 700 or below on the SAT.

The athletic association explained that the decision was an effort to push high schools do to a better job of preparing their students. But it was immediately denounced as racist by many observers, and Georgetown University basketball coach John Thompson made headlines by calling the tests "proven culturally biased" and walking off the court in protest at a subsequent game.

There are a great many people who believe with Coach Thompson that such bias is a proven fact. But, despite voluminous research on the question, the charge has not been substantiated by research.

The most extensive scholarly criticism of the SAT comes from James Crouse and Dale Trusheim of the University of Delaware, authors of *The Case Against the SAT* (University of Chicago Press, 1988). They do not question the validity of the test, but claim that its use with non-Asian minorities has a disproportionately negative impact while failing to improve selection decisions.

The number of students *taking both admissions tests each year is steadily growing.*

The authors say that when admissions decisions are made on a "colorblind" basis, combining SAT scores with high school grades "reduces the number of blacks admitted to colleges, sometimes more than half," while having no effect on whites. In colleges that have established black quotas they want to fill, says Crouse, using SAT scores in addition to grades does not result in any improvement in admissions decisions over the use of grades alone. In other words, with blacks, the additional winnowing done by adding SAT scores to the equation produces no better results than random rejection of the same additional number.

ETS and the College Board recently issued a statement saying they "profoundly disagree" with Crouse's conclusions. They claim that, in fact, the use of the SAT improves selection for blacks even more than it does for whites. For example, if the aim is to accept students likely to achieve a 2.5 freshman grade average, use of the SAT increases correct predictions by 11%, from 44% (based on high school record alone) to 55%—or 11,000 black applicants a year, according to the statement. The comparable increase for whites is closer to 3%.

Robert Linn of the University of Colorado (Boulder) agrees with ETS that the test predicts equally well for blacks and whites. He and others have found that the SAT, if anything, favors blacks, since those who get the same SAT scores as whites do not do as well in college. Linn says that both test scores and high school grades yield predictions for black freshman grades that are about one-quarter of a grade point higher than those actually obtained. High school grades tend to overpredict because blacks come from lower-ranking schools, but there is no consensus on why the tests also overpredict. The same pattern pertains for students from low socioeconomic backgrounds, but to a far lesser extent.

In any case, the potential impact of admissions test scores on black college acceptance has been greatly diminished by the widespread adoption of affirmative action programs, which are particularly strong in the most selective institutions. According to the ETS, being black increases the probability of being accepted by up to 50% at some highly selective colleges.

Recent controversies have helped further the impression that the SAT is falling out of favor. This is far from the truth. Spokesmen for both the SAT and the ACT say the number of students taking both college admissions tests each year is steadily growing.

Larry Litten of the College Entrance Examination Board, the main user of the SAT, says a "substantial majority" of admissions officers say the test makes the selection process more efficient, and are increasingly using the tests for student guidance and placement. A few small colleges have dropped the requirement for SAT scores but there is no sign of a trend away from college admissions tests. A recent survey of experts in academia revealed that 90% of the respondents favored continued use of the SAT in college admissions decisions.†

Many test critics believe that admissions

†Results of a survey of experts in education, psychology, sociology, cognitive science, and behavior genetics are reported in M. Snyderman and S. Rothman, *The IQ Controversy* (Transaction Books, New Brunswick, NJ, 1988).

tests would be fairer for minorities if they were achievement tests directly based on the content of high school instruction. But the evidence is not encouraging, given the high correlations between standardized achievement and ability tests. For example, says Elliott, the mean of the best three achievement tests correlates 0.83 with the SAT.

Crouse and Christopher Jencks of Northwestern University have pressed the case for moving to achievement-based admissions tests, primarily as a spur to raising the quality of high school education. Jencks points out that schools that attempt to prepare students for admissions tests would then at least be prepping them in calculus rather than in "test-taking skills."

Crouse suggests that the lack of uniformity in high school preparation could be remedied by inaugurating a program along the lines of the College Board's Advanced Placement program, which is based on uniform course descriptions for participating high schools.

The chief argument against moving to standardized achievement tests is one of the main reasons the SAT was created in 1926—to reduce discrimination against students with poor academic preparation by offering a means of uncovering ability that is independent of high school curriculum. The argument that achievement tests would lead to better high schools is met with skepticism by those who point out that many other factors, like budgets, exert more influence.

In absence of standardized high school curricula, Elliot observes, "an achievement test that everybody can take is going to be the lowest level." Other objections are that schools would "teach to the tests," emphasize knowledge amenable to the multiple-choice format, and encourage cramming.

Very few observers, FairTest excepted, believe the cause of better education for minorities would be served by eliminating standardized tests. More moderate critics believe changes in tests can exert only an indirect influence, as leverage to compel improvements in high school education.

By this reasoning, lowering standards for applicants to athletic scholarships would have as its major effect the perpetuation of the discrepancy between black and white academic performance. Nonetheless, the NCAA, after a week-long furor over its vote, decided to back down—in late January it announced that it would shelve the new rule pending further study of test scores and academic performance.

Hostility to standardized tests is a classic case of "blaming the messenger" of bad news, but it is unlikely to abate as long as solutions to the problem remain elusive.

■ CONSTANCE HOLDEN

AIDS Panel Urges New Focus

Ever since the AIDS epidemic began, most of the research attention has gone to biomedical scientists and their attempts to find the cause and cure of the disease. Last week, an expert panel of the National Research Council called for a broad array of programs to focus on the social behaviors that spread the disease.

The panel issued a 589-page report outlining initiatives to better monitor the spread of human immunodeficiency virus (HIV), to begin unprecedented efforts to collect data on intimate sexual behavior and drug abuse, and to support frank and aggressive educational programs designed to change risky behavior.

The AIDS virus promises to be around well into the next century, the report says, and improved information about its spread and about the effectiveness of programs to contain it are necessary to limit the crisis.

"AIDS is spread by social behavior. The only way to combat the spread of HIV is through changing behavior," Lincoln E. Moses, chairman of the Department of Statistics at Stanford University and the panel's chairman, told a news conference.

Efforts to monitor the epidemic should include anonymous surveys of all newborn infants and women treated in abortion clinics. These data are vital to establishing a baseline infection rate, the panel says. Although the Centers for Disease Control (CDC) estimates that 1 million Americans are now infected with HIV, "we really have no reliable incidence data," Moses said. CDC is already planning to test newborns, for example, in 30 metropolitan areas.

Data on Americans' sexual habits are equally lacking, the panel says in urging vastly expanded federal research efforts in this area. Much current knowledge of sexual behavior comes from the Kinsey studies of the 1930s and early 1940s, which were methodologically flawed. Little is known, for example, about how many men have sex with men, or about the prevalence of extramarital affairs, or the practice of anal sex among heterosexuals. The National Institutes of Health currently is planning just such a survey (*Science*, 20 January, p. 304).

The panel is especially concerned about data on teenage sexuality. Some recent studies suggest that sexual behavior is increasing among teenagers. One in particular indicates that three-fourths of all girls have had sex during their teenage years and that 15% have had four or more partners. Sex education programs should begin in early adolescence and should include explicit language about AIDS, the panel says. In Europe, where sex education generally is more frank, the rate of unwanted adolescent pregnancies and venereal diseases is lower, said Thomas J. Coates, of the University of California at San Francisco School of Medicine, and a panel member.

That view runs smack up against that of former Secretary of Education (and now Bush Administration "drug czar") William J. Bennett, who has long taken the high moral ground by arguing chastity and criticizing condom use. Moses said that the report "is in direct contradiction with that point of view."

A "Just Say No" approach to sexuality reflects "a kind of moral mean-spiritedness," said panel member Marshall Becker of the University of Michigan. "You're saying 'If you refuse to play the game, die!' That's a terrible attitude."

Public education programs should also include explicit language "in the idiom of the target audience" and should be designed for specific audiences. Television networks should air more public service spots about AIDS and accept commercial condom advertising, the panel says.

Because intravenous drug users, their sexual partners, and children are at high risk for contracting the virus, the panel says studies of their sexual behavior are a high priority. The panel also endorses wider use of needle-exchange and needle-sterilization programs and increased availability of drug treatment programs.

HIV antibody testing should be widely available, and laws are needed to prevent discrimination against those who test positive, the panel reports.

The number of behavioral and social scientists at federal agencies responsible for AIDS-related programs should be "substantially increased," the panel says.

The panel's report has a familiar ring. Many of its recommendations echo those of earlier reports by the National Academy of Sciences/Institute of Medicine and the Presidential Commission on the Human Immunodeficiency Virus. The Reagan Administration gave both reports the cold shoulder. It remains to be seen how the Bush Administration will react to this one.

■ GREGORY BYRNE