Standardized Tests as a Tool of Exclusion:

Improper Use of the SAT in New York

Kary L. Moss†

INTRODUCTION

Christine Perlin,[1] a senior at a high school in Queens, is second in her class of 316 students. She has had an exceptional high school career; her grade point average is ninety-five, and she ranked first in her class in the tenth and eleventh grades. Christine will be the first in her family to attend college and hopes to go to Columbia University to study molecular engineering. She scored 1000 on the Standardized Aptitude Test (SAT) out of a possible 1600. Eve Hoyt, a senior at a competitive New York public high school which does not rank its students, has also had an exceptional career. She graduated with a ninety-two grade point average, and is a founder of Children for Worldwide Peace, an organization dedicated to promoting social justice. Eve is involved in technical theater at her high school, and has written for and edited a student newspaper. She is also a staff member and contributing member of the school's literary magazine, and has toured the Soviet Union, singing in a group called Peace Child, which required her to learn Russian. She scored 1290 on the SAT.

Despite their impressive records in high school, it is not likely that either young woman would qualify for a New York State Regents College Scholarship—which awards $250 each to 25,000 students and is

---

† Ms. Moss, a staff attorney at the Women's Rights Project (WRP) of the American Civil Liberties Union, was one of several attorneys representing the plaintiffs in Sharif v. New York State Education Department, 709 F. Supp. 345, 347-48 (S.D.N.Y. 1989). The other attorneys were Isabelle Katz Pinzler, Director of the WRP, Joan Bertin, Associate Director of the WRP, Deborah Ellis, staff attorney at the WRP, Robert Levy, staff attorney at the New York Civil Liberties Union, and Helen Hershkoff, Associate Legal Director of the A.C.L.U. The law firm of Stroock, Stroock and Lavan provided valuable pro bono assistance. Pat Campbell, Peggy Chase, Phyllis Rosser, Martin Shapiro, Carol Tittle, FAIRTEST, and the New York Public Interest Research Group (NYPIRG) were indispensable. Personal thanks go to my parents, my husband Doug Baker, Cindy Stagoff, and Dinesh Khosla.

[1] Names have been changed to protect the privacy of the women discussed herein.

renewable for up to five years—or an Empire State Scholarship of Excellence—which awards $2,000 to the top 1,000 students and is also renewable for up to five years—as long as the state considers only their SAT scores. Those scores, while good, are not good enough.

Whether a student wins a scholarship can determine which college she will attend or whether she will be able to attend at all. The criteria used to award scholarships is determinative: New York's practice of relying exclusively on the SAT has cost females, who in recent years have averaged approximately ten verbal points and forty to fifty math points less than males on the SAT,[2] an estimated three million dollars per year.[3] We do not know the racial breakdown of that distribution, but since we know that most women of color do less well than white women on the SAT,[4] scholarships distributed solely on the basis of SAT scores presumably hurt the former most seriously.

The unfairness of New York state's practice is clear from figures which indicate that females had a higher mean grade point average than males in 1988.[5] Grade point average is widely regarded by most educators, including The College Board and the Educational Testing Service (ETS), the creators of the SAT and other standardized tests, as the best measure of high school achievement.[6] When used alone, test scores are *not* viewed as an adequate measure of aptitude and achievement. ETS has warned:

> In order to make useful comparisons of students' performance on tests, a common test given to *all* students would be required. Because the percentage of students taking the SAT and Achievement Tests varies widely and the test takers are self-selected, the SAT is inappropriate for these purposes.[7]

Indeed, the Code of Ethics promulgated by the National Association of

---

[2] EDUCATIONAL TESTING SERVICE, COLLEGE-BOUND SENIORS: 1988 PROFILE OF SAT AND ACHIEVEMENT TEST TAKERS iii (1988) [hereinafter 1988 PROFILES]; M. J. CLARK AND J. GRANDY, SEX DIFFERENCES IN THE ACADEMIC PERFORMANCE OF SCHOLASTIC APTITUDE TEST TAKERS 1 (College Board Report No. 84-8, 1984).

[3] This figure was calculated by the New York Public Interest Research Group, one of the major forces behind the reform of New York state's scholarship program. [Telephone interview with Blair Horner, New York Public Interest Research Group (June, 1989)].

[4] See *infra*, note 18.

[5] 1988 PROFILES, *supra* note 2, at 10; J. Lowen, P. Rosser and J. Katzman, Gender Bias in SAT Items 14 (April 5, 1988) (available from the American Educational Research Association) [hereinafter Rosser]. For students with an A+ average, 53% were female; for an A average, 58% were female; for an A− average, 57% were female; for a B average, 54% were female; for a C average, 44% were female; and for a D average or less, 36% were female. 1988 PROFILES, *supra* note 2, at 2. For statistics from 1973 to 1983, *see* EDUCATIONAL TESTING SERVICE, COLLEGE BOUND SENIORS: ELEVEN YEARS OF NATIONAL DATA FROM THE COLLEGE BOARD'S ADMISSIONS TESTING PROGRAM (1984).

[6] ETS tells students that "[y]our high school record is probably the best single indicator of how well you will do in college, but a combination of your high school grades and test [SAT] scores is an even better indicator." EDUCATIONAL TESTING SERVICE, TAKING THE SAT: THE OFFICIAL GUIDE TO THE SCHOLASTIC APTITUDE TEST AND TEST OF STANDARD WRITTEN ENGLISH 4 (1988-89 ed.).

[7] 1988 PROFILES, *supra* note 2, at iii.

College Admission Counselors (NACAC), requires member institutions to refrain from using minimum test scores as the sole criterion for admission, to use test scores in conjunction with other data such as school record and recommendations, and to avoid discriminating against students whose scores may reflect bias.[8] The SAT needs the corroboration of other data precisely because it excludes so many qualities which are integral to future success.[9]

Moreover, exclusive use of the SAT as a measure of high school performance is inappropriate because the SAT has never been validated as a tool to assess high school achievement; it has only been validated to predict first-year performance in college. Well-established professional standards require test users to undertake statistical analysis ("validation") in order to ensure that the major types of inferences drawn from a test are valid.[10] Courts have long rejected use of tests which fail to follow proper validation procedures.[11] It is therefore not appropriate to award scholarships—which in New York were designed to reward high school

---

[8] NATIONAL ASSOCIATION OF COLLEGE ADMISSIONS COUNSELORS, STATEMENT OF PRINCIPLES OF GOOD PRACTICE C (December 1988). NACAC's members include 4,000 educational institutions. The Code of Ethics was developed in tandem with the American Association of Collegiate Registrars and Admission Officers and The College Board, and endorsed by the American Council on Education, the National Association of Secondary School Principals, and the American School Counselor Association.

[9] ETS and The College Board explained their policy proscribing exclusive use of SAT scores thus:

> [T]he basis for the guidelines is the professional judgment that types of information other than that available from the SAT, and other sources of information (besides the SAT) about students' cognitive and other abilities, are valuable. . . . For example, other information may reflect creativity, industriousness, or self-discipline that are not measured by standardized tests, such as the SAT. In addition, the guidelines respond to a variety of other significant educational policy considerations, including the general importance of recognizing talent and achievement in many forms.

Brief for The College Board and Educational Testing Service as *Amici curiae* at 9, Sharif v. New York State Education Dept., 709 F. Supp. 345 (S.D.N.Y. 1989) (88 Civ. 8435) [hereinafter ETS Brief].

[10] *See* AMERICAN PSYCHOLOGICAL ASSOCIATION, STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING (1985). The College Board requires that tests be validated periodically, "to ensure that they predict the expected outcome at a level acceptable for the institution's particular purpose." THE COLLEGE BOARD, ATP GUIDE FOR HIGH SCHOOLS AND COLLEGES: SAT AND ACHIEVEMENT TESTS 22 (1987). ETS, which adopts the APA guidelines, but establishes the standards for ETS products specifically, requires that tests be validated by procedures that are most appropriate to the intended uses of test scores, that the study describe the appropriateness of the validity evidence to the intended use of the test, and that the study document the results of the analysis and validation. EDUCATIONAL TESTING SERVICE, ETS STANDARDS FOR QUALITY AND FAIRNESS 7-9 (1987).

[11] *See* Anderson v. Banks, 520 F. Supp. 472, 490 (S.D. Ga. 1987), *appeal dismissed*, 730 F.2d 644 (11th Cir. 1984) (use of CAT test must be validated when it is used for a purpose other than that for which it is designed); York v. Ala. State Bd. of Educ., 581 F. Supp. 779, 786 (M.D. Ala. 1983) (use of NTE should be validated if there is evidence of adverse racial impact). Even where a particular use of a test has been upheld, courts have acknowledged that tests which are demonstrably invalid for the purpose for which they are used are unlawful if they have a discriminatory impact. Debra P. by Irene P. v. Turlington, 730 F.2d 1405, 1409 (11th Cir. 1984). *See also* United States v. LULAC, 793 F.2d 636, 647 (5th Cir. 1986) (invalidity of test can be evidence of discriminatory purpose).

performance—by using a vehicle which does not in fact accurately measure high school performance.

The tendency to place inappropriate emphasis on test results is not a new phenomenon.[12] The most frequent challenges to the use of tests have been on the basis of racial bias.[13] However, due to an absence of data on the impact of the New York scholarship competition on people of color, the most recent case, *Sharif v. New York State Education Department*,[14] was premised solely on charges of gender bias. The plaintiffs were the two young women described above, eight others, and two non-profit organizations. Based on Title IX of the Education Amendments of 1972,[15] which prohibits discrimination in educational institutions that receive federal funds,[16] and the equal protection clause of the fourteenth amendment, they were able to secure a preliminary injunction prohibiting exclusive reliance on the SAT in selecting New York scholarship winners.

This discussion focuses on *Sharif* and how testing practices have posed, and continue to pose, a barrier to post-secondary education for women and people of color. I will devote specific attention to the phenomena of gender bias in the SAT. Failure to confront the complexity of the issue means that children and teenagers will be saddled with the cost of our failure to develop selection processes which are gender and color blind.

THE SAT AND GENDER BIAS

The young women plaintiffs in *Sharif* are not unusual; generally, females trail males on the SAT by an estimated sixty points. African-American women, who constitute nearly sixty percent of African-American test takers, score an estimated eight points lower on the verbal section and twenty-four points lower on the math section than African-American men.[17] Generally, women of color perform less well than

---

[12] *See* J. CROUSE AND D. TRUSHEIM, THE CASE AGAINST THE SAT (1988); S. GOULD, THE MISMEASURE OF MAN (1981); D. OWEN, NONE OF THE ABOVE: BEHIND THE MYTH OF SCHOLASTIC APTITUDE (1985).

[13] *See* Georgia State Conf. of Branches of NAACP v. State of Ga., 775 F.2d 1403, 1416 (11th Cir. 1985) (achievement grouping); Larry P. by Lucille P. v. Riles, 793 F.2d 969 (9th Cir. 1984) (IQ tests); United States v. Gadsen County School District, 572 F.2d 1049, 1051 (5th Cir. 1978) (per curiam) (ability grouping); Morales v. Shannon, 516 F.2d 411, 414 (5th Cir. 1975) (ability grouping); Moses v. Washington Parish School Bd., 456 F.2d 1285 (5th Cir. 1972) (per curiam) (ability and achievement tests); Montgomery v. Starkville Mun. Separate School Dist., 665 F. Supp. 487, 495 (N.D. Mass. 1987) (achievement grouping).

[14] 709 F. Supp. 345 (S.D.N.Y. 1989).

[15] 20 U.S.C. § 1681(a).

[16] Title IX provides in relevant part: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance. . . ." 20 U.S.C. § 1681(a).

[17] The College Board, Press Release (Sept. 20, 1988).

white women.[18]

These score differences are particularly meaningful when they do not accurately predict first-year college performance, which is what the tests are designed to do. ETS and The College Board argue that the correlation between SAT scores and first year college grades for females is higher than it is for males.[19] However, they do not dispute that the SAT *underpredicts* the first-year college scores of females and *overpredicts* the first-year college scores of males.[20] This means that females must score much higher than males on the SAT in order to be competitive in the college admissions process.

A look at what is called the "regression line"—a statistical measurement used to indicate if a test is biased—illustrates the problem. Regression analysis predicts a "dependent variable" (in this case, first-year college grade point average) by using information contained in one or more independent variables (SAT scores or both SAT scores and high school grade point average).[21] If the SAT is gender blind, first-year college grades will be approximately the same regardless of whether they are computed with female or male student data. Instead, the regression line indicates that females must score higher on the SAT to get the same predicted grades as males.[22] Consequently, scholarship or educational

---

[18] According to 1988 statistics released by The College Board, the mean scores for women on the verbal portion (v) and math portion (m) of the SAT are as follows:

	Mean Verbal	Mean Math
Asian-American Females	404	499
White Females	439	468
American-Indian Females	387	418
Mexican-American Females	373	410
African-American Females	350	374
Puerto Rican Females	348	384

The College Board, 1988 SAT Profile: Profiles of SAT and Achievement Test Takers 6-7 (available from The College Board) (1988).

[19] ETS Brief, *supra* note 9, at 4.

[20] *Id.* at 6. They ascribed this overprediction to differences in course selection by males and females at the college level and different grading practices in the courses selected.

[21] Performance on a test is checked against a "criterion," a direct and independent measure of that which the test is designed to predict, which indicates the effectiveness of a test in predicting an individual's behavior in specified situations. *See* A. ANASTASI, *Validity: Basic Concepts*, in PRINCIPLES OF PSYCHOLOGICAL TESTING 140 (1976). For a scholastic aptitude test, such as the SAT, the criterion tends to be first-year college grades or grade point average (GPA).

[22] For example, the common regression line might indicate that both males and females who score 1100 will receive at least a 3.0 grade point average. However, due to the SAT's gender bias, females must score *higher* than males to have the same predicted grades. This creates a sort of double bias against females; not only must they score higher than males to get the same *predicted* grades, but they must score higher still to reflect their *actual* grades, since females tend to actually receive higher grades than males with the same SAT scores. Similar results have been found in the American College Testing Program (ACT) test which is used primarily by students living in the Mid-West. *See* Gamache and Novick, *Choice of Variables and Differentiated Prediction Within Selected Academic Programs*, 22 J. EDUC. MEASUREMENT 53 (1985). For a discussion of the relationship between gender and performance on mathematics questions on the ACT, *see* Doolittle and Cleary, *Gender Based Differential Item Performance in Mathematics Achievement Items*, 24 J. EDUC. MEASUREMENT 157 (1987). *See also* A. Doo-

decisions based on SAT scores will exclude many females simply because the test underpredicts their college grades.

Few agree on why women and people of color score lower on the SAT than do white males.[23] In the past, some educators and testing experts explained low test scores of people of color as due to their "inferior cognitive abilities."[24] While such an explanation is no longer acceptable for people of color, some educators and testing experts will still rely on it to explain lower math scores of women.[25] However, most ascribe low scores, both for women and people of color, to a complex matrix of social forces which include inadequate course preparation,[26] inadequate high school counseling,[27] language difficulties,[28] socioeconomic status,[29] lower parental education levels,[30] varied grading practices among selected courses,[31] and real cultural differences.[32] In addition, The Col-

---

little, Gender Differences in Performance on Mathematics Achievement Items (Aug. 1987) (paper presented at the Annual Meeting of the American Psychological Association).

23 While the gender problem results from underprediction of women's first-year college grades, the SAT appears to overpredict the first-year college grades of African-Americans. Apparently, the overprediction is reduced when high school grades are considered. *See* J. CROUSE AND D. TRUSHEIM, *supra* note 12, at 96-121. Nevertheless, this means that the test still is not doing what it should be doing, which is to predict performance *accurately*. Moreover, "overprediction" is not meaningful if the test itself reflects what one author has called "errors in measurement"; standardized tests, as currently designed, penalize those of the non-dominant culture, thus masking their accomplishments and abilities and magnifying their differences. White, *Culturally Biased Testing and Predictive Invalidity: Putting Them on the Record*, 14 HARV. C.R.-C.L. L. REV. 99 (1987).

24 *See* S. GOULD, *supra* note 12, for a classic and compelling discussion of this history.

25 Goldberg, *Numbers Don't Lie: Men Do Better Than Women*, N.Y. Times, July 5, 1989, at A21 (arguing that "SAT scores accurately reflect male superiority in math"). *But see* Sheehan, *SAT Supports Illusions of Male Grandeur*, N.Y. Times, July 21, 1989, at A28.

26 Hearings on H. 111 Before the Subcomm. on Civil and Constitutional Rights, 100th Cong., 1st Sess. 54 (1987) (statement of Gretchen Rigol, Executive Director of Access Services, The College Board).

27 J. Lubetkin, The Scholastic Aptitude Test: A Valid and Unbiased Predictor of College Performance? 31 (rev. ed. Oct. 1988) (unpublished senior thesis presented to the faculty of the Woodrow Wilson School of Public and International Affairs, Princeton University).

28 Hoover, Politzer, and Taylor, *Bias in Reading Tests for Black Language Speakers: A Sociolinguistic Perspective*, 38 NEGRO EDUC. REV. 81 (1987); Schmitt, *Language and Cultural Characteristics That Explain Differential Item Functioning for Hispanic Examinees on the Scholastic Aptitude Test*, 25 J. EDUC. MEASUREMENT 1 (1988); A.P. Schmitt and N.J. Dorans, Differential Item Functioning for Minority Examinees on the SAT 10 (1987) (paper presented at the Annual Meeting of the American Psychological Association) (summarizing studies assessing test-taking abilities of Asian-Americans); Taylor and Lee, *Standardized Tests and African-American Children: Communication and Language Issues*, 38 NEGRO EDUC. REV. 67 (1987) (identifying various forms of cultural and language bias).

29 *See* ETS Brief, *supra* note 9, at 5; B. Berne and E. Tobier, The SAT Scores of Seniors in New York City Public High Schools 17-20 (March 1988) (unpublished draft) [hereinafter N.Y.U. Report].

30 Rigol, *supra* note 26, at 54; Rosser, *supra* note 5, at 21-22; N.Y.U. Report, *supra* note 29, at 21-23; ETS Brief, *supra* note 9, at 5.

31 ETS Brief, *supra* note 9, at 6.

32 Nobles, *Psychometrics and African-American Reality: A Question of Cultural Antimony*, 38 NEGRO EDUC. REV. 45 (1987); White, *supra* note 23, at 97 ("[A]chieving high scores on standardized tests might require conformity to, or at least awareness of, a widely held set of assumptions and possession of widely shared knowledge. . . . Members of minority groups are unlikely to possess such 'standardized' knowledge and beliefs"). Factors that can produce cultural bias in standardized tests include intentional bias, as where tests are used to exclude one group, requiring or assuming knowledge of information that is more familiar to one group

lege Board credits lower test scores to the increasing numbers of women now taking the test.[33]

Perhaps the most striking problem with these arguments, and it cannot be underscored enough, is that the SAT is not a valid aptitude test precisely because gender, income, cultural background, and the other variables enumerated above do have an effect on test scores. Moreover, even after these factors are controlled for, there is evidence that bias remains. For example, in one study, SAT scores were correlated with variables such as high school grades and years of study devoted to various subjects. Gender bias was still found to persist among *both* college math and social science majors.[34] The success of coaching courses among women and people of color also belies the claim that the SAT is an accurate measurement of aptitude.[35]

The argument that the socioeconomic status, or poverty, of the parent or parents completely explains these lower scores is also troublesome.[36] While few dispute that socioeconomic status affects test scores, bias persists even at the highest income levels ($70,000 and above) for

---

than another, using subject matter that is more interesting to one group, using terms which are more familiar to one group, and imposing a time limit on test takers. Additionally, test takers from minority groups are more likely to be unfamiliar with standardized tests, and this can cause cultural bias as well. White, *supra* note 23, at 108-12.

These factors also apply to gender, as in the case of time limits. Assistant Professor of Mathematics Kate Sheehan observed that sex differences in mathematical problem solving may be ascribed to timed tests which reward "the facile test taker, not the thoughtful thinker who gathers information and organizes, evaluates, and expresses ideas clearly, perceives subtle relationships and patterns, and makes reasonable estimates and predictions." Sheehan, *supra* note 25, at A28.

[33] Rigol argues that "higher proportions of test takers . . . result in lower test scores for the group." Rigol, *supra* note 26, at 53. However, this argument does not address the fact that female high school grade point average and rank have been consistently higher than their test scores indicate. M. J. CLARK AND J. GRANDY, *supra* note 2, at 5-6.

[34] M. J. CLARK AND J. GRANDY, *supra* note 2, at 4. Recent figures released by The College Board bear this out: of those taking the 1988 SAT, 88% of the females had four years of English, as compared to 86% of the males, 60% of the females had four years of math, as compared to 68% of the males, 84% of the females had three years of social sciences, as compared to 83% of the males, 73% of the females had three years of natural science, as compared to 79% of the males, and 88% of the females had two years of foreign language, as compared to 82% of the males. 1988 PROFILES, *supra* note 2, at v.

[35] *See* R. Curran, The Effectiveness of Computerized Teaching for the PSAT and SAT (1988) (unpublished doctoral thesis available at Boston University School of Education); J. Zuman, The Effectiveness of Special Preparation for the SAT: An Evaluation of a Coaching School (Apr. 1988) (paper presented at the 1988 meeting of the American Educational Research Association); Educational Testing Service, *Minority Students in Higher Education*, 22 FOCUS 1, 13-15 (1988) [hereinafter FOCUS 22]; N.Y.U. Report, *supra* note 29, at 7-8.

[36] *See* Larry P. v. Riles, 495 F. Supp. 926, 956 (N.D. Cal. 1979), *aff'd in part and rev'd in part on other grounds*, 793 F.2d 969, 975 (9th Cir. 1984), a case in which African-American parents challenged a school's use of intelligence tests to place their children in "educable mentally retarded" classes, because of the test's known racially discriminatory impact. The district court rejected the defendant's argument that African-American children tend to be retarded more often than white children because more African-American people tend to be poor, and poor pregnant women tend to suffer from inadequate nutrition, ultimately causing brain damage to their children. The court reasoned that this argument did not explain why severe mental retardation does not occur in greater proportion among African-Americans and poor sections of the population.

both women and people of color.[37]

It is thus clear that the SAT cannot be relied upon as a perfect measure of aptitude and that some of the problem lies with the test, as ETS has conceded:

> [T]he research literature either finds no difference between men and women in performance on cognitive skills, or finds a slight advantage for females on verbal skills and a slight advantage for males on mathematical and spatial skills. Male-female biological differences do not appear to explain the observed difference in cognitive functioning; experiences, stereotypes, and expectations no doubt play a role, but it has been difficult to identify specific ways in which they may account for difference in academic performance. In addition, the measures we use may contribute to the differences we observe.[38]

Accordingly, ETS and The College Board review their tests to minimize these and other problems,[39] and claim to weed out biased questions in the SAT on a regular basis, although some testing experts have questioned their good faith.[40] We know that past efforts to eliminate gender bias from intelligence tests have been successful.[41] There is no reason why they should not also be successful in the case of the SAT.

In the meantime, the question becomes who should bear the cost of the test's problems—the test user or the test taker? ETS believes that the discriminatory effects of the test should not be the test users' problem, as

---

[37] Mean verbal and math scores for those in the income bracket of $70,000 and above, which are the highest scorers, were as follows for 1988:

	Mean Verbal	Mean Math
White Males	476	542
White Females	465	502
Puerto Rican Males	421	489
Puerto Rican Females	420	463
African-American Males	419	461
African-American Females	421	440

The College Board, *supra* note 18, at 6-7.

[38] M. J. CLARK AND J. GRANDY, *supra* note 2, at 4.

[39] Fiske, *Changes Planned in Entrance Test Used By Colleges*, N.Y. Times, Jan. 3, 1984, at 1A.

[40] Owen describes the review process as cursory and inadequate. The process is conducted by "an actual member of a minority" who:

> simply counts the number of items that refer to each of the "population subgroups" and enters these numbers on a Test Sensitivity Review Report Form. On the verbal SAT administered in May 1982, actual minority member Beverly Whittington found seven items that mentioned women, one that mentioned Black Americans, two that mentioned Hispanic Americans, none that mentioned Native Americans, and four that mentioned Asian Americans (actually, she was stretching here; these particular Asian Americans were Shang Dynasty Chinese, 1766-1122 B.C.). Two items overlapped, so Whittington put a "12" in the box for Total Representational Items. She also commented "OK" on the exam's test specifications, "OK" on the subgroup reference items, and "OK" on item review. She made no other remarks. If she had discovered the word "nigger" in one of the questions, presumably she would have scratched it out. ETS made Whittington take a three-day training program in "test sensitivity" before permitting her to do all of this. When her report was finished, it was stamped E.T.S. CONFIDENTIAL and SECURE. Then it was filed and forgotten.

D. OWEN, *supra* note 12, at 217.

[41] *See Larry P.*, 495 F. Supp. at 955.

shown by its failure to enforce its own guidelines for test use.[42] It is not clear why these guidelines exist at all if they may be followed solely at the discretion of a test user, particularly if the test user is already disinclined to worry about any discriminatory impact that its use of a test may have, as was the case in the New York scholarship competition. ETS' reluctance to police the users of its products thus shifts the burden to the courts who, arguably, should not have to monitor scholarship competitions.[43] New York state similarly decided to let women and people of color pay the price when it chose to use a scholarship selection device that it knew would have a discriminatory impact. It was that decision that led to *Sharif*.

THE NEW YORK CASE: *SHARIF V. NEW YORK STATE EDUCATION DEPT.*[44]

Due to the bias in the SAT, it was foreseeable that if SAT scores were used alone, scholarship distribution would not reflect actual high school achievement, as both the Regents and Empire Scholarships were designed to do.[45] Nevertheless, since 1977 the State Education Department (SED), formally charged by the Board of Regents with operating the scholarship program, has awarded state scholarships based exclusively on applicants' SAT scores with a predictable discriminatory result. In 1986-87, the last year in which scholarships were distributed solely on the basis of SAT scores, females received only forty-three percent of the Regents Scholarships and twenty-nine percent of the Empire Scholarships, although they comprised fifty-three percent of the applicants.[46]

---

[42] For example, in its *amicus* brief in *Sharif*, ETS argued that its own guidelines prohibiting use of the SAT alone to make educational decisions should not constitute a binding obligation on the test users. They said:

> The guidelines were adopted by the organizations to establish particular standards and goals for themselves and users of the tests they sponsor and develop, but they have not been adopted by a public body and were not intended to incorporate legal requirements. To convert voluntary, private guidelines like these into binding legal requirements irrespective of the particular factual circumstances confronting the test user would be unfair to all parties concerned and operate as a serious disincentive to the development and improvement of such private guidelines.

ETS Brief, *supra* note 9, at 10.

[43] ETS' unwillingness to monitor test users is understandable, however, given the amount of money ETS makes a year from use of its tests. ETS' 1983 annual report, for example, lists its tax-free revenue at $133 million, half of which was generated by The College Board programs. D. OWEN, *supra* note 12, at 7.

[44] 709 F. Supp. 345 (S.D.N.Y. 1989).

[45] The Regents Scholarship was created by Act approved Apr. 16, 1913, ch. 292, 1913 N.Y. Laws 527. The legislature would reaffirm this purpose for the Regents Scholarship throughout the years. Act of 1974, ch. 942, 1974 N.Y. Laws 602, 604, 605; Act approved Apr. 18, 1986, ch. 56, 1986 N.Y. Laws 140; Act approved Aug. 7, 1987, ch. 837, 1987 N.Y. Laws 1, 2. *See Tenth Annual Report of the Education Department* (March 16, 1914); *Report of the Select Committee on Higher Education,* State of New York Legislative Doc., No. 16 (1974). For information on the rationale behind the establishment of the Empire Scholarship which was created in 1987, *see* Governor's Approval Memo on ch. 56, appendix, 1986 N.Y. Laws 3153.

[46] University of the State of New York and the State Education Dept., Statistical Review of the

Citing evidence of gender bias in the SAT, the legislature amended the Education Law, in July 1987, to require SED to award scholarships "based on a formula which includes high school performance and which may include nationally competitive examinations" for one year.[47] SED settled on a formula weighing test scores and grade point average equally. In March 1988, the new scholarship winners were announced. Women, who comprised fifty-three percent of the applicants, won more scholarships than they ever had. In the case of the Empire State Scholarships, they received thirty-eight percent, rather than the twenty-nine percent they received when only SAT scores were considered. For the Regents Scholarship, women received forty-nine percent of the scholarships, rather than the forty-three percent they received when SAT scores alone were used.[48] No statistics were made available to the public on the impact of the new selection device on people of color.

The new scholarship distribution method did not, as many critics charged, displace "deserving" men. Instead, one expert witness testified that the new policy only took scholarships away from those who had high SAT scores and low grade point averages and gave them to those with slightly lower SAT scores and higher grade point averages.[49] For both scholarships, the mean grade point average for males and females was approximately the same: 85 for females and 84.4 for males. Females averaged 916 as compared to 981 on the SAT.[50] Thus, the inclusion of grades substantially reduced the impact of the scholarship program's gender bias and rewarded those who had performed consistently well in their high school years.

Based upon this statistical evidence, plaintiffs sued in federal district court charging SED with violations of Title IX and the equal protection clause of the United States Constitution. In a precedent setting ruling, the court held that the plaintiffs did not have to show discriminatory intent to establish a prima facie case of discrimination under the Title IX implementing regulations.[51] Judge Walker compared Title IX regulations to Title VI regulations, which explicitly forbid programs receiving federal financial assistance from discriminating on the basis of race, color, or national origin, and have been found to prohibit practices with a

---

Awarding of the 1988 New York State Scholarships 3 (Apr. 1988). [hereinafter Statistical Review].

47 Act approved Aug. 7, 1987, ch. 837, 1987 N.Y. Laws 1599.

48 Statistical Review, *supra* note 46, at 4.

49 Testimony of Dr. Martin M. Shapiro at Preliminary Injunction Hearing (Jan. 23, 1989), Sharif v. New York State Education Dept., 709 F. Supp. 345 (S.D.N.Y. 1989) (88 Civ. 8435).

50 Statistical Review, *supra* note 46, at 6.

51 *Sharif*, 709 F. Supp. at 361. *See also* 34 C.F.R. § 106.3 (1988). Subsection (c)(1) requires each recipient educational institution to evaluate the effects of its current policies and practices. Subsection (c)(2) requires recipients to modify any policies and practices which do not meet the regulations' requirements, and subsection (c)(3) requires recipients to take remedial steps to eliminate the effects of any discrimination.

discriminatory impact.[52] Judge Walker found the Title IX regulations to be similar,[53] and held that those regulations also prohibit practices with discriminatory consequences.[54]

Having established that the challenge to the SAT-only policy could be reviewed using disparate impact analysis under Title IX implementing regulations, the court turned to traditional Title VII methods of analyzing disparate impact claims because of the considerable body of available Title VII testing cases—in particular, *Griggs v. Duke Power Co.*[55] and *Albemarle Paper Co. v. Moody*.[56]

The plaintiffs were required to demonstrate the test's discriminatory impact in order to establish a prima facie case. The court held that the plaintiffs met their burden through their presentation of undisputed statistical evidence and expert testimony. SED blamed the results on "neutral" variables, such as socioeconomic status and parental education levels, but the court rejected this attack:

> [S]tatisticians have attempted to explain the score differentials between males and females by removing the effect of "neutral" variables, such as ethnicity, socioeducational status (parental education), high school classes, and proposed college major. However, under the most conservative studies presented in evidence, even after removing the effect of these factors, at least a thirty point combined differential remains unexplained.[57]

To rebut the prima facie case, the court required SED to show that the practice was justified by "educational necessity"—that is, that it bore "a manifest demonstrable relationship to classroom education."[58] Thus, SED had to show a manifest relationship between use of the SAT—which is designed to predict performance in college and as such is properly called an "aptitude test"—and recognition and reward of academic achievement in high school, the stated intent of the scholarship.[59]

SED argued that the SAT is really an achievement, not an aptitude,

---

[52] *See* Guardians Association v. Civil Service Commission, 463 U.S. 582 (1983).

[53] 34 C.F.R. § 106.3(c) (1988) generally requires, for example, recipient educational institutions to evaluate the effects of their policies and practices. 34 C.F.R. § 106.21(b)(2) (1988) prohibits recipients from "administer[ing] or operat[ing] any test or other criterion for admission which has a disproportionately adverse effect on persons on the basis of sex unless the use of such test or criterion is shown to [validly predict] success in the education program or activity in question."

[54] *Sharif*, 709 F. Supp. at 361.

[55] 401 U.S. 424 (1971).

[56] 422 U.S. 405 (1977). For an educational testing case which uses a Title VI analysis, *see* Georgia State Conference of Branches of NAACP v. State of Ga., 775 F.2d 1403 (11th Cir. 1985). While the Court's decision this year in Wards Cove Packing Co. v. Atonio, 109 S.Ct. 2115 (1989) is extremely troublesome for future Title VII litigation, it should not prove problematic for Title IX litigation primarily because the Court's fear that employers would enact quotas to avoid rigorous legal standards is not a concern presented by Title IX cases.

[57] *Sharif*, 709 F. Supp. 345, 355 (S.D.N.Y. 1989).

[58] *Id.* at 361, citing Georgia State Conference of Brotherhood of NAACP v. State of Ga., 775 F.2d at 1418. *See* Board of Education, N.Y.C. v. Harris, 444 U.S. 130, 151 (1979); Larry P. by Lucille P. v. Riles, 793 F.2d 969, 982 n.9 (9th Cir. 1984).

[59] *See* United States v. LULAC, 793 F.2d 636, 649 (5th Cir. 1986); Debra P. by Irene P. v. Turlington, 730 F.2d 1405, 1409 (11th Cir. 1984).

test. The court rejected this defense, reasoning that the math and verbal sections of the SAT constitute only twenty percent of a student's high school studies. The SAT does not measure a student's achievements in foreign languages, the sciences, or social studies, or a student's diligence, creativity, social development or work habits, all of which are integral to high school performance. The court reflected: "[a]fter a careful review of the evidence, this Court concludes that SAT scores capture a student's academic achievement no more than a student's yearbook photograph captures the full range of her experiences in high school."[60]

Even if SED had been able to show a correlation between SAT performance and academic achievement in high school, the court noted that SED failed to satisfy the validation requirement, and defied professional guidelines, by using SAT results alone to make scholarship decisions. For all of these reasons, the court held that the classification of scholarship applicants solely on the basis of SAT scores was not "rationally related" to the state's goal of rewarding high school achievement. Accordingly, SED was found to have violated the equal protection clause of the fourteenth amendment. The court granted the preliminary injunction and ordered SED to not rely exclusively on SAT scores but instead to also consider high school grades in determining scholarship winners. However, the court left the weighting of grades and test scores, as well as the development of future alternatives for awarding scholarships, to SED's discretion.[61]

SED decided to weigh SAT scores and grades equally in this year's competition. Females, who make up 53.2% of the test takers, will receive approximately forty-nine percent of the Regents Scholarships and thirty-eight percent of the Empire Scholarships. Seven of the ten plaintiffs, including Christine Strutt and Eve Hoyt, have received Regents Scholarships; SED conceded that three of the plaintiffs would not have received a scholarship under an SAT-only policy. None of the plaintiffs will receive an Empire Scholarship. The solution is not perfect. However, it is an improvement over past practices and establishes a precedent for future challenges.

---

[60] *Sharif*, 709 F. Supp. at 362.

[61] SED has maintained that use of grades is unworkable because school districts will cheat and inflate their grades, and students will avoid difficult courses to get better grades. *Sharif*, 709 F. Supp. at 352-53. This argument is problematic for several reasons. First, SED offered no evidence that such suppositions were true except letters from school districts that disagreed with the policy. *See, e.g.*, Letter from William McPhee to SED (Sept. 9, 1987); Letter from Raymond J. Cenni to SED (Sept. 10, 1987); Letter from Richard Hibshman to SED (Sept. 11, 1987). In fact, some of the difficulties caused by the use of grades may have been a result of SED's failure to issue specific instructions to schools on how grades should be reported. *Sharif*, 709 F. Supp. at 352-53. Second, the argument misses the point that those school districts that inflate their grades should be the ones punished, not the students who have demonstrated superior academic achievement in high school.

IMPLICATIONS

Despite the court's ruling, three of the ten plaintiffs—two of whom were African-American and without financial resources—did not qualify for either scholarship award because their SAT scores still were not good enough. One of them ranked eleventh in her class of 255, and received an Adam Clayton Powell leadership award, yet scored in the low 900s on the SAT. She has received an offer of a partial scholarship from New York University, but will not be able to attend without more financial aid. The other ranked fifth in her class of 339 and was named class valedictorian, yet scored in the low 700s.

As long as women and people of color perform less well than white men on standardized tests, the use of these tests will have a foreseeable discriminatory impact that will limit their access to higher education. It is therefore incumbent on test users to minimize this impact by relying on tests only for the purposes for which they were designed, and by never using them alone to make important educational decisions. When tests are misused, test makers, such as ETS, should withhold test scores until the test user complies with professional testing guidelines. If ETS had done this in New York state, *Sharif* would have been avoided.

Fortunately, the principles of validation and non-exclusive use of SAT scores, as applied in *Sharif*, apply equally to other selection devices which limit access to higher education. Intelligence examinations, for example, formerly used to track students in the first years of their education, are no longer defended as a vehicle to predict future performance.[62] Similarly, the bar examination has been attacked because of its questioned ability to accurately measure those qualities deemed essential to good lawyering.[63] Although only one court has addressed the "validity" of the bar examination, it found this failure troublesome:

> While the Bar Examiners do not concede that they would lose under this requirement [Title VII], we believe the record is inadequate to demonstrate either "criterion" ("predictive"), "content," or "construct" validity under professionally acceptable methods. Thus, if we were to determine that Title VII standards were applicable, it would be necessary to reverse and declare the South Carolina Bar Examination constitutionally invalid.[64] (citations omitted)

However, the real issue presented here is the nature of our responsi-

---

[62] In *Larry P.*, the defendants had to show that intelligence tests predicted specifically that Black elementary schoolchildren who score at a certain level on IQ tests are mentally retarded and incapable of learning the regular school curriculum. 763 F.2d at 980.

[63] *See* Emsellem, *Racial and Ethnic Barriers to the Legal Profession: The Case Against the Bar Examination*, N.Y. St. B. J., Apr. 1989 at 44.

[64] Richardson v. McFadden, 540 F.2d 744, 746-47 (4th Cir. 1976), *rev'd in part, on rehearing en banc*, 563 F.2d 1130 (4th Cir. 1977), *cert. denied*, 435 U.S. 968 (1978). *See generally* ABA Section on Legal Education and Admissions to the Bar, Report and Recommendations of the Task Force on Lawyer Competency: The Role of Law Schools 8-13 (1979) (discussing those qualities thought essential to lawyer competency).

bility for the foreseeable discriminatory impact of these tests. Test misuse has real financial consequences: national statistics indicate that the student college debt burden has been growing at an alarming rate throughout the 1980s. A decade ago, loans comprised only seventeen percent of all financial aid, and scholarships and grants accounted for the remainder.[65] Today, loans comprise fifty percent of all financial aid.[66] With women earning sixty-six percent of what men earn,[67] and African-American employees with five or more years of college earning $7,000 less than their white counterparts,[68] the repayment burden is even more onerous for these groups.

Thus, scholarship selection devices which unduly burden these groups have a long-term impact which can affect the very nature of our labor force. As of 1986, women comprised only six percent of our engineers, four percent of our dentists, seventeen percent of our doctors and eighteen percent of our lawyers.[69] Women of color, whose participation rate in the labor force is approximately the same as that of white women (56.6% compared to 55%), still predominate at the very lowest of earning occupations.[70]

There are social costs as well: students must carry these test scores with them for years as measurements of their intelligence.[71] Where overreliance on test scores results in the exclusion of particular groups from colleges, college students will be increasingly isolated from those whose backgrounds are different from their own. For example, more white men (27%) and white women (26%) between the ages of eighteen and twenty-four attended college in 1980 than did Hispanic men and women (16%), African-American women (21%) or African-American men (17%).[72] African-Americans, Hispanics and Native Americans have been attending college in decreasing numbers since 1974, although a greater number have been graduating from high school.[73] The news is even worse for

---

65 FOCUS 22, *supra* note 35, at 5.

66 *Id.*

67 Women's Equity Action League, Report on the Economic Status of Women 2 (1988) [hereinafter W.E.A.L. Report].

68 FOCUS 22, *supra* note 35, at 5.

69 W.E.A.L. Report, *supra* note 67, at 2.

70 *Id.* at 4.

71 For example, the Rosser study surveyed 1112 students who took an SAT coaching course with Princeton Review. The students were mostly white (75.3%) but included 13.2% Asian Americans, 5.2% Blacks, and 2.4% Hispanics. They came from the five boroughs of New York and a variety of schools. Rosser and her colleagues discovered that both males and females perceive their abilities to be more in line with their test scores than their grades. However, even girls who perform well on the math portion tend to have a lower perception of their abilities than do boys. Rosser, *supra* note 5, at 4.

72 W.E.A.L. Report, *supra* note 67, at 4.

73 The proportion of Black high school graduates entering college has dropped from 48% to 44% while the number graduating from high school has increased from 72% to 77%. The attendance rate for Hispanic students has dropped from 49% to 47%, although the number of Hispanics graduating from high school has increased from 55% to 60%. FOCUS 22, *supra* note 35, at 2.

people of color in law school, who constituted only eleven percent of J.D. students and fourteen percent of the applicant pool for the 1986-87 academic year.[74] Social isolation prevents communication and students' familiarization with different cultures; it also prevents them from learning in an environment which is a real microcosm of the world. One recent consequence may be the rise in incidents of racial harassment at campuses throughout the country.[75]

Decreased reliance on biased selection methods known to facilitate social isolation and to limit early educational opportunities will enable colleges to enrich the educational experience for all students by improving their access to diversity. Even more, however, decreased reliance on biased selection methods will force all of us to confront the complexity of the real problems in our educational system, a system which has allowed so many to fall between the cracks.

---

[74] Law School Admission Services, Inc., Access 2000 Data Book: U.S. Minority Educational Enrollment: Law School Application, Enrollment, Placement, and Teaching Patterns 43 (1988); Emsellem, *supra* note 63, at 44.

[75] *See* Weiner, *Reagan's Children: Racial Hatred on Campus*, The Nation, Feb. 27, 1989, at 260. The Nation reported that at Columbia University, "[t]he litany is that black people tend to be criminals, drug addicts, and welfare cheats; that they don't want to work; and that black students aren't as smart as whites."