Standardized Tests: The Continuation of Gender Bias in Higher Education

Andrea L. Silverstein

Follow this and additional works at: http://scholarlycommons.law.hofstra.edu/hlr

Part of the Law Commons

Recommended Citation
Available at: http://scholarlycommons.law.hofstra.edu/hlr/vol29/iss2/8

This document is brought to you for free and open access by Scholarly Commons at Hofstra Law. It has been accepted for inclusion in Hofstra Law Review by an authorized administrator of Scholarly Commons at Hofstra Law. For more information, please contact lawcls@hofstra.edu.
NOTE

STANDARDIZED TESTS: THE CONTINUATION OF GENDER BIAS IN HIGHER EDUCATION

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.1

Just as the Manhattan Project had split the atom, the Educational Testing Service . . . would decode the mind. . . . ETS would measure all abilities, not just aptitude or intelligence. It would map and code the personality. . . . Human nature itself would be reformed.2

I. INTRODUCTION

Gender bias in higher education established its roots centuries ago. In the 1800s, higher education was considered to be dangerous for women, and in accordance with widely held views about a woman's proper place, the first colleges established in the United States accepted only men.3 Assertions about differences between the sexes have been repeatedly advanced to rationalize the denial of educational opportunities for women.4 Although most of these asserted biological "differences" have been proven invalid and discredited, men and women still do not stand on equal ground with respect to higher education.

3. See United States v. Virginia, 518 U.S. 515, 536-37 (1996) (citing EDWARD H. CLARKE, SEX IN EDUCATION 38-39, 62-63 (1873)). Doctor Clarke "maintained that the physiological effects of hard study and academic competition with boys would interfere with the development of girls' reproductive organs." Id. at 536 n.9.
4. See Brief Amici Curiae in Support of Petitioner at 3, United States v. Virginia, 518 U.S. 515 (1996) (No. 94-241). For example, the "craniology" movement of the nineteenth century sought to prove that intelligence was a function of brain size in order to establish male intellectual dominance over females and to rationally deny educational opportunities to females. See id. at 2.
One inequality that lingers between the sexes involves the scores that males and females receive on the standardized tests that are required for admission into colleges and universities. Males consistently score significantly higher than females on the SAT. This disparity in scores may create unequal opportunities for men and women when applying for admission to college and obtaining scholarships. However, because taking standardized tests is such a well-established ritual in the lives of college-bound teenagers, the idea of standardized tests remains largely unchallenged.

This Note examines whether the use of standardized tests for higher education may be successfully challenged under Title IX of the Educational Amendments of 1972. Title IX was enacted in 1972 to ensure that females receive the same educational opportunities as males. The language of this statute is modeled after Title VI of the Civil Rights Act of 1964, the educational counterpart to the equal employment statute, Title VII of the Civil Rights Act of 1964. Part II of this Note delves into the history of standardized tests as they relate to higher education, focusing primarily on the SAT. This Part also examines the format of the SAT, the pattern of test scores through the years, the effect of preparation for the SAT on scores, and how the SAT became
ingrained in American society. Part III of this Note reviews Title IX and Title VI, briefly explains the legal theories available to challenge practices of an educational institution, and then describes in more detail the disparate impact doctrine and the viability of this doctrine in the educational context. Subsequently, this Part provides an overview of the legal challenges that standardized testing has faced.

Finally, Part IV applies the legal theories of discrimination to a hypothetical case which challenges the mere existence of the SAT and analyzes the fate of such a case. This Part also proposes several nondiscriminatory alternatives to the SAT and reviews the remedies available under Title IX and Title VI.

II. THE HISTORY OF THE SAT

Every year nearly one-and-three-quarter million high school students take America’s oldest and most widely used college entrance exam, the SAT.12 These students spend a great deal of time and energy taking such standardized tests, worrying about how the tests will impact their future, and anxiously awaiting the arrival of the envelope that contains their scores. Once the envelope arrives, its contents may dictate a student’s mood for the following hours and even days. Negative results on standardized tests may even affect a student’s self-image and cause her to lose confidence in her abilities when taking subsequent high-pressure tests.13

The institution that writes and administers the SAT, as well as a multitude of other standardized tests, is the Educational Testing Service (“ETS”). According to critics, ETS perpetuates the view “that people’s positions in society should be determined by their scores on a series of multiple-choice tests.... [H]uman superiority and inferiority can and should be measured scientifically and rewarded accordingly.... ETS plays a significant role in determining who gets ahead in America and who falls, or stays, behind.”14

In the late nineteenth century, colleges administered their own entrance exams. However, the lack of uniform standards for college admissions, combined with the absence of cooperation among colleges,
led to the establishment of the College Entrance Examination Board ("College Board") in late 1899.15

The College Board began to experiment with comprehensive examinations, called "intelligence tests," in order to ascertain whether candidates for college possessed the essential intellectual qualities: alertness, power, and endurance.16 World War I witnessed a tremendous expansion in the field of intelligence testing, with millions of young draftees serving as psychologists' lab rats.17

The American Psychological Association and the National Research Council, with the approval of military authorities, administered the Army Alpha and Army Beta tests to draftees, which ostensibly sorted soldiers according to their abilities and potential.18 These intelligence quotient ("IQ") tests for soldiers enabled the military commanders to distinguish readily (albeit superficially) "between morons on the one hand and bright 'officer material' on the other[,]"19 and also helped the IQ test movement to build statistical evidence.

By 1926, the Army test had metamorphosed into the SAT,20 and had begun to be administered to college applicants only to validate the test, not to decide who was to be admitted into college.21 Validity is the measurement of a test's ability to predict a future outcome.22 The validity of the SAT was established by correlating the scores received on the test with the takers' freshman grades.23 Validity was measured on a zero to one scale, with a validity of .00 meaning that there was no relationship between test scores and first-semester grades, and a validity of 1.0 indicating a complete congruence between test scores and first-semester grades.24

15. See generally CLAUDE M. FUESS, THE COLLEGE BOARD: ITS FIRST FIFTY YEARS (1950) (discussing the development of the College Entrance Examination Board ("College Board").
16. See id. at 101.
17. See id. at 102.
18. See id.
19. Id.
20. See LEMANN, supra note 2, at 32. The official date of the introduction of the SAT into American life is June 23, 1926. See id. On that day, 8040 high school students, primarily from the northeast, took the SAT and had their scores sent to the colleges where they planned to apply. See id. However, the SAT was not required to be taken by college applicants until the 1968-69 school year. See id. at 173.
21. See id. at 173.
22. See id. at 32.
23. See id. at 32-33.
24. See id. at 32 (discussing "reliability," another criterion by which test makers measure a test, and showing that reliability is demonstrated by the consistency of a person's score over repeated administrations of the test). However, the reliability of the SAT is not as significant an
However, the predictive validity of the SAT was not as high as the testing promoters hoped. The validity of the SAT was reported to be in the 0.40 range, but the predictive validity of secondary school grades was higher, and a combination of grades and test scores yielded a higher predictive validity than from either one alone.\textsuperscript{25} Therefore, the predictive value of the SAT on its own was significantly lower than the test makers originally envisioned.\textsuperscript{26}

The idea of mass testing flourished in times of war. World War I was the first instance of large-scale testing,\textsuperscript{27} and World War II allowed for another round of even more widespread testing. Because wartime situations required people to be quickly routed into army positions, objective tests were an obvious tool for accomplishing that task.\textsuperscript{28} During World War II, every Army inductee, totaling ten million before the end of the war, took an IQ test called the Army General Classification Test.\textsuperscript{29}

After World War II, the notion of mass testing remained popular, not for routing soldiers into army positions, but for defining who was fit for higher education. Soon after the War, the GI Bill was passed,\textsuperscript{30} providing veterans with many benefits, including money for college education.\textsuperscript{31} As a result, colleges became affordable for more citizens. The affordability of higher education pushed colleges to be more selective in choosing a student body. The SAT proved a useful tool in distinguishing among the many applicants.

\section*{A. How the SAT Became Ingrained in American Society}

With the passage of the GI Bill, influential test proponents worried about the societal ramifications of allowing anyone to attend any university. These proponents believed in Jefferson's vision of a "natural aristocracy," based on the qualities of "virtue and talents."\textsuperscript{32} The Great Depression damaged people's confidence in the country's leadership,
and President Roosevelt, with his innovative plans for the country, incited a reformist spirit across the nation.  

This spirit was the ideal segue for test proponents to introduce mass testing into peacetime America. Thus, although everyone had the opportunity to obtain free schooling, people underwent a strict selection process for college, a selection process designed to produce the country's new leaders. In other words, the SAT made possible the creation of a "natural aristocracy."  

On January 1, 1948, ETS opened for business. At first, colleges lacked enthusiasm for the SAT because they were not comfortable with the objective testing format; the colleges were accustomed to the essay examinations they had been administering. However, by 1961, the number of students nationwide taking the SAT had increased to 802,500.  

ETS was able to increase this number by taking advantage of the political movement of equality for minorities, emphasizing the objective nature of the SAT and the measurement of mental aptitude based on ability, not on race or color. Efforts to increase opportunities for minorities grew in the late 1960s; one marked change was the passage of the Civil Rights Act of 1964. This Act mandated, among other requirements, that the United States Commissioner of Education survey educational opportunity nationwide. ETS exploited this situation to advertise its tests, claiming the tests' purpose was to create equal opportunity according to merit. The stated objective of ETS aligned well with commonly held views about intelligence testing for occupational selection: that no person should be employed in work either above or below his or her ability.

---

33. See id.  
34. See id. One of the most influential test proponents, James Bryant Conant, the President of Harvard University in the 1930s and 1940s, wholeheartedly believed in the creation of a natural aristocracy and had a clear idea of how to create one. See id. His plan was to establish a strict selection scheme once people completed elementary and high school, and to lower college population by "weed[ing] out 'perhaps one-half' of the people in college." Id. (quoting James Bryant Conant, president of Harvard University).  
35. Id.  
36. See id. at 65.  
38. See id. at 35.  
40. See CROUSE & TRUSHEIM, supra note 37, at 35.  
41. See id.  
42. See id. at 24.
Lewis Terman, developer of the Stanford-Binet Intelligence Test and perhaps the most influential testing psychologist of the period, stated this premise:

Preliminary investigation indicates that an IQ below 70 rarely permits anything better than unskilled labor; that the range from 70 to 80 is preeminently that of semi-skilled labor, from 80 to 100 that of skilled or ordinary clerical labor, from 100 to 110 or 115 that of semi-professional pursuits; and that above all these are the grades of intelligence which permit one to enter the professions or the larger fields of business. Intelligence tests can tell us whether a child's native brightness corresponds more accurately to the median of (1) the professional classes, (2) those in semi-professional pursuits, (3) ordinary skilled workers, (4) semi-skilled or (5) unskilled laborers.  

ETS has repeatedly drawn on Terman's ideas to justify the SAT. Stated Terman's way, standardized testing appears fair because objective, multiple-choice tests measure whether a person will succeed in college without being biased by the quality of the individual's elementary and high school education. Thus, the widespread introduction of the SAT into American education was readily accepted because the logic behind the test reinforced widely-held beliefs. Consequently, beginning in the 1968-69 school year, all university applicants were required to take the SAT.

This requirement greatly expanded ETS's clientele. ETS is "responsible for maintaining the cult of mental measurement." ETS has become the largest testing company in the world. It publishes not only the SAT, but also the Graduate Record Examinations ("GRE"), the Graduate Management Admission Test ("GMAT"), the National Teacher Examinations ("NTE"), parts of the Law School Admission Test ("LSAT"), and other certification and licensing exams in the United States and abroad. ETS is not a government agency; rather, "ETS is a private, autonomous, tax-exempt corporation whose revenues in fiscal 1983 exceeded $130 million."
B. The SAT: An Evolving or Stagnant Concept?

The name of the SAT has changed several times since its inception. However, the format of the SAT has not been altered. Changes have been implemented, but these modifications are merely superficial.

The SAT, in terms of its name, is an evolving concept. When the test was first introduced, SAT was an acronym for "Scholastic Aptitude Test." However, in 1963, one of its creators explicitly posited it to be an intelligence rather than an aptitude test: "'Intelligence tests and scholastic aptitude tests[]' ... 'have the same purpose: to estimate the capacity of the student for school learning.... For all practical purposes, and in all of their school uses, they are the same kind of test.'" This confident declaration has since become an embarrassment to ETS because the idea that a person's mental capacity is inherent and unchanging has fallen out of vogue. Studies have shown that IQ scores can be changed by training, nutrition, or simply by having friendlier people administer the test.

The word "aptitude" in reference to the SAT has been similarly disregarded. For example, until 1982, the GRE was called the GRE Aptitude Test; it is now called the GRE General Test. Aptitude tests became known as "ability" tests instead, which is merely a superficial change, as ability and aptitude are synonymous in the thesaurus. Subsequently, the full name of the SAT was changed to the "Scholastic Assessment Test." Presently, "SAT" does not stand for anything due to the "uneasiness [of] ETS and the College Board about defining just what the test measures."

The content of the SAT, however, has been more stable than the name. Since its inception in 1926, the test has included analogies, sentence completions, reading comprehension, standard math, and quantitative comparisons. The SAT does not test advanced

50. Id. at 200 (second alteration in original) (quoting Henry Chauncey).
51. See FairTest, What's Wrong with Standardized Tests?, at http://www.fairtest.org/facts/whatwron.htm (last visited Nov. 14, 2000). In reality, intelligence quotient ("IQ") tests are nothing more than a type of achievement test, which primarily measures knowledge of standard English and exposure to the cultural experiences of middle-class whites. See id.
52. See id.
53. See Owen, supra note 14, at 200.
56. See id.
mathematics topics and it does not attempt to assess higher-order thinking or reasoning skills. 57

Similarly, early versions of the SAT included some mathematical equations, but were mostly comprised of word familiarity. 55 Although the original versions of the SAT were written with a more elitist, boarding-school type vocabulary, the basic format of the original questions bears a striking resemblance to the test millions of students take today. 59 In addition, the original test and today's SAT share the intimidating qualities of being simple and confusing simultaneously, of requiring guessing and second-guessing on the part of the test taker, and of being limited in time. 60

Here are a few samples from the original version of the SAT:

Pick out the antonyms from among these four words:

Obdurate  spurious  ductile  recondite

Say which word, or both or neither, has the same meaning as the first word:

Impregnable  terile  vacuous

Nominal  exorbitant  didactic

Find the wrong word and change it to the right word in the following passage:

In the citron wing of the pale butterfly, with its dainty spots of orange, he sees beyond him the stately halls of fair gold, with their slender saffron pillars, and is taught how the delicate drawing high upon the walls shall be traced in tender tones of orpiment, and repeated by the base in notes of graver hue. 61

In 1990, substantial revisions to the SAT, the Preliminary SAT ("PSAT"), and the National Merit Scholarship Qualifying Test ("NMSQT") were proposed. 62 These changes would have placed more

58. See Lemann, supra note 2, at 31.
59. See id.
60. See id.
61. Id. (Looking at these examples makes the Author wonder how she was ever accepted into college!).
emphasis on reading skills, decreased the number of multiple-choice questions, and permitted calculators to be used. Such changes also would have included an additional twenty minutes on the verbal section and an expanded critical reading portion to replace some antonym questions.

However, the College Board decided against the more sweeping changes, such as mandatory essays, because of the additional costs to administer and score the essays. The suggestion to allow calculators received severe criticism by some minority groups, who claimed that the burden to buy and learn to use the calculators would put them at an additional disadvantage. Another criticism of the changes was that the changes were merely "cosmetic," because the tests would continue to be used for the same reason—to predict college performance; this prediction itself results in bias against women and minorities.

In 1994, in response to a nationwide decrease in scores, the College Board decided to recenter the scores and grade the tests on a new curve in order "to reflect the changing student population." Until 1994, the average scores of the student population had been calculated based on the scores received in 1941. The College Board recalibrated the scores so that the average scores for today's students would be based on the students' performance in 1990, as opposed to the 1941 test scores.

Simply put, a high-scaled score in 1994 reflects a lower score than the same score in 1941. In other words, students in 1994 were unable to achieve scores as high as the students from 1941. In order to compensate for this decrease, the scores were recentered. Thus, for example, achieving a score of 800 (the highest score on one section of the SAT) today is equivalent to receiving a 700 in 1941.

This change, however, only eliminates the score deficit on paper; nothing else has changed. Rather than identifying the source of the problem which resulted in the score decrease, such as a deterioration of secondary schools, the scores were merely redefined to maintain the appearance of the highest scores.

63. See id.
64. See id.
65. See id.
66. See id.
67. See id.
69. See id.
70. See id.
71. See id.
72. See id.
The final idea that ETS proposed to implement, after implicitly agreeing with critics of the SAT that the SAT is unfair, is a "strivers index." This index is designed to identify those students in minority groups who outperform the expectations of their backgrounds. "[T]hose who score [two hundred] points higher than expected will be labeled 'strivers'—and presumably college admissions officers will note their better-than-expected performance and admit them when they" would not have simply based on their scores.

The goal of the "strivers index" is to uncover a student's true ability, regardless of the test taker's education. However, this proposed solution neglects to take into account the possibility that minorities and women may not receive lower scores than men based on their prior education. Rather, they may receive lower scores because the tests are not written objectively, and the questions reflect the writers' subjective and cultural experiences.

By admitting that there are different score expectations for different minority groups, ETS also implicitly admitted that the score disparity leads to differential treatment of these groups in college admissions. Creating the "strivers index" to remedy the effects on minority (and gender) groups who, as a whole, receive lower scores than other groups, is an insignificant, superficial reform. In addition to being insufficient, it exemplifies ETS's steadfast unwillingness to address and mend the root of the problem: the SAT itself.

73. See Michael Kramer, Editorial, The SAT Is (A) Racist (B) Dub (C) Reforming (D) Hopeless: The Controversial College Test Is Trying to Even the Playing Field by Rewarding 'Strivers,' DAILY NEWS (N.Y.), Sept. 5, 1999, at 53.
74. See id.
75. Id.

In labeling strivers, the testing service will consider 14 factors. Among them are these:

Family: Low socioeconomic status of the student's family, as measured by the parents' education and total family income. So while race itself can be another factor (that will be left to the colleges to evaluate), well-off minorities will not be eligible for striver status simply because they are minorities.

Language: Since the SAT is a "speeded" exam, meaning it must be completed in a given time period, students for whom English is a second language can be given a break.

Academics: If the student's school is deemed inferior, as measured by such factors as low percentage of previous-year graduates entering a four-year college and few if any rigorous academic courses, that too will be part of the striver evaluation.

School location: A public school in a depressed, inner-city neighborhood or an economically disadvantaged region of the country would lower SAT expectations and could lead to a student being deemed a striver.

Id.

76. See id.
C. SAT Scores Through the Years

The SAT is composed of two sections, Verbal and Math. Each is scored on a 200-800 point scale, and the questions are almost solely multiple-choice. The SAT's main competing examination, the American College Testing Program Assessment ("ACT"), is administered to approximately half the number of students than those who take the SAT, and is primarily taken by students in the Midwest. The ACT consists of four sections: English, Mathematics, Reading, and Scientific Reasoning.

There is a consistent and significant gap in scores between male and female test takers on the SAT. Presently, the gender gap is forty-two points, which is the largest gap since 1995. In 1997 there was a forty point difference; the average score for men was 1037, and for women it was 997. The gender gap in ACT scores is much narrower. For example, in 1994, women's ACT scores averaged only 0.2 points lower than men's scores.

Psychologists have speculated about the reasons for this gap, and have listed several contributing factors. Such factors include biased test questions, the multiple-choice format, the guessing penalty (which the

78. See id. Recently, ten "student-produced response" questions were ... added, which require the student to 'grid in' the answer (like filling in the name and address section); the remaining 128 questions are multiple-choice. Id.
79. See id.
80. See id.
81. The gender gap in scores on the SAT will be focused on primarily because the gap between males and females on the American College Testing Program Assessment ("ACT") is significantly narrower. See id.
85. See id. This Note will not focus on the substantive material of the tests, but rather will analyze the SAT based on the disparity of scores between males and females and between white and minority test takers. However, the speculated reasons for the disparity based on psychological study is worth mentioning briefly.
ACT does not have and which may be one reason for such a narrow gap in the scores of men and women on that test), and the "speeded" nature of the test, which provides an average of fifty-one seconds for each question (even for those with lengthy reading passages). Another factor may be the test makers' excuse that the gender gap is caused by the fact that more females take the test than males; that a larger group of women includes more low-scoring students, which negatively skews the average score for females.

The format of the SAT has not changed throughout the years. However, the format of the PSAT was changed to include a "Writing Skills" section in response to a gender bias civil rights complaint in 1994. The following administration of the test resulted in a forty percent reduction of the gender gap, and in the second year with this change, the gender gap shrank by another twenty-six percent. PSAT scores are the sole criterion for the National Merit Scholarship, which provides college tuition aid awards for eligible students. The alteration in the PSAT, simply the addition of a writing section, narrowed the gender gap and directly affected the number of female students eligible for the scholarship.

Because the nature of the PSAT and the SAT are identical, the failure to similarly revise the format of the SAT makes the SAT gender gap even more apparent. Although the effect of SAT scores may not be analogized to the direct effect of PSAT scores on female college or
scholarship applicants respectively, the disparity in scores deprives otherwise eligible women from admission to universities of their choice.

D. If the SAT Measures Aptitude, Does Preparation Help?

The idea of mental tests and the SAT was that they measure the physical property of the brain, akin to the taking and testing of a blood sample. Thus, by definition, preparation for the SAT could not affect test results. Nevertheless, in the 1950s, Stanley H. Kaplan built a successful business tutoring high school students for the SAT. "The word on the street was that Brooklyn kids tended to do a whole lot better on the SAT if they had studied with Stanley Kaplan than if they did not. So they did." When ETS learned of Kaplan's business, it categorized test preparation as organized cheating and attempted unsuccessfully to put Kaplan out of business via legislators in Albany and prosecutors in New York City. In 1981, Kaplan's commercial coaching classes cost $600. In that year, Stanley H. Kaplan Educational Centers grossed over $22 million.

Today, Stanley H. Kaplan Educational Centers are nationwide and boast average score improvements of one hundred points. Thousands of hopeful students enroll in private tutoring, Kaplan's courses, and in the courses of Kaplan's main competitor, The Princeton Review. A basic SAT course today costs almost $900, and private tutoring costs about $2000 for approximately twenty hours.

The Princeton Review even guarantees a score improvement of at least one hundred points. These courses may improve scores, but have only a minimal effect, if any, on the gender gap, because only those able

93. See LEMANN, supra note 2, at 112.
94. See id.
95. See id.
96. Id.
97. See id. at 114.
98. See CROUSE & TRUSHEIM, supra note 37, at 69.
99. See id.
101. See id.
to afford the courses have this opportunity to improve.\textsuperscript{104} The availability of expensive preparatory courses may put pressure on schools in areas where most people are not able to afford such courses to prepare the students for the SAT during classroom time. However, doing so could have a damaging effect on the quality of education because less time would be spent on general education.

III. \textbf{TITLE IX, TITLE VI, AND LEGAL THEORIES TO REMEDY DISCRIMINATION}

Title IX of the Educational Amendments of 1972 prohibits sexual discrimination in education.\textsuperscript{105} It provides that "[n]o 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."\textsuperscript{106} Congress patterned Title IX after Title VI of the Civil Rights Act of 1964,\textsuperscript{107} and courts strive to construe the statutes so they will coincide with each other in scope and effect.\textsuperscript{108}

Neither Title IX nor Title VI explicitly address standardized testing and test use.\textsuperscript{109} However, regulations implementing Title IX prohibit the administration and application of tests which have a disproportionate and adverse effect on people on the basis of sex. For example, educational programs or activities which fall under Title IX (those that receive federal financial assistance) may:

\begin{itemize}
  \item \textsuperscript{104} The commercial coaching courses arguably have an effect on the disparity of scores on an income-based scale. Such a disparity does exist. For example, in 1997, students in a family whose income ranged from $10,000 to $20,000 a year averaged 166 points lower than a student whose family income ranged from $80,000 to $100,000 a year. See FairTest, 1997 SAT and ACT Scores, at http://www.fairtest.org/satscr97.htm (last visited Oct. 3, 1999). In 1998, the same category of incomes differed by 171 points. See Press Release, FairTest, College Board Hides Growing SAT Gender Gap: Users of Biased Test Scores Risk Legal Sanctions (Sept. 1, 1998), at http://www.fairtest.org/pr/satact98.htm (last visited Oct. 3, 1999) (on file with the Hofstra Law Review). Additional details of the economic disparity are beyond the scope of this Note.
  \item \textsuperscript{105} 20 U.S.C. § 1681 (1994). Coincidentally, this statute was passed only four years after the SAT was required to be taken by every college applicant—four years after ETS achieved nationwide success. See \textbf{LEMANN}, supra note 2, at 173.
  \item \textsuperscript{106} 20 U.S.C. § 1681(a).
  \item \textsuperscript{107} 42 U.S.C. § 2000d (1994). This statute is worded identically to Title IX, except that the word "sex" was substituted for "race, color, or national origin." See 20 U.S.C. § 1681; 42 U.S.C. § 2000d. Title VI provides: "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance." 42 U.S.C. § 2000d.
\end{itemize}
not administer or operate 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 predict validly success in the education program or activity in question and alternative tests or criteria which do not have such a disproportionately adverse effect are shown to be unavailable.¹¹⁰

In addition, other regulations, which do not specifically refer to testing, prohibit discrimination in areas where test scores are often applied, such as providing financial assistance or awarding scholarships.¹¹¹ Thus, although these regulations do not explicitly address testing, such areas are encompassed under Title IX jurisprudence.¹¹²

Such regulations mirror the prohibitions on the use of discriminatory tests in employment, which are codified in Title VII of the Civil Rights Act of 1964.¹¹³ Title VII prohibits the use of selection devices that have a disparate impact on a protected class.¹¹⁴ This doctrine was first established by the United States Supreme Court in Griggs v. Duke Power Co.¹¹⁵ Disparate impact discrimination refers to facially neutral practices which do not intentionally discriminate, but nonetheless have a discriminatory effect.¹¹⁶ This Part explains disparate impact and its application to the educational context.

A. The Disparate Impact Theory: Development and Application

The United States Supreme Court, in Griggs v. Duke Power Co.,¹¹⁷ established the standard for a disparate impact discrimination claim under Title VII¹¹⁸ and was codified as part of the Civil Rights Act of 1991.¹¹⁹ At issue in Griggs was whether an employer is prohibited “from requiring a high school [diploma] or passing of a standardized general intelligence test as a condition of employment.”¹²⁰ Neither requirement

¹¹¹. See id. § 106.37(a); see also supra notes 90-91 and accompanying text (discussing the use of the PSAT as the sole criterion for the National Merit Scholarship).
¹¹⁴. See id. § 2000e-2(a).
¹¹⁶. See Connor & Vargyas, supra note 112, at 42.
¹¹⁷. 401 U.S. 424 (1971). This case was decided in 1971, a year before Title IX was passed. See id.
was shown to be significantly related to job performance and both disqualified black applicants at a significantly higher rate than white applicants.\textsuperscript{121}

The Supreme Court held that the employer’s requirements violated Title VII, and that an absence of discriminatory intent does not preclude a cause of action for discrimination if the employment practice is shown to have discriminatory results.\textsuperscript{122} The employer has the burden of showing that any requirement must be related to job performance.\textsuperscript{123} The Court stated, “[w]hat Congress has commanded is that any tests used must measure the person for the job and not the person in the abstract.”\textsuperscript{124}

\textit{Griggs’} progeny established that for a disparate impact discrimination claim, the plaintiff must first establish “that a particular employment practice has caused a significant adverse effect on a protected group.”\textsuperscript{125} In order to show such an effect, there are specific guidelines, called the Uniform Guidelines on Employee Selection Procedures (“Uniform Guidelines”),\textsuperscript{126} which were created by the Equal Employment Opportunity Commission, the agency that enforces Title VII.

The Uniform Guidelines define “adverse impact” as “[a] substantially different rate of selection... which works to the disadvantage of members of a race, sex, or ethnic group.”\textsuperscript{127} The group ultimately selected and the group who was qualified to be selected should be compared to determine if there has been a disparate impact.\textsuperscript{128} However, there are no similar guidelines to ascertain an adverse effect in education cases.\textsuperscript{129}

\begin{footnotes}
\item[121.] See id. at 426. The question regarding the SAT is whether SAT scores are related to performance in the first year of college.
\item[122.] See id. at 432-33.
\item[123.] See id. at 432.
\item[124.] Id. at 436.
\item[125.] United States v. City of Warren, 138 F.3d 1033, 1091 (6th Cir. 1998).
\item[126.] 29 C.F.R. pt. 1607 (1998). The Uniform Guidelines on Employee Selection Procedures (“Uniform Guidelines”) are useful for advancing the basic purposes of Title VII, and should always be considered by the court, but they are not regarded as conclusive unless the facts of the particular case support that conclusion. See Guardians Ass’n of N.Y. City Police Dep’t, Inc. v. Civil Serv. Comm’n, 630 F.2d 79, 90-91 (2d Cir. 1980).
\item[127.] 29 C.F.R. § 1607.16(B).
\item[128.] See Connor & Vargyas, supra note 112, at 50.
\item[129.] See id. at 51. In application, cases in the educational context require the adverse impact of the test to be statistically and readily apparent. See id.; see also Sharif v. N.Y. State Educ. Dep’t, 709 F. Supp. 345, 362 (S.D.N.Y. 1989) (“Plaintiffs have met their burden of establishing a \textit{prima facie} case through persuasive statistical evidence and credible expert testimony that the composition of scholarship winners tilted decidedly toward males and could not have occurred by a random distribution.”).
\end{footnotes}
Once the plaintiff has shown a significant adverse effect, the burden shifts to the employer to produce evidence that the practice in question is a business necessity.\textsuperscript{130} Necessity has been interpreted in education and employment testing analyses to be based on the concept of validity. To be valid, the test must measure what it purports to measure, and the test must address the skills necessary for a successful performance in the job.\textsuperscript{131}

Another requirement, according to the Uniform Guidelines, is that the test itself be fair. A test may unfairly deny opportunities to a particular group "[w]hen members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance."\textsuperscript{132}

\textbf{B. The Disparate Impact Doctrine in the Educational Arena}

Neither Title IX nor Title VI specifically addresses educational practices which are facially neutral but discriminatory in effect. It is unclear whether the disparate impact theory is viable under either statute. One landmark case which arose under Title VI attempted to answer this question.

In \textit{Guardians Ass'n v. Civil Service Commission},\textsuperscript{133} the Supreme Court analyzed a challenge to an examination for would-be police officers.\textsuperscript{134} Entry-level appointments were made in order of test scores, which caused minorities to be hired later than otherwise similarly situated whites and lessened the minorities' seniority and benefits.\textsuperscript{135} In addition, the police department laid off police officers on a "last-hired,

\begin{thebibliography}{135}
\bibitem{130} "Educational necessity" and "business necessity" are considered analogous, and the Title VII formulations for business necessity have been applied to Title IX cases. \textit{See Sharif}, 709 F. Supp. at 361-62.

\[ T \]o prevail, defendants must show a manifest relationship between use of the SAT and recognition and award of academic achievement in high school. The court finds that defendants have failed to show even a reasonable relationship between their practice and their conceded purpose. The SAT was not designed to measure achievement in high school and was never validated for that purpose. \textit{Id.} at 362.

\bibitem{131} \textit{See} Connor & Vargyas, \textit{supra} note 112, at 52.

\bibitem{132} 29 C.F.R. § 1607.14(B)(8)(a). This requirement, if applicable in the educational context, could have a dispositive effect on a claim against the SAT because of the clear disparity in scores between male and female test takers. \textit{See discussion supra} Part II.C (discussing the consistent disparity in SAT scores between men and women).

\bibitem{133} 463 U.S. 582 (1983).

\bibitem{134} \textit{See id.} at 585.

\bibitem{135} \textit{See id.}. 

http://scholarlycommons.law.hofstra.edu/hlr/vol29/iss2/8
first-fired” basis, so that the officers with the lowest examination scores, the minorities, were laid off first, causing further racial disparity among the employees.\textsuperscript{15}

However, a deeply divided plurality of the Court did not clearly decide whether the disparate impact was remediable under Title VI. Only two Justices, Justice White and Justice Marshall, agreed that Title VI itself proscribes unintentional, disparate impact discrimination.\textsuperscript{137} Justices Stevens, Brennan, and Blackmun agreed that although Title VI expressly requires proof of discriminatory intent, the administrative regulations promulgating the statute that incorporate a disparate impact standard are valid.\textsuperscript{138}

In sum, the plurality of the Court found that a violation of Title VI itself requires proof of discriminatory intent. In order to avoid the requirement of proving such intent, a plaintiff must allege that the defendant has violated the Title VI implementing regulations, which expressly prohibit practices that have a discriminatory effect.\textsuperscript{139}

The Guardians decision provides a muddy picture for plaintiffs who wish to challenge an educational practice under Title VI, and an even muddier one for those challenging a practice under Title IX, because the decision did not directly apply to Title IX claims. The first court to address disparate impact discrimination allegations under Title IX utilized a Title VI analysis for the decision.\textsuperscript{140} Therefore, at least according to one court, the Guardians decision applies to both Title VI and Title IX claims.

Subsequent to Guardians, courts have interpreted Title VI and Title IX to require a showing of invidious intent in order to establish a prima facie violation of the statutes.\textsuperscript{141} This requirement may be circumvented

\textsuperscript{136} See id.

\textsuperscript{137} See id. at 593 (White, J., plurality opinion); id. at 623 (Marshall, J., dissenting) (“Title VI bars practices that have a discriminatory impact and cannot be justified on legitimate grounds.”).

\textsuperscript{138} See id. at 642-43 (Stevens, J., dissenting). Justices Stevens, Brennan, and Blackmun recognized that when Congress explicitly authorizes an administrative agency to create regulations in order to enforce a federal statute, those regulations are valid and “have the force of law so long as they are ‘reasonably related to the purposes of the enabling legislation.’” Id. at 643 (Stevens, J., dissenting) (quoting Mourning v. Family Publ’ns Serv., Inc., 411 U.S. 356, 369 (1973)). Here, these Justices held that the “effects” standard in the regulations is a reasonable method for the agency to enforce Congress’ prohibition against discrimination. See id. at 644-45 (Stevens, J., dissenting).

\textsuperscript{139} See id. at 608 n.1 (Powell, J., concurring).

\textsuperscript{140} See Sharif v. N.Y. State Educ. Dep’t, 709 F. Supp. 345, 360-61 (S.D.N.Y. 1989) (“This court finds no persuasive reason not to apply Title VI’s substantive standards to the present Title IX suit.”).

\textsuperscript{141} See, e.g., Larry P. v. Riles, 793 F.2d 969, 981 (9th Cir. 1984) (“[V]iolation of Title VI requires[s] proof of discriminatory intent.”); Sharif, 709 F. Supp. at 360 (“[V]iolation of Title VI itself requires proof of discriminatory intent.”).
if plaintiffs allege a violation of regulations enforcing the statutes which explicitly incorporate a discriminatory effect. In such situations, proof of disparate impact will generally suffice as prima facie evidence of discrimination.\(^{142}\)

If a suit is brought under Title IX or Title VI regulations containing an "effects" standard, the disparate impact claim is analyzed as it would be under Title VII. Once a prima facie case is established based on disparate impact, the burden switches to the defendant to demonstrate educational necessity for the practice.\(^{143}\) If the defendant demonstrates necessity, the plaintiff may still prevail by presenting a comparable, less discriminatory alternative, or by providing "proof that the legitimate practices are a pretext for discrimination."\(^{144}\)

C. Legal Challenges to Testing Under Title IX

The first case to apply the disparate impact theory of discrimination under Title IX was *Sharif v. New York State Education Department*.\(^{145}\) Female applicants alleged that distributing college scholarships based solely on SAT scores violated the gender discrimination prohibitions of Title IX.\(^{146}\) The purpose of the scholarships was to recognize and reward past academic achievement.\(^{147}\) The plaintiffs argued that the SAT was not designed to measure academic performance, but even if it was, the SAT nevertheless discriminates against female applicants for scholarships, because it underpredicts performance for females.\(^{148}\)

First, because this was a case of first impression on the issue of whether the disparate impact theory could be applied to Title IX cases, the court analogized it to Title VI.\(^{149}\) The court proceeded on a discriminatory-effect analysis because the plaintiffs sued under the Title IX regulations, not just under the statute itself.\(^{150}\) Several of these

\(^{142}\) See, e.g., *Sharif*, 709 F. Supp. at 360 ("[P]roof of discriminatory effect suffices to establish liability when a suit is brought to enforce the regulations promulgated under Title VI, rather than [the] statute itself.").

\(^{143}\) See id. at 361.

\(^{144}\) Id.


\(^{146}\) See id. at 348. Plaintiffs also alleged that the discrimination violated the regulations promulgated under Title IX and the Equal Protection Clause of the Fourteenth Amendment to the United States Constitution. See id.

\(^{147}\) See id.

\(^{148}\) See id.

\(^{149}\) See id. at 360.

\(^{150}\) See id.
regulations, including the one relevant to testing, specifically prohibit facially neutral policies.\textsuperscript{151}

The court found that the plaintiffs established a prima facie showing of discriminatory effect because males consistently received substantially more scholarships than females, and eligibility was based solely on SAT scores.\textsuperscript{152} In 1987, males made up only forty-seven percent of the scholarship competitors, but received seventy-two percent of the Empire State Scholarships and fifty-seven percent of the Regents Scholarships.\textsuperscript{153} The probability that these results would occur by chance was less than one in a billion.\textsuperscript{154}

After the plaintiffs established that the facially neutral practice had a disproportionate effect on women, the burden shifted to the defendants to prove there is a \textit{manifest} relationship between the use of the SAT and recognition of high school achievement (that the use of the SAT is an educational necessity).\textsuperscript{155} The court found that defendants failed to show even a \textit{reasonable} relationship between their practice and their conceded purpose because the SAT does not even purport to measure high school achievement, but rather predicts success in college.\textsuperscript{156} Subsequently, the plaintiffs offered an alternative to sole reliance on the SAT scores: a combination of grade point averages and SAT scores.\textsuperscript{157} Based on the foregoing analysis, the court held that the plaintiffs were likely to succeed on the merits of the Title IX regulations claim, and thus enjoined the state from awarding college scholarships based on SAT scores alone.\textsuperscript{158}

Standardized tests came under direct attack in 1994 when the National Center for Fair & Open Testing ("FairTest") filed a complaint

\textsuperscript{151} See id. at 361. For example, the provision governing admissions procedures prohibits a recipient from:

\begin{quote}
"[A]dminister[ing] or operat[ing] any test or other criteria for admission which has a \textit{disproportionately adverse effect} on persons on the basis of sex unless the use of such test or criterion is shown to predict validly success in the education program or activity in question and alternative tests or criteria which do not have such a disproportionate adverse effect are shown to be unavailable."
\end{quote}

\begin{flushright}
\textit{Id.} (quoting 34 C.F.R. § 106.21(b)(2) (1975)) (emphasis added) (second and third alterations in original).
\end{flushright}

\textsuperscript{152} See id.

\textsuperscript{153} See id. at 355.

\textsuperscript{154} See id.

\textsuperscript{155} See id. at 362.

\textsuperscript{156} See id. "[T]here can be no serious claim that a test given on one single morning can take into account a student's diligence, creativity and social development and work habits in that student's environment—all part of high school achievement." \textit{Id.}

\textsuperscript{157} See id. at 362-64.

\textsuperscript{158} See id. at 364.
with the Department of Education’s Office of Civil Rights. FairTest charged ETS and the College Board with illegally discriminating against females. The charge was based on their role in designing and administering the PSAT/NMSQT. However, the claim was never tried in court because the complaint was settled; the test makers agreed to add a multiple-choice “writing” section to the exam in 1999. As a result of this change, scores in 2000 showed a forty percent reduction in the gender gap.

This significant result leads to several questions: “Why . . . have similar changes not been made on the SAT, the GRE and related exams which show comparable bias? Even more fundamentally, why are instruments on which results can be so quickly “adjusted” ever used as the sole or primary factor to determine college admissions or award scholarships?”

IV. A HYPOTHETICAL CASE: CHALLENGING THE EXISTENCE OF THE SAT

To date there have been no cases which challenge the mere existence of the SAT, and no suits against a university for using the SAT as a decisive factor in its admissions decisions. The hypothetical case in this Part alleges that ETS has violated the regulations promulgating Title IX as well as Title IX itself because of its involvement in creating and administering the SAT and other standardized tests.

The requirement that the discrimination occur under “any education program or activity receiving [federal financial assistance]” could create an obstacle to this hypothetical suit. ETS may not be considered an


160. See id. Scores from this three-hour, multiple-choice exam are the sole factor in determining eligibility for the scholarships. See id. In addition, in the past, women received less than forty percent of the scholarships, while constituting more than fifty-six percent of the exam-taking population, despite earning higher grades than similarly situated men in both high school and college when compared in identical courses. See id.

161. See id.

162. See id.

163. Id. (quoting FairTest Public Education Director, Bob Schaeffer).

164. This hypothetical will not address the viability of a constitutional claim. However, a suit alleging a violation of the Equal Protection Clause of the United States Constitution, U.S. CONST. amend. XIV, § 1, or alleging a violation of a state constitution may be successful.

educational program because it does not specifically receive federal financial assistance. Nevertheless, this obstacle may be overcome by arguing that the SAT is an educational program based on its close relationship to college admissions decisions and because the test is designed specifically to predict success in an educational program.

In addition, although ETS does not receive federal financial assistance per se, for tax purposes "ETS is classified as an 'organization exempt from income tax'" in the Internal Revenue Code.166 This section of the Internal Revenue Code includes schools, churches, and other organizations that function exclusively for educational, religious, and other similar purposes.167 Thus, it can be argued that ETS receives federal assistance in the form of a tax subsidy.

However, if this obstacle is not surpassed, then an alternative to the hypothetical case described above is to bring an action against colleges and universities that use the SAT as a determinative criterion in making admissions decisions. This Part will analyze the initially proposed hypothetical case, although the same examination would apply to this alternate action as well.

First, the plaintiff must have standing to sue. Although neither Title IX nor Title VI expressly authorize private suits, the Supreme Court inferred a private right of action to enforce Title IX in Cannon v. University of Chicago.168 The Court also attempted to answer whether individuals have a private right of action under Title VI in Guardians, but only a divided plurality resulted.169

Second, the plaintiff must sue to enforce the enabling regulations of Title IX because the statute itself does not expressly refer to the disparate impact theory.170 Rather, the statute requires, and courts consistently hold, that discriminatory intent be shown.171 Title IX regulations specifically prohibit the use of any test that has a

166. OWEN, supra note 14, at xx.
167. See id.
169. See Guardians Ass'n v. Civil Serv. Comm'n, 463 U.S. 582 (1983). The Court differed as to the nature of relief available and the basis for awarding relief. The various opinions can be summarized as follows: Victims of intentional discrimination may sue for compensatory and prospective relief, and victims of unintentional discrimination may sue for prospective relief only. See id. at 597, 602-03 (White, J., plurality opinion); see also id. at 625-27 (Marshall, J., dissenting); id. at 638 (Stevens, J., dissenting); Lora Silverman, Note, Unnatural Selection: A Legal Analysis of the Impact of Standardized Test Use on Higher Education Resource Allocation, 23 LOY. L.A. L. REV. 1433, 1448 (1990) (explaining in more detail the plurality decision in Guardians).
170. See discussion supra Part III.C.
171. See discussion supra Part III.C (discussing Sharif).
disproportionate and adverse effect on students on the basis of gender.\textsuperscript{172} Therefore, if the plaintiff sues to enforce the regulations of Title IX, she can then proceed to establish a prima facie case of disparate impact discrimination.

A. Establishing Disparate Impact

The first issue that arises when plaintiffs sue under the effects theory is how much of an impact is needed in order to establish a prima facie case. The regulations for Title VII incorporate clear guidance for plaintiffs suing under the effects theory.\textsuperscript{173} The Uniform Guidelines provide a formula known as the four-fifths rule to determine whether there is an adverse impact, under which a selection rate for any race, sex, or ethnic group which is less than four-fifths of the rate for the group with the highest selection rate will generally be regarded as evidence of adverse impact.\textsuperscript{174} However, the regulations of Title VI and Title IX do not provide such statistical guidance.

In educational cases, the standard is not clear; if the adverse impact of the test is readily apparent and statistically significant, this may suffice as a prima facie case. In \textit{Sharif}, the court did not use a specific mathematical computation of the effect because the disparity was glaring.\textsuperscript{175} Rather, the plaintiffs established an adverse impact "through persuasive statistical evidence and credible expert testimony that the composition of scholarship winners tilted decidedly toward males and could not have occurred by a random distribution."\textsuperscript{176}

Another related issue is determining which groups to use for statistical comparison. For example, if a university is sued for allegedly discriminatory admissions practices, it would urge the court to compare the percentage of women who apply to the university with the percentage of women enrolled. This comparison would not show a disparate impact because there are approximately equal numbers of men and women at most universities. If, however, the percentage of women

\textsuperscript{172} See 34 C.F.R. § 106.21(b)(2) (1998).
\textsuperscript{173} See 29 C.F.R. pt. 1607 (1998); \textit{see also} discussion \textit{supra} Part III.A (explaining what constitutes a sufficient adverse effect to establish a prima facie case under Title VII).
\textsuperscript{174} See 29 C.F.R. § 1607.4(D). However, the Uniform Guidelines do not preclude the use of other statistical analyses; it also considers practical and statistical significance in order to establish the requisite impact. \textit{See id.}
\textsuperscript{175} For an explanation of the proof that the court in \textit{Sharif} relied upon to conclude there was a significant adverse effect, \textit{see discussion} \textit{supra} Part III.C.
who take the SAT is compared with the percentage of women enrolled at the university the results would be significantly different.

The latter comparison would demonstrate how SAT scores affect where people decide to apply to college; that many women otherwise qualified to attend the college may self-select not to apply because they feel their SAT scores disqualify them from admission. However, the adverse effect of self-selection is not actionable under Title IX because the statute refers only to the decisions and practices of institutions receiving federal financial assistance, not to the decisions of individual students.177

Under Title VII regulations, the groups that must be compared are those ultimately selected and those qualified for selection. However, Title IX regulations allow a broader pool for comparison.178 One proper "comparison [is] females who were 'potentially available for training[,]' rather than those who had actually applied to the school or taken a particular test."

This standard of comparison allows for detrimental self-selection to be taken into account because a student with high grades may be "potentially available" for a particular school but will not even apply based on her SAT scores.

In this hypothetical case, the groups to compare would be the female test takers and the male test takers. However, this group may be too broad, in which case the groups to compare are female and male test takers who are similarly situated in terms of grades in high school. The adverse effect would be shown by the resulting scores and the percentage of those admitted into certain universities (those universities chosen for the statistical analyses).

B. Defendant's Rebuttal: Proving "Educational Necessity"

In response to the plaintiff's prima facie case of disparate impact discrimination, the defendant will attempt to demonstrate that the practice of administering and using tests in the admissions process is an "educational necessity." In order to do so, the defendant must show a rational relationship between the practice of creating the SAT and the purpose of using the SAT, namely to admit students who are most likely to succeed in college.

178. See, e.g., 34 C.F.R. § 106.36(c) (1998) (establishing disparate impact when there is a substantially disproportionate number of members of one sex in a particular course of study, classification, or class).
179. Connor & Vargyas, supra note 112, at 51 n.198.
The defendant will present studies which demonstrate the relationship between students' SAT scores and freshman year college grades. ETS runs the College Board's Validity Study Service, which encourages colleges to base admissions decisions on predicted performance and provides equations and computational aids to make such predictions. ETS also recommends that colleges conduct studies of the SAT's predictive effectiveness. These studies often show that including the SAT in the equation improves the accuracy of the prediction of first-year college grades.

However, for women, the SAT does not fulfill its purpose of predicting first-year college grades. Researchers at the College Board and individual universities have discovered that even though men receive higher SAT scores, women generally achieve higher grade-point averages in their first year of college. Thus, a single prediction for both men and women based on SAT scores, even if combined with high school grades, will underpredict women's success in college, and may cause fewer women to be admitted to the college than are qualified to attend. Furthermore, if the purpose in using the SAT for admissions is to select the students most likely to be successful at that college, then educational necessity will not be a successful defense because the prediction for women is flawed and thus is not rationally related to the conceded purpose of predicting first-year college performance.

180. See CROUSE & TRUSHEIM, supra note 37, at 41. Many colleges use these formulae to decide eligibility for admission. See id. For example, a College Board/American Association of Collegiate Registrars and Admissions Officers survey found that eleven percent of the public and three percent of the private four-year colleges use minimum predicted grades from ETS's computations as a cutoff for admissions. See id. Still another thirty-four percent of public and thirty-five percent of private colleges use the predictions as either one factor or guidance in admissions decisions. See id.

181. See id. at 42.

182. See id. at 43. Although these results seem impressive, they are not necessarily reliable and should not be accepted without skepticism because the studies are based on students who have enrolled and completed one year at the institution, not on students who were rejected and may have been as successful. Therefore, using information from the pool of applicants enrolled to make predictions for future applicants of that institution may be substantially misleading. See id. at 45.


184. See Press Release, Females Cheated Again by SAT Bias; SAT Gender Gap Grows While Narrowing on PSAT, ACT; Test-Maker "Accountability" Needed To Stop Illegal Discrimination (Aug. 31, 1999), at http://www.fairtest.org/pr/8-31/SATgap.html (last visited Oct. 3, 1999) (on file with the Hofstra Law Review). "According to a recent academic study published in 'Research in Higher Education,' [June 1999,] SAT underprediction ... 'arguably leads to the exclusion of 12,000 women from large, competitive, "flagship" state universities' each year because admissions offices rely on minimum 'cut-off' scores." Id. (quoting the recent academic study).
C. Providing Nondiscriminatory Alternatives

Even if the defendant successfully proves an educational necessity for the SAT, the plaintiff may still prevail by providing an alternative to the existing SAT reliance that has a less discriminatory impact.

One alternative would be to eliminate the SAT altogether from the admissions process. Four highly selective schools—Bates, Bowdoin, Muhlenberg, and Franklin & Marshall—have deemphasized standardized tests in their admissions processes, by eliminating or making the SAT optional. These colleges have found that the reform promotes equity and excellence and diversifies the applicant pool without any loss in academic quality. "Bowdoin College, a top liberal


- High school performance is the best available method for screening applicants;
- Tests add little useful information to the high school record;
- Moving away from reliance on admissions tests promotes sounder educational practices in high schools by downgrading the value of multiple-choice exam preparation; and
- Other colleges considering admissions reforms can learn from the experiences of the colleges profiled in the report.

Id. In addition, the report includes a step-by-step guide for colleges to reform their admissions procedures, and lists several questions a college should consider:

- Do tests have meaningful predictive validity for significant educational outcomes, such as graduation rates, at that particular institution?
- Does that validity hold for all ethnic, age and income groups, as well as for men and women?
- Do the tests add anything of significance to what admissions officers already know about applicants?
- Are current test score requirements deterring potential applicants who would make successful students, particularly those from underrepresented groups?

Id.

186. See id. Discussing the reformed admissions procedure, authorities at the schools who have implemented the test-optional policy expressed their views and the purpose of this change. The former Director of Admissions at Bowdoin College stated:

The message we should be sending to high schools is that admissions offices at selective colleges are capable of making informed decisions without relying heavily or at all on the Educational Testing Service, not that we want them to design their courses to what can be tested by multiple-choice exams.

Id. (internal quotation marks omitted). A Muhlenberg College Questions and Answers About Muhlenberg’s Test-Optional Policy states: “Our hope is that the decision to move to a test-optional admissions policy will give some of the power back to students in the college admissions process. This decision gives students a larger say in how to present themselves, what constitutes their strongest portfolio of credentials, etc.” Id. (quoting Muhlenberg College’s Questions and Answers About Muhlenberg’s Test-Optional Policy). The Director of Admissions at Franklin & Marshall noted: “If you’re wincing because you see a modest SAT score, then you’re not being fair
arts school in Maine, stopped requiring SAT scores in 1970. 187 Fifteen years later, in 1985, Bowdoin's academic reputation was as high as it used to be and the administrators decided to continue the test-optional policy. 188 Thus, the question that remains is: "If a college like Bowdoin can get by without requiring SATs, how many schools can convincingly argue that they can't?" 189

If eliminating the SAT from the admissions process is too radical to implement, there are several other alternatives which would lessen the adverse impact on women. One option is to create gender-specific prediction formulae, which have been shown to predict college performance more closely than gender-neutral equations. 190 This alternative will enable colleges to more accurately achieve their goal of accurate admissions decisions and increase the number of women admitted into colleges that would not have been accepted with gender-neutral interpretations of test scores.

Another alternative would be to change the format of the SAT to add a writing section, as the PSAT did in 1994. 191 This minor alteration resulted in a forty percent reduction in the gender gap on the PSAT, which is a virtually identical test to the SAT. 192 Such a change, while it would not eliminate the adverse effect of SAT scores on women, it would narrow the gender gap and ultimately lessen the discriminatory impact of standardized testing.

Yet another option is to guarantee the top ten percent of each graduating class from all of a state's high schools a spot at the state's public university, at a campus of their choice. This option became a law in Texas two years ago and has since been imitated by Florida and California. 193 This option does not address the gender issue specifically,

---

187. OWEN, supra note 14, at 241.
188. See id. at 241-42.
189. Id. at 242.
191. For a discussion of the change in the PSAT format, see supra notes 88-94 and accompanying text.
193. See Jodi Wilgoren, New Law in Texas Preserves Racial Mix in State's Colleges, N.Y. TIMES, Nov. 24, 1999, at A1. The Governor of Florida proposed automatic admission for the top twenty percent of graduates in that state; and the California proposal promises slots to each high school's top four percent (although, not to any particular campus). See id.
but it gives all students an equal chance, regardless of SAT scores. In addition, this alternative serves the defendant's goal of selecting students that will be successful in college or the goal of predicting first-year grades because, academically, these students are as successful as students who are admitted based on their SAT scores.

D. Remedies Under Title IX and Title VI

If the plaintiff prevails, there are both governmental and private remedies available to enforce legal prohibitions against gender discrimination. One such governmental remedy is to defund a recipient of federal financial assistance if the recipient fails to correct a violation of Title IX. However, courts are reluctant to apply this remedy. Remedies principally include prohibiting or restricting the use of an invalid test via injunctions, requiring that a test be validated or that a valid selection procedure be developed, or ordering the elimination of the discriminatory effects of an invalid test use. Remedies also may include monetary relief and legal fees.

Title IX is enforced primarily through the Office for Civil Rights ("OCR") in the Department of Education and through private rights of action. Through the OCR, the Department of Education has various options for enforcing Title IX including: conducting compliance reviews and investigating complaints, making findings of non-compliance with the law, conciliating claims, and referring cases to the Department of Justice for judicial enforcement. However, the regulations

194. Another important side effect from this program is that the university system is becoming more involved and devoted to improving the schools throughout the state, from kindergarten through twelfth grade. See id.

195. See id. For example, in the Texas public university system, students accepted by the ten percent rule have an average chemistry grade point average of 2.63, which is slightly below the 2.7 average in the larger classrooms, but much higher than the 2.1 average of the students with comparable standardized test scores. See id.

196. See Connor & Vargyas, supra note 112, at 77.

197. See Storey v. Bd. of Regents of Univ. of Wis. Sys., 604 F. Supp. 1200, 1202 (W.D. Wis. 1985); see also Guardians Ass'n v. Civil Serv. Comm'n, 463 U.S. 582, 601 (1983) ("The remedy of termination of assistance was regarded as 'a last resort, to be used only if all else fails,' because 'cutoffs of Federal funds would defeat important objectives of Federal legislation, without commensurate gains in eliminating racial discrimination or segregation.'") (quoting 110 Cong. Rec. 6544, 6546 (1964) (statement of Sen. Humphrey)).

198. See Connor & Vargyas, supra note 112, at 80.

199. See id.

200. See id. at 77.

201. See id.
promulgating Title IX lack coherent guidelines regarding the enforcement of discriminatory testing practices.

For Title VI cases, the remedies issue has not been resolved. In Guardians, the Supreme Court attempted to clarify the issue but the resulting plurality precluded a clear result.\textsuperscript{202} In Sharif, the first court to address the remedies issue in Title IX cases, the court enjoined the New York State Education Department from using the SAT as the sole criterion for college scholarships and required the defendant to use a combination of SAT scores and grades.\textsuperscript{203} This remedy was straightforward and is directly applicable to the hypothetical case.

The difficulty of Title IX and Title VI cases is that there is no guidance for the agency or the court to enforce the statutes and eliminate the discriminatory effect of the testing practice. This issue arises especially when the plaintiff is suing under her private cause of action rather than through the OCR. Although courts have construed congressional intent to indicate that Title IX and Title VI are coextensive,\textsuperscript{204} court decisions are inconsistent with this interpretation, which may thus lead to an unjust result.

V. CONCLUSION

Our society has placed all students on a scale based on scores from a test that lasts only three and a half hours. This scale is objective, which to many people is equivalent to being unbiased and providing equal opportunity. Although the scale is the same for everyone, regardless of race or gender, the disparity of test scores between males and females and between white and minority test takers reveals that the underlying test is flawed and discriminatory.

\textsuperscript{202} See Guardians, 463 U.S. at 601-05 (White, J., plurality opinion) (stating that victims of \textit{intentional} discrimination may sue for compensatory relief, prospective relief, and retroactive relief, exclusive of back pay and back benefits); \textit{id.} at 639-42 (Stevens, J., dissenting) (noting that victims of \textit{unintentional} discrimination may sue for prospective relief only); \textit{see also} Silverman, \textit{supra} note 169, at 1448 (explaining the varying opinions of the Guardians' plurality).

\textsuperscript{203} See Sharif v. N.Y. State Educ. Dep't, 709 F. Supp. 345, 363-65 (S.D.N.Y. 1989). A similar remedy was applied in an earlier Title VI case. \textit{See} Larry P. v. Riles, 793 F.2d 969, 984 (9th Cir. 1984) (enjoining a non-validated use of IQ tests to place educable mentally disabled children in classrooms, which resulted in a significant disparate impact on the basis of race).

\textsuperscript{204} See, e.g., Grove City Coll. v. Bell, 465 U.S. 555, 566-70 (1984) (citing Title VI congressional record to determine whether private college students' receipt of federal financial aid funds triggers Title IX coverage); Cannon v. Univ. of Chi., 441 U.S. 677, 683-85 (1979) (applying Title VI decisions to determine whether a private right of action exists under Title IX); \textit{Sharif}, 709 F. Supp. at 360-61 (utilizing Title VI standards to determine whether plaintiffs must prove discriminatory intent to prevail in a Title IX action, and to determine the other elements the plaintiffs and defendants must show in order to succeed in the action).
These standardized tests, particularly the SAT, have a profound impact on our lives. We spend only a few hours taking the SAT, yet we spend days or even weeks anxiously waiting for the results and celebrating or agonizing over the results once they arrive. And we never forget those scores no matter how hard we try.

In addition to having a personal impact on our lives, standardized tests have an impact on society as a whole. The tests have the unintended consequence of discriminating against women and minority test takers. Because the SAT has not been changed substantially over the years, the adverse impact on women and minorities remains consistent and significant.

The mere existence of the SAT may be legally and successfully challenged under the disparate impact theory, although the Supreme Court has not yet addressed this question. To have a valid claim, plaintiffs need to sue to enforce the regulations of Title IX or Title VI, and must prove that there is a disparate impact on the group in question based on the facially neutral test. Then, the defendant may rebut by showing that the testing practice is an educational necessity. Finally, the plaintiff may still prevail by offering an alternative to the existing practice that achieves the objective of the practice but which has a less discriminatory effect.

However, there is a lack of guidance in the regulations implementing the statutes on how to approach the issue of disparate impact. There are no standards to analyze the effect of the educational practice. How much effect is significant enough for a plaintiff to succeed? To achieve consistent and just results in court, this question must be answered and subsequently codified into federal law. The issue of remedies is another problem which must be addressed by the legislature and codified into law for uniformity and predictability.

Standardized test scores are a superficially objective portion of college admissions decisions, a process which in reality seems to be overwhelmingly subjective. However, the resulting disparity in scores between males and females and between white and minority test takers is too blatant to be random and too significant to be deemed to have a neutral impact on these groups of people when the test scores are used to

205. See discussion supra Part IV.
make such important decisions. The SAT is discriminatory in effect and must be either eliminated altogether or substantially transformed.

Andrea L. Silverstein*

* I would like to thank Professor Joanna L. Grossman for her insight, guidance, and expertise throughout the development of this Note; and the staff and particularly the Managing Editors of the Hofstra Law Review for their tireless efforts in their pursuit for perfection. I would also like to extend my deepest gratitude and love to my friends and family, and especially to David Glaser, for their unwavering support and understanding, especially during these past three crazy years. This Note is dedicated to my parents, Lenore and Arthur, and to my brother David, for their unconditional love and encouragement and for teaching me that obstacles, such as less-than-stellar standardized test scores, will not stand in the way of achieving my dreams.