September 2013

The Legal Implications of Gender Bias in Standardized Testing

Katherine Connor
Ellen J. Vargyas

Follow this and additional works at: http://scholarship.law.berkeley.edu/bglj

Recommended Citation
Available at: http://scholarship.law.berkeley.edu/bglj/vol7/iss1/2

Link to publisher version (DOI)
http://dx.doi.org/https://doi.org/10.15779/Z38GS2D

This Article is brought to you for free and open access by the Law Journals and Related Materials at Berkeley Law Scholarship Repository. It has been accepted for inclusion in Berkeley Journal of Gender, Law & Justice by an authorized administrator of Berkeley Law Scholarship Repository. For more information, please contact jcera@law.berkeley.edu.
The Legal Implications of Gender Bias in Standardized Testing

Katherine Connor†
Ellen J. Vargyas‡

TABLE OF CONTENTS

I. INTRODUCTION ....................................... 14
II. THE FACTUAL CONTEXT ............................. 17
   A. The Scope of the Problem ......................... 17
      1. Post-Secondary Admissions Tests .............. 18
      2. Vocational Aptitude Tests and Interest Inventories 21
   B. Causes of Gender Differences in Test Scores ........ 24
      1. Post-Secondary Admissions Tests .............. 24
      2. Vocational Aptitude Tests and Interest Inventories 27
   C. Validity of the Tests ............................ 28
      1. Post-Secondary Admissions Tests .............. 29
         a. Tests Which Predict Differently by Gender ... 29
         b. Unvalidated Test Uses ........................ 31
      2. Vocational Aptitude Tests and Interest Inventories 32
         a. Failure to Measure Aptitude .................... 35
         b. Interest Inventory Validation Problems ...... 37
   D. The Factual Context: Conclusions ............... 38

Copyright © 1992 by the National Women's Law Center, 1616 P Street, NW, Washington, D.C. 20036. All rights reserved.
† Ms. Connor co-authored this article while staff counsel with the National Women's Law Center in Washington, D.C. She is currently an associate with Comey & Boyd in Washington, D.C. She received her B.A. from Middleburg College in 1983, and her J.D. from the University of California at Los Angeles in 1988.
‡ Ms. Vargyas is Senior Counsel with the National Women's Law Center in Washington, D.C. She received her B.A. from Williams College in 1971, and her J.D. from the University of Pennsylvania Law School in 1974.

The authors thank Marcia D. Greenberger, Co-President of The National Women's Law Center, Caroline Avery, Staff Attorney, and Dana Martin and Samantha Forman, law students at the Center, for their assistance in writing this article, and Riley E. Schifflaeley for the word processing. They also thank the Spencer Foundation for its financial support for the research and development of this article.
III. THE APPLICABLE LEGAL ANALYSIS 38
   A. Title IX 39
      1. The Title IX Prohibition Against Gender Discriminatory Test Uses 40
      2. Title IX and Its Regulations Reach Both Intentional and Disparate Impact Discrimination 42
      3. Proving Disparate Impact Discrimination in Testing Cases: The Title VII Analogy 45
         a. Griggs v Duke Power Co. and Its Progeny 47
         b. The Uniform Guidelines on Employee Selection 50
         c. Implications of the Civil Rights Act of 1991 58
      4. The Particular Legal Issues Confronting Women of Color: The Intersection of Title IX and Title VI 63
   B. Federal Constitutional Analysis 67
      1. Heightened Scrutiny 68
      2. Rational Relationship 70
   C. State Constitutional Guarantees and Laws Prohibiting Gender Discrimination in Education 71
      1. State Constitutional Guarantees 72
         a. Equal Rights Amendments 72
         b. Equal Protection 74
      2. State Educational Equity Laws 75
   D. Remedies for Gender-Discriminatory Uses of Educationally Related Tests 77
      1. Government Enforcement and Remedies 77
      2. Private Enforcement and Remedies 79
         a. Injunctions Barring or Restricting the Use of Discriminatory Tests or Requiring Adoption of Valid Selection Procedures 80
         b. Remedies Addressing the Continuing Effects of Past Discrimination 83
         c. Voluntary Remedies 84
         d. Monetary Damages 86
         e. Attorneys Fees 87

IV. CONCLUSIONS 87

I. INTRODUCTION

Standardized testing plays a major role in allocating a wide variety of benefits which our society offers in the field of education.1 These range

1 "The term standardized test means that all examinees are given identical directions, time limits, and questions." National Commission on Testing and Public Policy, From Gatekeeper To Gateway: Transforming Testing In America 2 (Boston College, 1990) "From Gatekeeper To
from college admissions and scholarships, including athletic scholarships, to entry into vocational training programs, and access to programs for gifted and talented adolescents. Standardized tests are used to judge the comparative successes and competitiveness of schools ranging from the elementary level to the post-secondary and to evaluate students starting before kindergarten. Standardized tests have also assumed a major place in the current debate over education reform.

The Commission estimates that each year elementary and secondary students take 127 million separate standardized tests, id at 15, at a direct cost of between $70 million and $107 million annually. Id at 17.

The most widely used tests in this regard are the Scholastic Aptitude Test (SAT) and the Preliminary Scholastic Aptitude Test (PSAT), both published by the College Entrance Examination Board and the Educational Testing Service (ETS), and the American College Testing Program Examination (ACT), published by the American College Testing Program. Nearly 1,500 four-year colleges and universities nationwide require standardized test scores or use them as cutoff scores for admission. Phyllis Rosser, The SAT Gender Gap: Identifying the Causes and Effects (Preliminary Findings, 1989) ("SAT Gender Gap Study").

SAT and/or PSAT scores are used in awarding numerous college scholarships including, for example, National Merit Scholarships, Air Force, Army and Navy Reserve Officer Training Corps ( ROTC ) scholarships, and scholarships awarded by the states of New York, Maryland, Massachusetts, Nevada, and Rhode Island. See notes 25-27 and accompanying text.

The National Collegiate Athletic Association, for example, has linked eligibility for athletic scholarships to SAT scores through its Propositions 48 and 42. Rosser, SAT Gender Gap at 91 (cited in note 2).

According to a national survey of secondary schools, vocational aptitude tests and interest inventories are among the most widely used standardized tests in the schools: approximately 95% of schools administered at least one test of this type to some students, and 75% administered a vocational aptitude test or interest inventory to all students. Harold B. Engen, Richard R. Lamb, and Dale J. Prediger, Are Secondary Schools Still Using Standardized Tests?, 60 Personnel and Guidance J 287, 288 (1982). The most popular vocational aptitude test is the Department of Defense's Armed Services Vocational Aptitude Battery (ASVAB), which 66% of the schools reported using. Id. The ASVAB is provided to the schools free of charge by the Department of Defense to over 1.3 million students annually. Dept. of Defense, Counselor's Manual for the ASVAB Form 14 ix (Dept. of Defense, 1989) ("ASVAB Counselor's Manual"). Test results are made available to the schools and students and are used by the military for recruiting. Id at 1, 2. The second most popular test is the Differential Aptitude Test (DAT), initially published by Psychological Corporation in 1947, and used in 34% of the schools. Engen, Lamb, and Prediger, 60 Personnel and Guidance J at 288. Interest inventories are also widely used, although to a lesser extent than these two vocational aptitude tests. Id.

Admission to programs for academically talented junior high school students run by, among others, Johns Hopkins University, Duke University, Northwestern University, Arizona State University, several campuses of the University of California, and the University of Denver, is based on SAT scores. Gita Wilder and Patricia Lund Casserly, Young SAT-Takers: Two Surveys, College Board Report No 88-1 (College Entrance Examination Board, 1988) ("Young SAT-Takers"); Rosser, SAT Gender Gap at 89 (cited in note 2); Johns Hopkins University, Catalogue for the Summer Programs for the Center for the Advancement of Academically Talented Youth 6 (1990). In order to be admitted to the liberal arts courses of the Johns Hopkins program, a thirteen-year-old must achieve a 430 on the verbal section of the SAT. To participate in the math and science courses, he or she must also score at least 500 on the math section. Rosser, SAT Gender Gap at 89 (cited in note 2).

Test scores are even used to evaluate school superintendents and to determine levels of state funding, with schools receiving bonuses for annual score gains. See, for example, Amy Goldstein, Finding A New Gauge of Knowledge; Some States Are Designing Alternatives to Standardized Testing, Wash Post, A20 (May 20, 1990).

According to the National Commission on Testing and Public Policy, "prekindergarten tests are mandated in more than 16 states, widely used in seven states, and known to be used at the district level in more than 100 states. Kindergarten exit/first-grade entrance tests are used in at least 5 states and are known to exist at the district level in an additional 37." National Commission on Testing and Public Policy, From Gatekeeper To Gateway at 14 (cited in note 1).
In many of these tests there are substantial scoring differentials among various population groups. Differentials based on race and national origin are well documented and have been the subject of litigation9 and scholarly legal attention.10 However, many widely used standardized test scores — ranging from the Scholastic Aptitude Test (SAT) to the Armed Services Vocational Aptitude Battery (ASVAB) and beyond — also reflect substantial differences based on gender. Minority females suffer a double jeopardy as they often score lower than both white females and males of their own racial or ethnic group. Nonetheless, the law regarding both gender issues and combined gender and race issues is largely undeveloped. Only one lawsuit, Sharif v New York State Education Department,11 has been brought to challenge any use of a standardized test on the grounds of gender bias. Legal scholars have only recently begun to devote attention to this issue and have focused on Sharif.12 Virtually no attention has been focused on the particular issues raised in connection with minority girls and women.

Building on the legal principles which have been developed regarding test bias on the basis of race and/or ethnicity in employment and education, along with the analysis in Sharif, this article will present a legal model for analyzing claims of gender bias in standardized testing and test use. Key topics will include the impact of federal and state statutory and constitutional protections against sex discrimination as they apply to both liability and remedies. Throughout, we will pay particular attention to the framework in which to analyze the multiple discrimination suffered by minority females.

This article will first review the relevant empirical literature to

9 See, for example, Larry P. v Riles, 495 F Supp 926 (ND Cal 1979), aff'd in part and rev'd in part, 793 F2d 969 (9th Cir 1984) (challenge by Black students to the use of IQ tests to place students in classes for the educable mentally retarded); Debra P. v Turlington, 474 F Supp 244 (MD Fla 1979), aff'd in part and vacated in part, 644 F2d 397 (former 5th Cir 1981), on remand, 564 F Supp 177 (MD Fla 1983), aff'd, 730 F2d 1405 (11th Cir 1984) (challenge to minimum competency test with a substantial adverse impact on minority students); United States v LULAC, 793 F2d 636 (5th Cir 1986) (challenge by Black and Hispanic students to requirement that college students pass a skills test before taking professional education courses).


11 709 F Supp 345 (SD NY 1989). Sharif challenged New York State’s exclusive use of SAT scores in awarding state-sponsored scholarships. In granting plaintiff’s motion for a preliminary injunction, the court ruled that this practice violated Title IX of the Education Amendments of 1972 and its regulations, as well as the Fourteenth Amendment guarantee of equal protection.

establish the scope and nature of the problem. We will then turn to the legal questions at issue.

II. THE FACTUAL CONTEXT

The factual predicate for the legal consideration of the issues of gender-in-testing includes three basic elements. The first is the scope of the problem, including both the range of tests which reflect the differentials and the extent to which the uses of these tests adversely affect females, including females of color. The second is the available information regarding the underlying explanation — or lack thereof — for the gender differentials. The third is the extent to which these tests actually measure what they purport to measure, that is, whether the test uses are valid. We will consider these matters in turn.

A. The Scope of the Problem

Gender differentials in scoring are found in a broad array of standardized tests. They are particularly prevalent in connection with two broad categories of tests: (1) the complex of admissions tests for colleges, graduate schools, and professional schools, such as the Scholastic Aptitude Test and achievement tests, which are also used for many non-admissions purposes; and (2) tests used for vocational education course selection, placement, and career counseling, such as the Armed Services Vocational Aptitude Battery (ASVAB), the Differential Aptitude Test (DAT), and career interest inventories.

Other tests, particularly those administered to nationally representative, non-self-selecting samples of students, may not reflect such differences. Some researchers are of the view that these tests demonstrate that gender differences in both verbal and math areas are declining, possibly to the point of insignificance.\(^1\) The major exception in this analysis appears to be a persistent remaining difference at the high end of math

---

\(^1\) See, for example, Carol Nagy Jacklin, *Female and Male: Issues of Gender*, 44 Am Psychologist 127, 128 (1989) (current research indicates gender differences in verbal ability do not exist, and gender differences in other intellectual abilities are also beginning to decrease); Alan Feingold, *Cognitive Gender Differences are Disappearing*, 43 Am Psychologist 95, 101 (1988) (gender differences in verbal skills are now virtually nonexistent; gender differences in math skills have significantly decreased over the past decade, although a significant gap still exists in higher math skills at the high school level); Janet Shibley Hyde and Marcia C. Linn, *Gender Differences in Verbal Ability: A Meta-Analysis*, 104 Psychological Bull 53, 53 (1988) ("[m]any regard gender differences in verbal ability to be one of the well-established findings in psychology. To reassess this belief, we located 165 studies that reported data on gender differences in verbal ability. The weighted mean... indicated a slight female superiority in performance. The difference is so small that we argue that gender differences in verbal ability no longer exist"); and Gita Z. Wilder and Kristin Powell, *Sex Differences in Test Performance: A Survey of the Literature*, College Board Report No 89-3, 4-9 (College Entrance Examination Board, 1989) ("Sex Differences in Test Performance") (gender differences in verbal skills no longer exist, but gender differences in quantitative skills persist, especially in high-end mathematics).
testing, with very high-scoring boys substantially outnumbering very high-scoring girls.\textsuperscript{14} However, the scoring patterns on some tests administered to the general population suggest that broadly based gender differentials do persist. For example, a recent review of the findings of the National Assessment of Educational Progress (NAEP) over the past twenty years identifies ongoing test performance disparities between males and females in reading and writing (in favor of females) and in mathematics, science, history, civics, and geography (in favor of males).\textsuperscript{15}

The inconclusive and apparently conflicting state of the literature is indicative of the underlying problem of inattention to the issue of gender bias in testing. In light of this problem, this article focuses on the tests for which the disparities are well established.

1. Post-Secondary Admissions Tests

Post-secondary admissions tests include, for example, the SAT, the ACT, achievement tests, graduate record examinations, and law, medical, and business school entrance examinations. Gender scoring differentials in these tests are well established and, where it has been conducted, research has identified gender/race differentials as well. For example, for many years females have scored approximately sixty points lower than males on the SAT, with a female deficit in both the verbal and math sections of the test.\textsuperscript{16} Similar differences are found in the Preliminary Scholastic Aptitude Test (PSAT)\textsuperscript{17} and the ACT\textsuperscript{18} as well as most college entrance achievement tests and professional and graduate school

\textsuperscript{14} See, for example, Diane I. Halpern, \textit{The Disappearance of Cognitive Gender Differences: What You See Depends on Where You Look}, 44 Am Psychologist 1156, 1156-57 (1989); Feingold, 43 Am Psychologist at 101 (cited in note 13); Jacklin, 44 Am Psychologist at 128 (cited in note 13).

\textsuperscript{15} Ina V.S. Mullis, Eugene H. Owen, and Gary W. Phillips, \textit{Accelerating Academic Achievement} 53 (ETS, 1990). See also Ina V.S. Mullis and Lynn B. Jenkins, \textit{The Reading Report Card, 1971-1988: Trends From the Nation's Report Card} 17-18 (Dept. of Educ., 1990). Interestingly, Rosser notes that the NAEP is based on tests written by ETS and that the NAEP results are used to justify gender differentials in other ETS tests such as the SAT. Rosser, \textit{SAT Gender Gap} at 73 (cited in note 2).

\textsuperscript{16} Males have achieved higher math scores than females since the inception of the SAT. Since 1967, the math differential has ranged from a low of 43 to a high of 52 points. Until 1971, females outscored males on the verbal portion of the test, although by a much smaller factor than the difference in the math scores. Starting in 1972, females lost this modest advantage and have fallen behind males on the verbal scores by a factor of between 2 and 13 points in each ensuing year. Since 1972, the total female scoring deficit has ranged between 45 and 61 points. In 1991, it was 52 points. College Board, \textit{College Bound Seniors: 1991 Profile of SAT and Achievement Test Takers} iii (College Entrance Examination Board, 1991) ("College Bound Seniors").

STANDARDIZED TESTING

entrance examinations. For women of color the differences are even more pronounced. Women of color consistently score lower on both the SAT\textsuperscript{20} and the ACT\textsuperscript{21} than both their white female classmates and the men in their ethnic/racial group. Gender/race data is not generally available for other tests.

These scoring differences directly and concretely affect the distribution of the many valuable education-related benefits which are allocated on the basis of this complex of tests, principally scholarships and admissions. For example, year-in and year-out, between sixty and sixty-six percent of the prestigious National Merit Scholarships — totalling over $23 million annually — are awarded to young men.\textsuperscript{22} The National Merit Scholarship Program uses PSAT scores as the sole criterion for determining its semi-finalist pool from which all scholarship winners are selected.\textsuperscript{23} Similarly, when New York State relied exclusively on SAT scores to allocate state-sponsored scholarships, over seventy percent of its elite Empire State Scholarships were awarded to young men, as were approximately sixty percent of its Regents Scholarships.\textsuperscript{24} Many other scholarship programs also rely on these scores, at least in part, in making scholarship decisions. These include the Army, Air Force and Naval

\begin{itemize}
\item \textsuperscript{18} Rosser, \textit{SAT Gender Gap} at 26 (cited in note 2) (noting that in 1987-88, males received an average composite score of 19.9 while females received an average composite score of 18.6).
\item \textsuperscript{19} Studies show that males routinely outscore females on 10 of the 14 College Board Achievement Tests (the exceptions are English Composition, German, Hebrew, and Literature), and on the quantitative sections of the Graduate Record Examination, the Medical College Admissions Test, and the Graduate Management Admissions Test. Data from the Law School Admissions Service regarding the Law School Admissions Test shows females earning slightly lower test scores although their grade point averages in school are slightly higher. Wilder and Powell, \textit{Sex Differences in Test Performance} at 2-3 (cited in note 13). See also Patricia Wheeler and Abigail Harris, \textit{Comparison of Male and Female Performance on the ATP Physics Test}, College Board Report No 81-4, 1 (College Entrance Examination Board, 1981) (showing that men also outperform women on the ATP Physics Test).
\item \textsuperscript{20} On the 1991 SAT, Latin American men scored 60 points higher than Latin American women (math and verbal combined); Mexican American men scored 55 points higher than Mexican American women; Asian American men scored 55 points higher than Asian American women; Puerto Rican men scored 49 points higher than Puerto Rican women; Native American men scored 36 points higher than Native American women; and African American men scored 22 points higher than African American women. College Board, 1991 \textit{National Ethnic/Sex Data} (College Entrance Examination Board, 1991). College Board studies for earlier years have also found a consistent score gap between men and women within each racial/ethnic group. Leonard Ramist and Solomon Arbeiter, \textit{Profiles, College-Bound Seniors 1985} xix-xxiii (College Entrance Examination Board, 1986).
\item \textsuperscript{21} From highest scoring to lowest, the results for the 1987-88 ACT were as follows: white males, Asian American males, white females, Asian American females, Puerto Rican males, Puerto Rican females, Native American males, Mexican American males, Native American females, Mexican American females, African American males, and African American females. Rosser, \textit{SAT Gender Gap} at 26 (cited in note 2).
\item \textsuperscript{22} Id at 85.
\item \textsuperscript{23} The National Merit Scholarship Program does use geographic quotas in identifying semi-finalists. Cutoff scores are determined on a state-by-state basis to assure that semi-finalists will represent all areas of the country. The result is that qualifying scores diverge depending on the applicant's residence. See, for example, College Board, 1988 \textit{PSAT/NMSQT Student Bulletin} 39 (College Entrance Examination Board, 1988).
\item \textsuperscript{24} Sharif, 709 F Supp at 355.
\end{itemize}
Reserve Officer Training Corps (ROTC) programs, state merit scholarship programs, and other public and private scholarship programs.

Furthermore, the gifted and talented programs for junior high school students which base admissions on SAT scores are disproportionately comprised of male students. Indeed, at the first and most well-known of these programs, the Johns Hopkins Center for the Advancement of Academically Talented Youth, the math and science summer courses are approximately sixty-five percent male and only thirty-five percent female. This distribution exists despite the fact that more than half of the young test-takers are female.

The literature does not include analyses of the impact of these tests on the admission of women to colleges and graduate and professional schools. Although women comprise a slight majority of college students, they follow different patterns of college enrollment than do men. For example, women are disproportionately enrolled as part-time students in two- and four-year undergraduate programs and graduate programs. The evidence strongly suggests that students adjust their college expectations based on their SAT or ACT scores; lower-scoring females apply to less competitive colleges and universities than their grades would warrant. And once they are in a college or university, women tend to enroll in areas in which women have traditionally studied. In all racial and ethnic groups, women cluster in such fields as education, foreign languages, health, home economics, letters, liberal studies, and psychology, whereas many more men than women enroll in such fields as engineering, mathematics, physical sciences, and protective services.

---

26 Rosser, SAT Gender Gap at 85-86, 107-16 (cited in note 2).
27 Id at 85. These include, for example, Alcoa Foundation Scholarships, the International Brotherhood of Teamsters Scholarship Fund, the LULAC National Scholarship Fund, the National Achievement Scholarship Program for Outstanding Negro Students, National Presbyterian College Scholarships, the Navy Boost Program, and the Permian Honor Scholarship. The College Board, Registration Bulletin 1989-90, SAT and Achievement Tests 24 (College Entrance Examination Board, 1989).
28 Rosser, SAT Gender Gap at 22-23 (cited in note 2).
29 Telephone interview between John Chung, Research Coordinator of the Center for the Advancement of Academically Talented Youth, Johns Hopkins University, and Samantha Forman, National Women's Law Center (Feb 26, 1990).
30 See Wilder and Casserly, Young SAT-Takers at 4 (cited in note 6).
32 Id.
33 Rosser, SAT Gender Gap at 22 (cited in note 2) (citing Ernest L. Boyer, College: The Undergraduate Experience in America (Harper & Row, 1987)). See also Wilder and Powell, Sex Differences in Test Performance at 31 (cited in note 13) (noting that women tend not to go into science fields because of their lower test scores).
34 National Center for Education Statistics, Digest of Education Statistics at 243-44 (cited in note
programs. However, as is true with many aspects of the effect of tests on women, relatively little is known about the precise influence of tests on these patterns.

2. Vocational Aptitude Tests and Interest Inventories

It is well-documented that sex segregation in vocational training also remains a deeply troubling and persistent problem. According to the final report of the National Assessment of Vocational Education, there has been little change in the last fifteen years in the amount of sex segregation in the fields of agriculture, construction, mechanics and repair, health, and occupational home economics. Females are overwhelmingly concentrated in training for low-wage, non-technical, traditionally female job paths.

The most severe and persistent sex segregation is experienced by low-income and academically disadvantaged students, who are disproportionately students of color. For example, approximately half of all the vocational credits earned by disadvantaged women are in low-level service occupational courses or consumer and homemaking education. Further, disadvantaged females take even fewer technical and communications courses than do advantaged females, who themselves take a small number of such courses. By contrast, academically disadvantaged males enjoy higher quality vocational education than do academically disadvantaged females.

There are troubling connections between the gender segregation in vocational education and the use of vocational education tests. Substan-
tial sex-differences in scoring are reflected on the two most frequently used vocational aptitude tests in secondary schools, the ASVAB and the DAT. Moreover, the combined effect of gender and race results in particularly low scores for females of color on the ASVAB.

Although DAT results by both sex and race can be easily calculated, they are not made available by the test publisher, Psychological Corporation.

Career interest inventories, which are widely used in the secondary schools for vocational education counseling and placement, also result in substantial gender-based score differentials. For example, on the widely used Holland themes, women obtain higher scores on Social,
Artistic, and Conventional themes, while men obtain higher scores on the Realistic, Investigative, and Enterprising themes. As a result, the career suggestions based on these scores tend to be in careers that are traditional for each sex. Women are often directed toward education, social welfare, and office occupations, while men are commonly pointed toward careers in medicine, engineering, management, trades, or technical fields. Some test publishers have responded to this problem by developing same-sex norms, that is, comparing scores within each gender and not across genders.

Although same-sex norms may result in some women receiving career suggestions in nontraditional fields, they are not responsive to the larger criticism that interest inventories perpetuate stereotyped socialization patterns and a segregated workforce because they typically compare an individual's likes and dislikes to those of persons already in the workforce. Given the extreme sex and race segregation common in the workplace, this concern is significant. The issue is particularly acute for women of color, who may be doubly penalized by sex and racial/ethnic biases in interest inventories.

Despite the widespread use of vocational aptitude tests and interest inventories with sex-traditional score results, relatively little data is available indicating exactly how schools use the tests and whether they contribute to sex and race segregation in vocational education. To be sure, many factors influence students' vocational program choices; family pressure and socialization, self-image, peer pressure, and educational experiences, including guidance counseling and testing, may all contribute to a student's decision to enter a traditional or nontraditional program. However, numerous studies have found evidence of sex-stereotyped counseling in schools. A consistent and troubling finding has been that students who select nontraditional programs do not report receiving positive encouragement from guidance counselors in their choice.

48 Id.
49 Diamond and Tittle, Sex Equity in Testing at 180 (cited in note 45).
50 In 1989, 60% of all professional women worked in two traditionally female occupations: teaching and nursing. Over half of all African American and Hispanic women workers were employed in clerical and service occupations. In addition, in 1988, 46% of all women workers earned less than $10,000 per year, compared to 26% of all male workers, and 65% of minimum wage earners were women. National Commission on Working Women of Wider Opportunities for Women, Women and Work (Wider Opportunities for Women, 1990).
51 See, for example, Dianne Sauter, Ann Seidl, and Jacqueline Karbon, The Effects of High School Counseling Experience and Attitudes Toward Women's Roles on Traditional or Nontraditional Career Choice, 28 Vocational Guidance Q 241, 245 (1980) (finding that not a single woman taking nontraditional courses reported that guidance counseling influenced her choice, while 25% of women taking traditional courses reported being influenced by guidance counselors); Elizabeth H. Giese, Expanding Occupational Choices in Michigan's Secondary Vocational Education, in Sharon L. Harlan and Ronnie J. Steinberg, eds, Job Training for Women:
Moreover, the limited available evidence indicates that vocational tests and interest inventories reinforce some guidance counselors' practice of discouraging women from pursuing nontraditional programs. For example, a Michigan League of Women Voters survey of vocational education teachers, counselors, administrators, and students, strongly suggests that the tests directly contribute to that state's extreme sex segregation in vocational education programs.\textsuperscript{52}

[F]ully 40 percent of the teachers and counselors said students are required to provide evidence of vocational interest in a subject, such as favorable test scores or having taken prerequisites, before they are allowed to enroll. Indeed, 70 percent of students surveyed had taken a standardized test to identify their career interest, and 44 percent of the teachers and counselors said interviews were required before students could enroll in vocational education. And, 59 percent of the students pursuing non-traditional vocations felt there were admission criteria for entering a vocational education school or courses, compared to only 36 percent of traditional students who perceived such admission criteria.\textsuperscript{53}

Thus, rather than expand vocational options, aptitude tests and inventories heighten the other systemic pressures that make a young woman's pursuit of nontraditional vocational training extremely unlikely. Further, vocational education continues to serve as the training ground for a segregated workplace.

\section*{B. Causes of Gender Differences in Test Scores}

\subsection*{1. Post-Secondary Admissions Tests}

Despite the clear presence of gender scoring differentials in these tests, there are no clear answers regarding the underlying causes. As a recent analysis prepared by two Educational Testing Service researchers observes:

At the outset it should be noted that the conclusions about gender differences that can be reached at the current time are limited. For all of the attention that the subject has received, the data that support many of the contentions made about gender differences and their causes are inconclusive and often contradictory. The majority of studies lack generalizability, based as they are on different populations or on performance in limited domains by small samples of individuals. . . . Complicating the issue still further are the different conclusions that researchers have managed to reach even when they work from the same data.\textsuperscript{54}

\begin{thebibliography}{99}
\bibitem{52} Giese, \textit{Expanding Occupational Choices} at 21-24 (cited in note 51).
\bibitem{53} Id at 322.
\bibitem{54} Wilder and Powell, \textit{Sex Differences in Test Performance} at 1 (cited in note 13).
\end{thebibliography}
The President of the College Board has reached the same conclusion regarding his premier test: "We do not pretend to be able to explain fully the reasons why men and women perform the way they do on the SAT." Similarly, the race and ethnicity differentials on the SAT are not fully explained, and virtually no research has focused on the reasons for the particularly low scores of minority women.

The available literature — which focuses almost exclusively on the SAT — examines the issue from multiple perspectives, but the bottom line is that the differences, even for this one test, are only partially understood. One recurring theme is that non-gender demographic characteristics may influence SAT test scores. The female test-taking population differs demographically from the male test-taking population in a number of ways: more females than males take the test; the females are disproportionately members of racial and ethnic minority groups; and the females are disproportionately from families with lower incomes and levels of parental education. The argument is made that these differences, rather than gender differences, account for the scoring differential. To the extent that this argument assumes that differences attributable to race and ethnicity reflect real differences in ability, it is misguided. There is substantial evidence of racial, ethnic, and cultural biases in standardized tests.

Moreover, recent studies of score differentials on the basis of gender which analyze the impact of the demographic differences strongly suggest that they do not account for the full magnitude of the observed score differences. The College Board’s studies show that, without regard to gender, Blacks, Hispanics, and Native Americans score substantially lower than whites, and that SAT scores vary directly both with family income and level of parental education. Id.

The recent report of the National Commission on Testing and Public Policy documents research showing that the way test content is oriented — toward the topics and culture of the dominant group in society as opposed to minority groups — can significantly affect test scores. National Commission on Testing, From Gatekeeper to Gateway at 11-13 (cited in note 1). See also Orlando L. Taylor and Dorian Latham Lee, Standardized Tests and African-American Children: Communication and Language Issues, 38 Negro Educ Rev 67 (Apr-Jul 1987) (situational, linguistic, and communicative style, cognitive style, and interpretation biases in the use of standardized tests harm students of different cultural and linguistic backgrounds); Mary Rhodes Hoover, Robert L. Politzer, and Orlando L. Taylor, Bias in Reading Tests for Black Language Speakers: A Sociolinguistic Perspective, 38 Negro Educ Rev 81 (Apr-Jul 1987) (documenting specific types of language-related bias in tests and the consequences of the biases, including the tracking of students into low-level classes).
differentials, although they may explain some of the differences. Further, the data clearly show that even when demographic factors are held constant, males outscore similarly situated females. That is, males of every racial and ethnic group for which there is data outscore comparable females; males outscore females at each level of family income; and males outscore females at each level of parental education.

Another frequently advanced argument is that females earn lower test scores because they take fewer high-level math and science courses and otherwise pursue a less rigorous preparatory curriculum. The available literature suggests, however, that differences in academic preparation are minimal and, in any event, do not fully explain the scoring differentials. In fact, SAT scoring differentials comparable to those achieved by high school students are reflected in the scores of male and female junior high school students who participate in the various talent search programs. Yet their course-taking patterns are virtually identical.

Other proffered explanations center on the effects of artifacts of the tests themselves, such as the context of the questions, the "speeded" or time-constrained nature of the tests, the impact of the guessing penalty, and the differential impact of particular test items on male and female and majority and minority test-takers.

60 See, for example, Clark and Grandy, Sex Differences in Academic Performance at 1 (cited in note 55) (while male and female SAT-candidates differ in course preparation, grade point average, major field, career interests, and socioeconomic background, neither these differences nor evidence on differential cognitive functioning are sufficient to account for all of the observed sex differences in performance on the SAT).

61 Rosser, SAT Gender Gap at 49 (cited in note 2).

62 Id at 171-72.

63 Id.

64 Id at 24-25.

65 See, for example, Clark and Grandy, Sex Differences in Academic Performance (cited in note 55); but see Wheeler and Harris, Comparison of Male and Female Performance on the ATP Physics Test at 37 (cited in note 19) (for takers of the ATP Physics Test, sex differences were not altered by the number of years of math taken, but the male/female discrepancy was significant when the number of semesters of physics taken was increased beyond two).

A study of secondary school students taking the same math classes found that males received substantially higher math test scores than females, despite the fact that females obtained higher grades in math classes. Susan Gross, Participation and Performance of Women and Minorities in Mathematics E-4 (Dept. of Educ. Accountability, 1988).

66 See Wilder and Casserly, Young SAT-Takers at 43-44 (cited in note 6).

67 See, for example, Cathy L.W. Wendler and Sydell T. Carlton, An Examination of SAT Verbal Items for Differential Performance by Women and Men: An Exploratory Study (paper presented at the 1987 annual meeting of the American Educational Research Association in Washington, D.C.) (on file with the National Women's Law Center) (data suggest that women do better on test questions that are related to each other and are related in a larger context, on items dealing with human relationships and humanities rather than the world of practical affairs, on antonyms with nouns, and on items that are more abstract, general, or intangible); Marcia C. Linn, et al, Gender Differences in National Assessment of Educational Progress Science Items: What Does I Don't Know Really Mean?, 24 J Res Sci Teaching 267, 267 (1987) (females more likely to use the "I don't know" option than males); and Rosser, SAT Gender Gap at 64-65 (cited in note 2) (females more likely to encounter problems with the speeded nature of the test, especially in math sections). See also R.J.L. Murphy, Sex Differences in Objective Test Performance, 52 Br J Educ Psych 213 (1982); Carol A. Dwyer, The Role of
forthrightly that test publishers control the differentials through choices made in test construction.\(^{68}\)

In sum, the research, fairly analyzed, does not fully explain why females, and particularly minority females, achieve substantially lower scores on the SAT than do their male peers. Theories abound, but the bottom line is that the score differentials remain in large part unexplained. The literature is virtually silent on the causes underlying the score differentials in the other post-secondary admissions tests.

2. Vocational Aptitude Tests and Interest Inventories

The limited available research in this area suggests score differentials between males and females on the ASVAB and DAT reflect the different socialization patterns and experiences of the two groups.\(^ {69}\) Although the DAT and the ASVAB are called “aptitude” tests, they actually measure experience or learned abilities rather than aptitude. For example, on the ASVAB, “the inability of a high school girl to recognize a pipe-cutter or to say what a thermocouple is used for indicates nothing about the careers for which she can be trained.”\(^ {70}\) Because aptitude tests are routinely administered early in a student’s education, a school’s use of tests for vocational education placement may result in the channeling of students into low-level and sex-traditional classes and careers based on the socialization that occurred in grade school or the first years of junior high school. Thus, rather than expand opportunities, the use of aptitude tests for counseling and placement may constrain them. Moreover, there is virtually no research devoted to examining the

---

\(^{68}\) See, for example, Thomas F. Donlon, Marilyn H. Hicks, and Madeline M. Wallmark, Sex Differences in Item Responses on the Graduate Record Examination, 4 Applied Psych Measurement 9 (1980) (arguing that test constructors have the power to substantially vary scaled score differences depending on the type of questions included); and Dwyer, Role of Tests at 340 (cited in note 67).

\(^{69}\) See, for example, Thomas F. Donlon, Marilyn H. Hicks, and Madeline M. Wallmark, Sex Differences in Item Responses on the Graduate Record Examination, 4 Applied Psych Measurement 9 (1980) (arguing that test constructors have the power to substantially vary scaled score differences depending on the type of questions included); and Dwyer, Role of Tests at 340 (cited in note 67).


large score differentials that operate against women of color on vocational aptitude tests.

The different socialization experiences of males and females also have a significant influence on interest inventory results. Interest inventories typically require test-takers to rate themselves in a number of areas, including interests, skills, abilities, values, needs, and occupational preferences. What test-takers have been socialized to believe are appropriate interests and occupations for their sex greatly influence test results. Research reveals that occupational stereotypes — the belief that certain careers are appropriate for males and others for females — are prevalent at a young age. Thus, while elementary-age boys indicate a wide variety of occupational preferences, mostly in male-dominated occupations, elementary-age girls list a much smaller number of occupations, with most choosing nursing and teaching. This occupational stereotyping persists among college-age and adult populations.

Interest inventories also reinforce sexual stereotyping because women tend to underrate their skills and abilities, particularly when it comes to traditionally male-oriented tasks. Claims of ability are often based more upon interest and self-confidence than upon aptitudes, with females displaying low confidence in their math abilities and high interest in their ability to serve others and males displaying high confidence in their ability to improve the performance of machinery and low interest in providing service.

This research begins to explain why interest inventory results magnify rather than diminish the differences between males and females. Again, there is virtually no research focusing on the score patterns of women of color on interest inventories.

C. Validity of the Tests

Given the demonstrated score differentials and the negative impact of these differentials on equal educational opportunities for girls and women, the final question is whether there is evidence that the tests are, nonetheless, being used in justifiable ways. That is, it must be determined whether the identified test uses are valid.

"Validity" as a concept in testing refers to whether a test actually accomplishes what it purports to do. Tests do not exist in a vacuum,
STANDARDIZED TESTING

neatly and conveniently measuring "true" abilities and aptitudes in a manner suitable for a wide range of applications. Rather they are designed — and must be justified — for specific purposes. According to the Standards for Educational and Psychological Testing, validity is "[t]he degree to which a certain inference from a test is appropriate or meaningful."

The standards elaborate further: "[V]alidity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. . . . The inferences regarding specific uses of a test are validated, not the test itself." As explained by the College Board:

Validity is . . . a concept that describes a correspondence. It examines the relation between what one would like to know, such as future performance in college, called a criterion, and what actually is measured, such as a predictive test score or high school grade point average (GPA) called a predictor.

A review of the literature shows that there are two basic factual issues concerning the validity of uses of tests showing gender and gender/race differentials in scoring. First, while there are validity studies supporting certain test uses, available analyses also suggest that certain tests predict differently by sex or race. Thus, a white male's score may predict a different outcome than the same score would for a white or minority female. Second, there are no studies or analyses to support the validity of many uses to which these tests are put.

1. Post-Secondary Admissions Tests

a. Tests Which Predict Differently By Gender

The SAT presents a prime example of the first problem, a test which predicts differently by gender. The test is validated as a predictor of first year college grades; however, the evidence shows that despite their lower SAT scores, females receive, overall, higher college grades than their SAT scores predict. The weight of authority supports the propo-

76 The concept of validity is well-established in the context of the law regarding employment testing. See discussion at notes 180-92 and accompanying text.
78 Id at 9.
79 College Board, Guide to the College Board Validity Study Service 5 (College Entrance Examination Board, 1988) (emphasis in original).
80 See, for example, Kenneth M. Wilson, A Review of Research on the Prediction of Academic Performance After the Freshman Year, College Board Report No 83-2, 1 (College Entrance Examination Board, 1983) ("Review of Research").
81 See, for example, Robert G. Cameron, The Common Yardstick: A Case for the SAT 17 (College Entrance Examination Board, 1989) ("Common Yardstick"); Clark and Grandy, Sex Differences in Academic Performance at 19 (cited in note 55). See also Kate Ruth Sheehan, The Relationship of Gender Bias and Standardized Tests to the Mathematics Competency of University Men and Women (Apr 1989) (unpublished doctoral dissertation) (on file at American University library and with the National Women's Law Center). Dr. Sheehan compared
tion that the test predicts differently for males and females:

Validity studies generally compare the admission test scores . . . of various groups with their first-year grade-point average. Such studies generally find women's test scores to be underpredictive of their performance and men's overpredictive. These studies also show women's test scores to be more strongly correlated with and more predictive of performance measures than men's.82

The College Board itself has acknowledged the predictive differences by gender and has recommended that they "can be eliminated by using separate prediction equations for each sex, rather than a single equation based on the total group."83 No studies have been identified regarding whether admissions officers follow this advice, and it is accordingly not known whether or to what extent females are penalized in admissions by the differential predictiveness of the SAT. Certainly, grave problems are presented by those institutions which use across-the-board SAT cutoff scores for determining admissions or for other purposes, such as the awarding of scholarships.

Differences in the predictive value of test scores also present serious problems for minority students. For example, studies have shown that

grades and SAT scores of incoming American University male and female freshmen as well as their subsequent academic performance. She found that entering females had significantly higher grade point averages and significantly lower SAT scores than their male peers. Id at 57. She also found that female American University students went on to receive significantly higher grades than their male classmates. Id at 72. Aware of the argument that higher female grades result from the fact that females take courses which tend to award higher grades, Dr. Sheehan went on to explore this question. She found that female students did tend to take different courses than male students but that there were no significant differences in grades among the varying fields of study. Id at 83. A senior thesis written by a Princeton University student reached a similar conclusion. It found that female members of the Princeton class of 1990 had slightly higher SAT scores on the verbal section and considerably lower math SAT scores than their male classmates. However, their average first-year grades were slightly higher than those of their male counterparts. SAT Gender Gap at 92 (cited in note 2).

82 Wilder and Powell, Sex Differences in Test Performance at 29-30 (cited in note 13). In addition, studies show that the predictive value of the SAT differs by the type of institution and program involved, which can also have an impact on the differential prediction by gender. For example, grades are less well predicted by the SAT where the institution is large, is a community college, has a diverse curriculum, or is an urban school enrolling many part-time, working, and commuting students. Grades are better predicted for students at four-year colleges, students at private colleges, students at high-cost institutions, and students who live in college-controlled residences. Leonard L. Baird, Predicting Predictability: The Influence of Student and Institutional Characteristics on the Production of Grades, College Board Report No 83-5 (College Entrance Examination Board, 1983). Certain of these characteristics, particularly the part-time/full-time dichotomy, have a major impact on female students who are disproportionately represented in the part-time category. However, institutions have shown little interest in the diminished levels of predictive validity for part-time students. Jerilee Grandy and Rosalea Courtney, A Look at Part-Time Undergraduates: Enrollment Trends, Admission Requirements, and Characteristics of Those Taking the SAT, College Board Report No 84-4 1, 2 (College Entrance Examination Board, 1984). See generally, Thomas F. Donlon, ed, The College Board Technical Handbook for the Scholastic Aptitude Tests and Achievement Tests (College Entrance Examination Board, 1984).

83 Cameron, Common Yardstick at 17 (cited in note 81). Indeed, the College Board's own researchers have observed that "the under-prediction of women's first-year college grades has been reported consistently in the research literature." Clark and Grandy, Sex Differences in Academic Performance at 21 (cited in note 55).
the predictive validity of SAT verbal and math scores of students with non-English-speaking backgrounds vary with the students' proficiency level in English. The academic performance of students with lower levels of English proficiency is likely to be underpredicted by their SAT scores, while high school rank or grade point average is a slightly more accurate predictor. Studies of the academic performance of Black students have similarly revealed that test scores exaggerate differences in performance between Black and white students, and that grades and non-standardized tests are much more accurate predictors of success among Black students. Again, there is little research focusing on the predictive validity of post-secondary admissions tests for minority females.

b. Unvalidated Test Uses

Of as much concern as the demonstrated gender differences in the SAT's prediction of first-year grades is the fact that there appear to have been no efforts to validate the SAT at all for numerous uses to which it is put. The Johns Hopkins Center for Academically Talented Youth, for example, points to no specific evidence to establish the SAT's validity in identifying junior high school age students with high math and/or science potential. The Army ROTC cadet command is not aware of any

84 Richard P. Duran, Testing of Linguistic Minorities in Robert L. Linn, ed, Educational Measurement 573, 582-83 (Am Council on Educ/MacMillan, 1989) (reviewing the body of literature documenting less accurate prediction of Hispanic students' college grades from test scores than from high school grades or class rank).

85 See, for example, National Commission on Testing, From Gatekeeper to Gateway at 13 (cited in note 1) (despite large test score differences between minorities and non-minorities, indicators of actual performance in education, such as grade point averages, do not show similarly large group differences); Timothy L. Walter, et al, Predicting the Academic Success of College Athletes, 58 Res Q for Exercise and Sport 273 (1987) (study of admission and graduation rates of football players with scholarships at the University of Michigan between 1974 and 1983 in relation to "Proposition 42"-like academic standards found SAT scores to be unrelated to college GPA for Blacks, and only weakly related for non-Blacks; while high school GPA correctly predicted success in 84% of cases, test scores accurately predicted success only 30% of the time).

86 The Johns Hopkins Center for Academically Talented Youth explains its use of the SAT by making the following claims: it is an objective evaluation of reasoning ability; it is a nationally recognized test administered across the United States under controlled conditions; it is economical in terms of time and cost; and it is difficult and thus identifies the upper limits of math and verbal activity. Undated Letter from John Chung, Research Coordinator of the Center for the Advancement of Academically Talented Youth, Johns Hopkins University to National Women's Law Center (received on or about Mar 1, 1990) (on file with the National Women's Law Center). The letter also states the opinion of the Johns Hopkins Office of the Study of Mathematically Precocious Youth that while the SAT "may be an underpredictor for females on the whole, it does not necessarily take away from those that have been identified as academically talented students." Id. Julian Stanley and Camilla Benbow, who pioneered the effort to identify gifted adolescents through high SAT scores, argue that the predictive value of the SAT for this population is demonstrated by the subsequent high achievement of the youngsters who have been identified. Julian C. Stanley and Camilla Persson Benbow, Youths Who Reason Exceptionally Well Mathematically in Robert J. Sternberg and Janet E. Davidson, eds, Conceptions of Giftedness 377 (Cambridge U Press, 1986). Even Stanley and Benbow implicitly acknowledge the highly circular nature of this reasoning by noting that "some students may be missed by this criterion." Id at 362.
study addressing the use of Scholastic Aptitude Test scores as a predictor of success in a military environment. The basis of National Merit’s reliance on PSAT and SAT scores is unclear — and thus difficult to evaluate — but insofar as the National Merit program is seeking to identify students who will excel throughout their college careers and in later life, the SAT’s and PSAT’s predictive values are limited at best. This problem is exacerbated by National Merit’s policy of imposing different cutoff scores by state. There is no evidence suggesting that the predictive value of the score is affected by a test-taker’s state of residence, but the different cutoff scores substantially affect scholarship awards.

2. Vocational Aptitude Tests and Interest Inventories

Vocational aptitude batteries appear to be used in several closely related ways in the secondary schools: to select students for vocational programs, to assist students in selecting vocational education programs, and to counsel students regarding possible vocations. A showing of validity for any of these uses would require evidence linking test performance with performance in specific vocational education programs and occupations.

The validity studies cited by the publishers of the aptitude tests, however, provide little, if any, evidence of such linkage in the secondary school context. Indeed, even the evidence supporting the ASVAB’s use as a classification mechanism for the military — the purpose for which it was designed — is weak. Moreover, the evidence regarding the ASVAB’s validity for military uses shows substantial variances in the predictive value of the test for different training programs and occupations, as well as variances based on the gender and race of the test-taker. For example, a study of the validity of particular composites for

---

88 See, for example, Wilson, Review of Research at 36 (cited in note 80) (concluding that the little evidence available regarding the predictive value of SAT scores for post-freshman grades suggests a gradual decline in predictive validity over time). See also Leonard L. Baird, The Role of Academic Ability in High-Level Accomplishment and General Success, College Board Report No 82-6 21, 24 (College Entrance Examination Board, 1982).
89 Diamond and Tittle, Sex Equity in Testing at 174 (cited in note 45).
90 Id.
91 See Kevin Murphy, Armed Services Vocational Aptitude Battery in Daniel J. Keyser and Richard C. Sweetland, eds, 3 Test Critiques 61, 68 (Test Corp of America, 1984) (“ASVAB”). The minimal correlation between ASVAB test scores and success in military occupations was demonstrated when a calibration error in 1976 resulted in the enlistment of more than 300,000 recruits who normally would have been rejected because of their low test scores. Several studies of the recruits admitted by mistake showed that many performed as well or better than those who received passing scores and that the performance of the remainder was only minimally below that of those not admitted by mistake. National Commission on Testing, From Gatekeeper to Gateway at 9-10 (cited in note 1). See also Bernard R. Gifford, The Political Economy of Testing and Opportunity Allocation, 59 J of Negro Educ 58, 64-65 (1990).
STANDARDIZED TESTING

predicting performance in Air Force training programs shows variations depending on the individual's race and sex.93 For all of the positions analyzed except one, the ASVAB composites were most predictive either for white males or white females. For many of the occupations, the composites had little predictive value at all for any group except white males. The correlations were particularly weak for Black females.94

Although it has been used in secondary schools since 1966, there is virtually no validity evidence supporting the use of the ASVAB in civilian student testing programs.95 Only three studies, all unpublished and all using an earlier version of the ASVAB, are referenced in the counselor's manual.96 A Freedom of Information Act request revealed that these reports were never published and cannot now be located by the Department of Defense.97 The request did produce three unpublished reports not cited in the Counselor's Manual.98 These studies, however, are preliminary and appear to be ongoing. In fact, the Department of Defense has conceded that there is no direct validity evidence for the student testing program.99

In the absence of such evidence, the Department of Defense relies on the theory of "validity generalization"100 to justify the test.101 Validity generalization theory generalizes validity evidence from one particular test or test use to another test or test use. This theory is controversial

---

94 Id. In fact, Black females were not even analyzed in two of the six job categories because they were too few in number. Id.
97 Letter from W.S. Sellman, Director, Accession Policy, Office of the Assistant Secretary of Defense, to Katherine Connor, National Women's Law Center (Sep 6, 1990) (on file with the National Women's Law Center).
99 John Welsh, Review of the National Women's Law Center Report on the Legal Implications of Gender Bias in Standardized Tests 2 (Nov 27, 1991) (unpublished review on file with the National Women's Law Center) ("Welsh Review"). Welsh writes that the Department of Defense "has attempted to obtain such evidence since 1985, but restraints imposed by the Office of Management and Budget have prohibited the direct collection of civilian validity information." According to Welsh, validity studies are now under way and results will be available within the next two years. Currently, there is no way to evaluate these studies (which presumably are those referenced in response to the Freedom of Information Act request discussed in note 97 and accompanying text) because they are not yet completed.
100 Dept. of Defense, Technical Supplement to the Counselor's Manual at 48-49 (cited in note 41).
and is by no means an adequate substitute for direct validity evidence or, at least, a rigorous analysis demonstrating the comparability of skills, abilities, and/or interests predicted by the respective test or test use.\textsuperscript{102} Yet despite the inadequacy of generalization theory, the Department of Defense invokes the theory in two ways. First, it generalizes the validity evidence accumulated in support of the General Aptitude Test Battery (GATB) to validate the ASVAB.\textsuperscript{103} Secondly, it generalizes the validity evidence accumulated in support of the ASVAB’s use to classify military recruits in order to support the civilian uses of the ASVAB.\textsuperscript{104}

Validity evidence offered in support of the ASVAB’s use as a military classification device is weak.\textsuperscript{105} Similarly, validity evidence supporting even the intended uses of the GATB is modest and must be viewed in light of the test’s well-established disparate impact on minorities.\textsuperscript{106} Studies of the GATB, which has been used throughout the United States Employment Service for job referral, have revealed that because minorities score much lower than whites on the test, selection errors — rejection of applicants who could perform the job successfully — “weigh more heavily on minority workers than on majority workers.”\textsuperscript{107}

To avoid this adverse impact, the Department of Labor instituted a policy for certain adaptations of the test under which “minority applicants were referred to employers in proportion to their relative numbers or ratio to nonminorities in the local office applicant pool.”\textsuperscript{108} Elsewhere, the same principle was incorporated through the method of within-group scoring.\textsuperscript{109} Also known as “race-norming,” this policy became quite controversial. In reaction to the controversy, and because of a decline in GATB validities in recent studies as well as lower validities for Blacks, on July 24, 1990, the Department of Labor put forth a proposal to suspend use of the GATB and to conduct a two-year study to address the underlying problems of limited validity and adverse

\textsuperscript{102} For example, see generally Richard T. Seymour, \textit{Why Plaintiffs’ Counsel Challenge Tests, and How They Can Successfully Challenge the Theory of “Validity Generalization,”} 33 J Vocational Behav 331, 350-63 (1988). See also \textit{EEOC v Atlas Paper Box Co.}, 868 F2d 1487, 1490 (6th Cir 1989) (“[t]he validity of the generalization theory utilized by Atlas . . . is not appropriate. Linkage or similarity of jobs in dispute in this case must be shown by such on-site investigation to justify application of such a theory”).

\textsuperscript{103} \textit{Welsh Review} at 2-3 (cited in note 99); Dept. of Defense, \textit{Technical Supplement to the Counselor’s Manual} at 49 (cited in note 41).

\textsuperscript{104} \textit{Welsh Review} at 2-3 (cited in note 99); Dept. of Defense, \textit{Technical Supplement to Counselor’s Manual} at 48-49 (cited in note 41).

\textsuperscript{105} See note 91 and accompanying text.


\textsuperscript{107} Hartigan and Wigdor, \textit{Fairness in Employment Testing} at 7 (cited in note 92).

\textsuperscript{108} 55 Fed Reg 30162 (cited in note 106).

\textsuperscript{109} Id.
impact.\textsuperscript{110}

The Department of Labor also proposed to discontinue its use of other test batteries which are derived from and/or base their validity on the GATB.\textsuperscript{111} According to the Department of Defense, the civilian use of the ASVAB falls into this category.\textsuperscript{112} Any Department of Labor-related use of the ASVAB would have thus been terminated under the proposed policy.

This proposal was never implemented. Instead, on December 13, 1991, the Department of Labor announced that, in response to a provision in the Civil Rights Act of 1991,\textsuperscript{113} it would end the practice of within-group scoring.\textsuperscript{114} At the same time, it announced that it would institute a multi-year study of the GATB to improve its performance but would continue to permit the use of the test.\textsuperscript{115} The 1991 announcement confirmed that studies have demonstrated both that the relationship between GATB scores and job performance is "modest" and that there has been a drop in validities in recent years.\textsuperscript{116} Nonetheless, the Department noted many comments from a variety of test-users opposing its earlier proposal to discontinue use of the GATB pending a study and concluded that it would not prohibit use of the GATB while the research is being conducted. At the same time, however, in an implicit acknowledgment of the GATB's disparate impact, it cautioned test-users that they are responsible for complying with applicable laws, including civil rights laws.\textsuperscript{117} While there are substantial concerns regarding the continued use of the GATB pending the outcome of the study, the Department of Labor remains on record as recognizing that the GATB is a flawed test.

\section{Failure to Measure Aptitude}

Even if there were strong validity evidence in support of the military use of the ASVAB or the use of the GATB generally, the fundamental problem presented by the use of the ASVAB in the student testing program remains: the tests fail to distinguish among different aptitudes in different areas and hence cannot be helpful in counseling students about different career possibilities. All of the academic and occupational composites of the ASVAB essentially measure the same thing — general aca-

\textsuperscript{110} Id.
\textsuperscript{111} Id at 30163.
\textsuperscript{112} Welsh Review at 2-3 (cited in note 99).
\textsuperscript{114} 56 Fed Reg 65746 (cited in note 106).
\textsuperscript{115} Id.
\textsuperscript{116} Id at 65747.
\textsuperscript{117} Id.
demic ability.118 The average correlation coefficient119 between each of the composites and the academic ability composite is .95; each composite expresses almost exactly the same information about a test-taker. Thus, a student who does well on the academic ability composite will also score high on all of the occupational composites (mechanical and crafts, business and clerical, electronics and electrical, and health, social, and technology). Conversely, a student with a low academic ability composite score is also likely to score low on all of the occupational composites. This flaw is fatal. As one reviewer noted, "[t]he acid test of a test battery is its ability to provide information about several distinct abilities. The ASVAB fails this test."120 In order for counselors to rely on test scores for vocational counseling, the test must be capable of distinguishing between different aptitudes in different subject areas.121 Because the ASVAB only measures general academic ability, it is, by definition, not competent to perform this function.

The DAT has similar problems. Test reviewers have repeatedly noted that, like the ASVAB, the DAT lacks the very ability to differentiate between different aptitudes that would support its valid use for counseling purposes.122 Yet the DAT score report explicitly steers students toward and away from particular occupations and classes based on scores on different subtests and a career planning questionnaire completed by the student.123 Moreover, test reviewers have found little evidence of the

118 Murphy, *ASVAB* at 65 (cited in note 91).
119 A correlation coefficient denotes a relationship between measures. David Nachmias and Chava Nachmias, *Research Methods in the Social Sciences* 142 (St. Martin's Press, 2d ed 1981). A positive number denotes a positive relationship, with 1.0 indicating a perfect predictive relationship between the measures and a negative number indicating that one measure does not predict the other. See discussion at notes 219-21 and accompanying text.
120 Murphy, *ASVAB* at 68 (cited in note 91).
121 Anne Anastasi, *Psychological Testing* 378-79 (MacMillan, 5th ed 1982). It appears that the Department of Defense has also concluded that the ASVAB has little ability to discriminate between different aptitudes in different areas. Beginning in July of 1992 the ASVAB workbook will explicitly correlate civilian occupations with ASVAB scores on academic composites only; the occupational composites will not be correlated with occupations. Interview with Anita R. Lancaster, Assistant Director, Office of the Assistant Secretary of Defense and Katherine Connor, National Women's Law Center (Apr 4, 1990); see also *Welsh Review* (cited in note 99).
123 For example, if a student indicates an interest in the "engineering and applied science" occupation group and it correlates with her other interests and educational plans, but not with her score on the mechanical reasoning subtest, her score report will state:

People who do well in this field [of] work usually like the school subjects and activities you like. Also, the kind of education they have matches your plans for school. However, their scores on the aptitude tests that are related to this field are often higher than yours. In view of this, you may wish to reconsider this occupational choice and look into other fields of work that would be more suited to your particular abilities.

DAT’s validity for predicting job success. As for success in school, the validity studies do show correlations between the academic subtests and grades in academic high school courses. However, vocational education classes are lumped into the “miscellaneous courses” category and the publishers are unable to draw “firm conclusions” about the predictive validity of the DAT for these courses.

Perhaps most troubling is the fact that the predictive validities of both the DAT and the ASVAB have never even been explored for students of color in the secondary school setting. The significant score differences by gender and race on the ASVAB make it all the more important to carefully analyze predictive validity for this population.

b. Interest Inventory Validation Problems

Interest inventories also present significant validity problems. Foremost is that the link between interests and abilities has never been established. Thus, for example, although a person may have interests similar to those of a lawyer, this does not necessarily mean that he or she has the ability to succeed in this profession. Conversely, a person who does not have interests similar to lawyers currently in the profession may still have the ability to be a successful lawyer. Despite this lack of a demonstrated correlation between interests and abilities, interest inventories are used heavily in the guidance counseling process.

In addition, a fundamental issue for women, and particularly women of color, is whether the inventories simply perpetuate a segregated status quo in vocational education and the workplace, or whether they actually expand opportunities. Because women tend to underrate their abilities in nontraditional tasks and occupations, interest inventories can close doors to nontraditional occupations rather than open them. Moreover, since many inventories are based on the interest profiles of persons presently in the workforce, the tests may perpetuate existing gender and racial segregation by suggesting that interests similar to those of men are necessary for success in male-dominated occupations. Based on the concern about the negative effects of interest inventories, different models for career exploration have been suggested. Some researchers maintain that an interest test can and should suggest an expanded range

124 See, for example, Sanders, Review of Differential Aptitude Tests at 506 (cited in note 122); Anastasi, Psychological Testing at 378 (cited in note 121); and Ronald K. Hambleton, Review of Differential Aptitude Tests, in Mitchell, 1 Ninth Mental Measurement Yearbook 505 (U of Neb, 1985).
126 Id. at 34-35.
127 DAT scores for minority students are not made available by the test publisher. Telephone interview with Patty McDivitt (cited in note 44).
128 Walsh and Betz, Tests and Assessment at 230 (cited in note 46).
of options and not simply reflect socialized experiences and the patterns of a segregated workforce.\textsuperscript{129}

These issues are particularly acute for minority women. Although the National Institute of Education Guidelines recommended that inventories be normed for different racial/ethnic groups and that separate interpretive materials should be developed if there are differences in the way that groups respond,\textsuperscript{130} this generally has not been done, despite the fact that preliminary investigation has revealed the likelihood of a mismatch between the interest structures of minority women and those of the norming groups used by the test publishers.\textsuperscript{131} Thus, minority women may experience both gender and racial bias on existing interest inventories and further restriction of educational and employment opportunities.

Despite the wide-ranging use of vocational education tests and interest inventories for counseling students about careers and placing them in particular vocational programs, there has been little attempt to validate the tests with success in specific occupations or vocational education programs.

D. The Factual Context: Conclusions

In sum, a review of the empirical literature establishes a clear, factual predicate for the consideration of the legal framework in which to understand gender-in-testing issues. First, a wide range of standardized tests reflect gender differentials in scoring which work to the concrete disadvantage of females in general and females of color in particular. Second, the causes of these differentials are not fully understood and are not explained by the literature. Indeed, there is barely any analysis at all of the reasons underlying the particularly low scores of girls and women of color. Third, serious validity concerns are raised in connection with many uses of these tests. These range from the uses of tests which predict differently by gender and/or a combination of gender and race to test uses for which no meaningful validity evidence exists at all.

III. The Applicable Legal Analysis

Federal prohibitions against gender discrimination in education are principally based in Title IX of the Education Amendments of 1972 and the Equal Protection Clause of the Fourteenth Amendment to the

\textsuperscript{129} Diamond and Tittle, \textit{Sex Equity in Testing} at 181 (cited in note 45).

\textsuperscript{130} See Carol Tittle and Donald Zytowski, \textit{Sex Fair Interest Measurement: Research and Implications} (National Institute of Education, 1978).

United States Constitution. In addition, state equal rights amendments, other state constitutional provisions, and federal and state statutes and regulations address these issues. Both on its own and by incorporating closely related principles which have been developed to limit race and gender discriminatory testing in employment, this jurisprudence provides the framework for substantially limiting, if not eliminating altogether, the gender discriminatory educational use of standardized tests. This section will explore the nature and applicability of these legal theories, including their remedial schemes.

A. Title IX

Title IX of the Education Amendments of 1972 (hereinafter “Title IX”) prohibits sex discrimination in any educational program or activity receiving federal financial assistance. It provides: “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 . . ..” Title IX’s broad prohibition against sex discrimination in education clearly encompasses many gender-discriminatory test uses. We will first review Title IX’s relevant substantive prohibitions and then turn to the key question of its applicability both to intentional discrimination and to discrimination which results from policies or practices which are neutral on their face but have a “disparate impact” based on gender. Further, because much discrimination in testing falls into this second category, we will consider the nature of the disparate impact analysis to be applied. Finally, we will discuss the potential of a combined Title IX/

---

132 20 USC §§ 1681-87 (1988), as amended by the Civil Rights Restoration Act of 1987, Pub L No 100-259, 102 Stat 28 (1988), codified at 20 USC § 1687. Following the passage of the Civil Rights Restoration Act in 1988, it is clear that Title IX applies to all education programs and activities conducted by recipients of federal financial assistance. See 20 USC § 1681 note, 1687, 1687 note, 1688, 1688 note. This coverage results regardless of whether the federal funds support the particular program or activity at issue. Title IX’s prohibitions thus apply to all public institutions at the elementary, secondary, and post-secondary level, including all private post-secondary institutions whose students receive any federal financial aid and all private institutions which receive any other form of federal financial assistance. In addition, many other institutions in the “education business” receive federal funds and are also covered by Title IX insofar as their education programs and activities are concerned. Examples of such institutions include the Educational Testing Service and the College Board, both of which receive substantial federal financial support. See, for example, letter from Renee Chilton, Grants and Contracts Service, US Dept. of Educ., to Ellen J. Vargyas, National Women’s Law Center, and supporting US Dept. of Educ. reports in response to Ms. Vargyas’ Freedom of Information Act request on this subject (May 18, 1989) (letter and reports on file with the National Women’s Law Center).

133 20 USC § 1681(a). Based on an extensive hearing record, Congress intended to enact “a strong and comprehensive measure [that would] provide women with solid legal protection from the persistent, pernicious discrimination which is serving to perpetuate second-class citizenship for American women.” 118 Cong Rec 5804 (1972) (Remarks of Sen Birch Bayh, D-Ind). See also Discrimination Against Women: Hearings on HR 106098 before the Special Subcommittee on Education of the House Committee on Education and Labor on HR 16098, 91st Cong, 2d Sess 1-2 (1970).
Title VI analysis to address the multiple nature of the discriminations suffered by women of color.

1. The Title IX Prohibition Against Gender Discriminatory Test Uses

While Title IX does not directly address testing and test use in its statutory language, its regulations explicitly prohibit the discriminatory use of tests in admissions, employment, and counseling and appraisal. Regarding admissions, recipients may not:

(2) . . . 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.

This section mirrors the regulation prohibiting the use of discriminatory tests in the employment practices of recipients. The counseling regulation is similarly broad. As originally proposed, the regulation only prohibited the use of different counseling materials on the basis of sex, or materials which permitted differential treatment on the basis of sex. However, the final regulation goes beyond facial differential treatment in materials and requires that schools reexamine their counseling practices

---

134 The agencies charged with enforcing Title IX have broad discretion in crafting the applicable regulatory framework. North Haven Bd. of Educ. v Bell, 456 US 512, 514 (1982).

135 Title IX applies to the overall admissions practices of the following educational institutions: institutions of vocational education, professional education, graduate higher education, and public institutions of undergraduate higher education. It does not cover the general admissions practices of elementary and secondary schools, private undergraduate institutions, or public institutions of undergraduate education which have traditionally and continually from their establishment had a policy of admitting only students of one sex. 20 USC § 1681(a)(1) and (5). But see 34 CFR § 106.35 (1991) (providing that a local educational agency shall not deny admissions on the basis of sex to any school or educational unit it operates "unless such recipient otherwise makes available to such person, pursuant to the same policies and criteria of admission, courses, services and facilities comparable to each course, service and facility offered in or through such schools"). Further, even where the general admissions exception may apply, the institution is covered by Title IX with regard to all of its other education programs or activities, including access to those programs or activities. Thus, if a secondary school uses a test for admissions to a particular course and the test has a disproportionate impact on the basis of gender, Title IX's prohibitions are fully implicated.

136 34 CFR § 106.52.

137 34 CFR § 106.36.

138 34 CFR § 106.21. This regulation applies only to admissions practices which are covered by Title IX.

139 34 CFR § 106.52. The comments accompanying the final version of 34 CFR § 106.21(2), which was slightly changed from the originally proposed version, explain that it is intended to "conform the provisions of the regulations dealing with students and those dealing with employees." Sex Discrimination Regulations, Hearings Before the Subcommittee on Postsecondary Education of the Committee on Education and Labor, 94th Cong, 1st Sess 14 (1975).

140 34 CFR § 106.36.

141 Access to Education Program or Activity, 39 Fed Reg 22235, § 86.34(c) (1974).
whenever there is disproportionate enrollment by sex in a particular class or program. It provides that:

Where a recipient finds that a particular class contains a substantially disproportionate number of individuals of one sex, the recipient shall take such action as is necessary to assure itself that such disproportion is not the result of discrimination on the basis of sex in counseling or appraisal materials or by counselors.\footnote{34 CFR § 106.36(c).}

Thus, whenever there is a gender-disproportionate enrollment in classes, recipients have an affirmative obligation to ensure that discrimination is not occurring in any aspect of the counseling process.\footnote{See also 34 CFR § 106.36(b) (recipients must ensure that counseling and appraisal materials, including tests, do not discriminate on the basis of sex).}

Other regulations, while they do not expressly address testing, prohibit discrimination in areas where test results are often used. For example, recipients are prohibited from discriminating on the basis of sex in the provision of financial aid or from assisting, in any fashion, any individual or organization which provides financial aid to any of the recipient’s students in a manner which discriminates on the basis of sex.\footnote{34 CFR § 106.37(a).}

While testing is not specifically referenced, the discriminatory use of tests in awarding scholarships, such as the use of the SAT in awarding National Merit Scholarships, is surely encompassed. Moreover, without limitation to financial aid programs, recipients are prohibited from “providing significant assistance to any agency, organization or person which discriminates on the basis of sex in providing any aid, benefit or service to students or employees.”\footnote{34 CFR § 106.31(b)(6). In Iron Arrow Honor Soc. v Heckler, 702 F2d 549, 555 (5th Cir 1983), the former Fifth Circuit upheld this section of the regulations both on its face and as applied to the relationship between the University of Miami and the Iron Arrow Honor Society which had an exclusively male membership. See Iron Arrow Honor Soc. v Hufstedler, 499 F Supp 496 (SD Fla 1980), aff’d 652 F2d 445 (5th Cir 1981), vacated and remanded for further consideration in light of North Haven Board of Educ. v Bell, 456 US 512 (1982), modified sub nom Iron Arrow Honor Soc. v Heckler, 702 F2d 549 (5th Cir 1983), dismissed as moot, 464 US 67 (1983).}

This principle bars, for example, recipients from assisting private programs for gifted and talented adolescents or private vocational education programs which use test scores in a gender-discriminatory fashion.

In addition to the regulations, the Department of Education also has promulgated “Guidelines for Eliminating Discrimination and Denial of Services on the Basis of Race, Color, National Origin, Sex, and Handicap in Vocational Education Programs” ("Vocational Education Guidelines").\footnote{34 CFR Part 100, App B (1991). Guidelines do not have the full authority of regulations but, nonetheless, are entitled to deference. Griggs v Duke Power Co., 401 US 424, 433-34 (1971).} These Vocational Education Guidelines adopt the standard incorporated in 34 CFR § 106.36(c) (1990), providing that where a vocational education program disproportionately enrolls members of one gender, recipients must assure themselves that this phenomenon is not the
result of unlawful discrimination in counseling activities. Furthermore, the Vocational Education Guidelines require that when a test used in connection with a vocational education program has a disproportionate impact on the basis of gender, a recipient must show that the test use is “validated as essential to participation in a given program” and that alternative criteria with a lesser discriminatory impact are unavailable.

In sum, Title IX applies across the board to prohibit the gender-discriminatory use of tests in federally assisted education except in limited, enumerated circumstances in the area of admissions. This prohibition includes virtually all test uses undertaken in connection with vocational and scholarship programs, assessment and placement decisions, covered admissions practices, or any other purpose.

2. Title IX and Its Regulations Reach Both Intentional and Disparate Impact Discrimination

The great weight of authority supports the conclusion that Title IX prohibits both intentional and disparate impact discrimination. Disparate impact discrimination refers to practices which, although neutral on their face and not intentionally discriminatory, are discriminatory in effect. The availability of disparate impact analysis under Title IX is particularly important in light of the fact that testing discrimination typically results from the discriminatory application of facially neutral practices rather than acts of intentional discrimination.

Because Title IX was expressly modeled on Title VI of the Civil Rights Act of 1964, the analysis of Title IX's scope is initially based in Title VI law. When Title IX was enacted in 1972, Title VI had explicitly been interpreted by federal agencies to reach disparate impact discrimination. Model regulations had been drafted, “and every Cabinet Department and about 40 federal agencies had adopted standards in which Title VI was interpreted to bar programs with a discriminatory

---

147 34 CFR Part 100, App B(V)(B).
148 34 CFR, Part 100, App B(IV)(K). The Vocational Education Guidelines set out examples of practices which must meet this test. These include “past academic performance, record or disciplinary infractions, counselors’ approval, teachers’ recommendations, interest inventories, high school diplomas and standardized tests, such as the Test of Adult Basic Education (TABE).”
149 See 34 CFR § 106.31 (broad prohibition against gender discrimination in education).
150 An exception is the separate norming of test scores by gender. See discussion at notes 356-59 and accompanying text. However, because Title IX explicitly permits “a recipient [to] take affirmative action to overcome the effects of conditions which resulted in limited participation [in an activity] by persons of a particular sex . . . “, 34 CFR § 106.3(b), separate norming may be permissible under the law if it comports with this purpose. See also 34 CFR § 106.3(a) (Assistant Secretary may order remedial action upon finding of discrimination).
152 Cannon v U of Chicago, 441 US 677, 695-96 (1979) (“[t]he drafters of Title IX explicitly assumed that [Title IX] would be interpreted and applied as Title VI had been during the preceding eight years”.

---
Indeed, the House of Representatives rejected a proposed amendment in 1966 that would have limited Title VI's coverage to intentional discrimination.\textsuperscript{154}

In \textit{Guardians Association v Civil Service Commission},\textsuperscript{155} a Title VI disparate impact challenge to examinations administered by the New York City Police Department, the United States Supreme Court confirmed that although Title VI itself reaches only intentional discrimination, Title VI regulations properly reach disparate impact discrimination. Therefore, the Court held that in actions brought under Title VI with reference to its regulations, a showing of intentional discrimination is not required.\textsuperscript{156} The Supreme Court reaffirmed this holding in \textit{Alexander v Choate},\textsuperscript{157} a case brought under § 504 of the Rehabilitation Act.\textsuperscript{158} In that case, a unanimous Court reiterated the framework for Title VI claims that resulted from the multiple opinions in \textit{Guardians}:

First, the \textit{Guardians} Court held that Title VI itself directly reached only instances of intentional discrimination. Second, the Court held that actions having an unjustifiable disparate impact on minorities could be redressed through agency regulations designed to implement the purposes of Title VI. In essence, then, we held that Title VI had delegated to the agencies in the first instance the complex determination of what sorts of disparate impacts upon minorities constituted sufficiently significant social problems, and were readily enough remediable, to warrant altering the practices of the federal grantees that had produced those impacts.\textsuperscript{159}

At least six circuits have explicitly recognized that a cause of action premised on the Title VI regulations does not require proof of discriminatory intent.\textsuperscript{160} Further, Title IX decisions reaching this question post-\textit{Guardians} have confirmed that Title IX regulations properly reach dispa-

\textsuperscript{154} Id.
\textsuperscript{155} 463 US 582 (1983).
\textsuperscript{156} Five separate opinions were issued in \textit{Guardians}. Justices Stevens, Brennan, and Blackmun concluded that although a violation of the statute requires proof of discriminatory intent, the regulations promulgated under the statute incorporate an effects standard. 463 US at 608 n 1. Justices White and Marshall took the view that no Title VI claim, either under the statute or under the regulations, requires a showing of intent. Id. Chief Justice Burger and Justices Powell, Rehnquist, and O'Connor reasoned that intentional discrimination is an element of any valid Title VI claim. Id.
\textsuperscript{157} 469 US 287 (1985).
\textsuperscript{158} 29 USC § 794 (1988). Section 504 of the Rehabilitation Act was also expressly modeled in part after Title VI. \textit{Alexander}, 469 US at 293 n 7.
\textsuperscript{159} 469 US at 293-94 (footnotes omitted). Based on an analogy to Title VI, the \textit{Alexander} court assumed, "without deciding[.] that Section 504 reaches at least some conduct that has an unjustifiable disparate impact upon the handicapped." Id at 299.
rate impact discrimination as well.\textsuperscript{161} Most recently, in \textit{Sharif v New York State Education Department}, the plaintiffs asserted that New York State's use of the SAT to determine eligibility for state merit scholarships had an unlawful disparate effect on females.\textsuperscript{162} The district court granted plaintiffs' motion for a preliminary injunction, holding that the Title IX implementing regulations in general prohibit practices with a discriminatory effect on one sex, and that plaintiffs need not prove intentional discrimination.\textsuperscript{163}

Thus, while Title IX's regulations may properly reach disparate impact discrimination, there is strong support for the argument that Title IX itself reaches such discrimination as well. The conclusion in \textit{Guardians} that Title VI's statutory prohibition is limited to intentional discrimination was based on the finding in \textit{University of California Regents v Bakke} that Title VI's prohibition is coextensive with that in the Fourteenth Amendment.\textsuperscript{164} The Fourteenth Amendment, of course, requires a showing of intent to establish a violation.\textsuperscript{165} However, there has never been a similar finding that Title IX incorporates constitutional standards and, in fact, nothing in Title IX's history suggests that it was designed as a statutory reflection of the Fourteenth Amendment's prohibition against sex discrimination. Indeed, when Title IX was enacted in 1972, the Supreme Court had just begun the lengthy process of dismantling the rational-basis analysis of sex discrimination cases and developing a heightened standard of review under equal protection for cases of gender discrimination.\textsuperscript{166} Given the absence of any express legislative history to the effect that Congress was seeking to reflect the contemporaneous and highly ambiguous constitutional standard for gender discrimination in its enactment of Title IX, such a conclusion is unlikely.

Moreover, the recent passage of the Civil Rights Restoration Act,
over a presidential veto,\textsuperscript{167} along with the enactment of the Civil Rights Remedies Equalization Act,\textsuperscript{168} further illustrates Congress' intent that Title IX should be broadly interpreted. Such an interpretation would include its application to disparate impact discrimination.

But regardless of whether Title IX itself reaches disparate impact discrimination, \textit{Guardians} and \textit{Alexander} clearly teach that its regulations may. As is apparent from the preceding discussion, Title IX's regulations, by their plain language, do reach disparate impact cases.\textsuperscript{169} Moreover, courts have broadly construed Title IX's regulations in line with their intent to eliminate discrimination. They have assured that disparate impact analysis is available under the Title IX regulatory scheme generally and is not limited to a cramped and narrow construction of the regulatory language.\textsuperscript{170}

\section{Proving Disparate Impact Discrimination in Testing Cases: The Title VII Analogy}

Since the Title IX regulations, and most likely Title IX itself, prohibit disparate impact discrimination without requiring proof of intent, the next issue is the standard for proving disparate impact discrimination under Title IX. Title VI disparate impact analysis, including that developed in Title VI testing cases, has principally relied on the law under Title VII, the section of the Civil Rights Act of 1964 prohibiting employ-

\textsuperscript{167} The Restoration Act reversed the Supreme Court's 1984 decision in \textit{Grove City College v Bell}, 465 US 555 (1984), restoring the effectiveness of four major civil rights statutes prohibiting discrimination by recipients of federal funds: Title IX, Title VI, Section 504, and the Age Discrimination Act. Specifically regarding educational institutions, the statute provides that where federal aid is extended anywhere within an institution, the entire institution is covered, and not just the specific program receiving assistance. Civil Rights Restoration Act of 1987, Pub L 100-259, 102 Stat 28 (1988), codified at 20 USC § 1687. In passing this legislation, Congress stressed that Title IX and the other three civil rights statutes are to be given the broadest interpretation in order to eliminate discrimination from institutions receiving federal financial assistance. Civil Rights Restoration Act of 1987, S Rep No 64, 100th Cong, 1st Sess 5 (1987).

\textsuperscript{168} The Civil Rights Remedies Equalization Act Amendment, Pub L 99-506, 100 Stat 1845 (1986), codified at 42 USC § 2000d-7 (Supp 1990), reversed the Supreme Court's decision in \textit{Atascadero State Hospital v Scanlon}, 473 US 234 (1985). In \textit{Atascadero}, the Court held that the Eleventh Amendment bars suits against states and state agencies in federal court for monetary relief under Section 504 of the Rehabilitation Act. Congress responded by promptly passing the Civil Rights Remedies Equalization Act Amendment, which expressly abrogates the Eleventh Amendment immunity of the states for violations of any federal statute prohibiting discrimination by recipients of federal financial assistance, including Title IX.

\textsuperscript{169} For example, the admissions regulation, 34 CFR § 106.21(b)(2), and the employment regulation, 34 CFR § 106.52, address tests which have a "disproportionately adverse effect on persons on the basis of sex" (emphasis added). See also 34 CFR § 106.36(b) and (c) (counseling appraisals and materials); 34 CFR Part 100, App B(IV)(K) (Guidelines for Eliminating Discrimination in Vocational Education).

\textsuperscript{170} See, for example, \textit{Sharif}, 709 F Supp at 361 (Title IX regulations generally reach disparate impact discrimination); and \textit{Haffer v Temple U}, 678 F Supp 517, 539-40 (ED Pa 1987) (disparate impact analysis available in claim of discriminatory award of athletic scholarships although applicable regulation does not use explicit disparate impact language).
ment discrimination. As discussed above, Title IX was modeled on Title VI and is typically interpreted and applied in a similar fashion. Moreover, the Title IX disparate impact testing case, Sharif, specifically incorporated a Title VII analysis.

At least one court has questioned the wisdom of applying the more limited Title VII framework for employment cases to the educational testing arena, suggesting that stronger prohibitions against discrimination would be appropriate. In Larry P. v Riles, the district court judge followed the Title VII standards of proof in evaluating a Title VI claim, but expressed the following reservations:

If tests can predict that a person is going to be a poor employee, the employer can legitimately deny that person a job, but if tests suggest that a young child is probably going to be a poor student, the school cannot on that basis alone deny that child the opportunity to improve and develop the academic skills necessary to success in our society.

In affirming this part of the decision in Larry P., the Ninth Circuit again noted that "the employment context is quite different from the educational situation." Thus, although Title VII testing law is certainly relevant to interpretation of Title VI and Title IX, it may not always be controlling because of the very different purposes underlying the use of tests in the workplace and in the school. These different purposes may suggest, in particular circumstances, a broader interpretation of Title IX than is offered by Title VII in order to prevent students from being improperly excluded from valuable educational opportunities. Indeed, in certain respects the Title IX regulatory scheme does provide more protections than are available under Title VII.

Nonetheless, because Title VII provides guidance for Title IX disparate impact analysis, we will turn to an examination of Title VII disparate impact law and its implications for discriminatory test uses in education. There is, however, a threshold issue which must first be addressed regarding the nature of the appropriate Title VII analogy to be drawn. This issue arises from the fact that the standard for establishing disparate impact discrimination under Title VII, initially established in 1971 in the unanimous Supreme Court decision in Griggs v Duke Power Co., has gone through two fundamental alterations since that time.

---

171 42 USC § 2000e (1981). See, for example, Latinos Unidos de Chelsea v Secretary of Housing, 799 F2d at 785-86; Castaneda v Pickard, 781 F2d at 465-66; Branches of NAACP v State of Ga., 775 F2d at 1417; and Larry P., 793 F2d at 982, 982 n 9.
172 Cannon, 441 US at 694-96.
173 Sharif, 709 F Supp at 361. See also Mabry, 813 F2d at 316 n 6, 317 (Title VII is the most appropriate analogue for defining the substantive standards for a Title IX claim of sex discrimination in employment); and Lipsett v U of Puerto Rico, 864 F2d 881, 896-97 (1st Cir 1988) (Title VII standard for proving sexual harassment applies to Title IX claim).
174 495 F Supp 926, 969 (ND Cal 1979), aff'd in part, rev'd in part, 793 F2d 969 (9th Cir 1984).
175 793 F2d at 980.
176 See notes 236-41 and accompanying text; see also note 248.
Initially, the standard was made significantly more difficult to meet in the Supreme Court's 1989 decision in *Wards Cove Packing Co. v Atonio*,\(^{178}\) which effectively reversed *Griggs*. Then, after a two-and-one-half year legislative battle, Congress rejected the decision in *Wards Cove* and codified a *Griggs*-based disparate impact cause of action as part of the Civil Rights Act of 1991.\(^{179}\) However, as is inevitable with a new statute, particularly one reflecting as complicated and contentious a legislative history as that of the Civil Rights Act of 1991, questions remain regarding its interpretation. These questions, which must now be resolved through the courts, include the precise state of the Title VII disparate impact standard.

While Title IX analysis has always reflected certain Title VII principles, the relevant Title IX interpretations, including both case law and administrative regulations, predate both the Civil Rights Act of 1991 and *Wards Cove* and are based squarely on *Griggs*. However, an unanswered question remains: namely, (1) whether Title IX disparate impact analysis is linked to disparate impact analysis under Title VII regardless of how Title VII law develops; or (2) whether Title IX disparate impact analysis has been based on Title VII because Title IX incorporates the doctrine of *Griggs* and should accordingly continue to follow the *Griggs* standard? Because, as shown below, the Title IX regulations closely track the *Griggs* formulation, deviating only to place additional burdens on the test-user, the better view is that Title IX incorporates, at a minimum, the *Griggs* standard for establishing disparate impact discrimination, regardless of subsequent changes in Title VII jurisprudence. Accordingly, the following discussion will focus on the *Griggs* line of cases and its application to gender discrimination in educational testing. However, because the Civil Rights Act of 1991 is so closely linked to the matters at hand — and because it confirms the view that Congress intended all along for courts to use the *Griggs* analysis in reviewing disparate impact discrimination — we will also address its implications for Title IX. The discussion will be informed throughout by an examination of the Title VI cases, as well as of *Sharif*, which have applied these principles to the discriminatory use of tests in education.

### a. *Griggs v Duke Power Co.* and Its Progeny

In 1971, the year before Title IX was enacted, the Supreme Court held unanimously in *Griggs v Duke Power Co.*, that Title VII prohibits not only intentional discrimination but also facially neutral practices.

---

which are discriminatory in impact.\textsuperscript{180} At issue in \textit{Griggs} was the employer's requirement that, in order to be hired, job applicants must either have a high school diploma or a passing score on one of two aptitude tests. The policy disproportionately excluded Blacks from employment and the employer produced no evidence that either the diploma requirement or the tests were related to the jobs at issue. In holding that Title VII prohibited these employment practices, the Court explained that Title VII "proscribes not only overt discrimination but also practices that are fair in form but discriminatory in operation. The touchstone is business necessity. If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited."\textsuperscript{181} Regarding the specific context of testing, the Court continued:

Nothing in the Act precludes the use of testing or measurement procedures; obviously they are useful. What Congress has forbidden is giving these devices and mechanisms controlling force unless they are demonstrably a reasonable measure of job performance. . . . What Congress has commanded is that any tests used must measure the person for the job and not the person in the abstract.\textsuperscript{182}

The standard for challenging the disparate impact of facially neutral practices was refined through a series of subsequent Supreme Court and lower court decisions, most of which, like \textit{Griggs}, involved testing. Moreover, as discussed below, this standard has been the reference point for the analysis of disparate impact discrimination in education law.

In the \textit{Griggs} line of cases, a plaintiff was first required to establish that the racial, ethnic, religious, or gender makeup of the pool of successful test-takers differed significantly from the pool of otherwise qualified applicants as a result of an employment practice or practices.\textsuperscript{183} The same requirements for making out a prima facie case apply under Title

\textsuperscript{180} 401 US at 431.
\textsuperscript{181} Id.
\textsuperscript{182} Id at 436. Title VII specifically endorses the use of a "professionally developed ability test" in employment, "provided that such test, its administration or action upon the results is not designed, intended, or used to discriminate because of race, color, religion, sex or national origin." 42 USC § 2000e-2(h). This section was not part of the original version of Title VII but was added during floor debate to assuage the fears of certain senators that Title VII would prohibit all employment testing and force employers to hire unqualified workers. \textit{Griggs} at 434-36. See also Barbara Schlei and Paul Grossman, \textit{Employment Discrimination Law} 82 (BNA 2d ed 1983).
\textsuperscript{183} See, for example, \textit{Albemarle Paper Co. v Moody}, 422 US 405, 425 (1975). See also Schlei and Grossman, \textit{Employment Discrimination Law} at 1326, 1326 n 126 (cited in note 182). As the Supreme Court has recently reaffirmed, "the 'proper comparison [is] between the racial composition of the at-issue jobs' and the racial composition of the qualified . . . population in the relevant labor market.'" \textit{Wards Cove}, 490 US 642, 650 (1989) quoting \textit{Hazelwood School Dist. v US}, 433 US 299, 308 (1977). While the differences must be significant, courts have not imposed a rigid test of statistical significance. In \textit{Watson v Ft. Worth Bank & Trust}, 487 US 977, 995-96 n 3 (1988), Justice O'Connor observed, "Courts appear generally to have judged the 'significance' or 'substantiality' of numerical disparities on a case-by-case basis. [Citations omitted.] At least at this stage of the law's development, we believe that such a case-by-case approach properly reflects our recognition that statistics 'come in infinite variety and . . . their
The burden then shifted to the employer to show that the employment practice or practices could be justified by "business necessity."\(^{185}\) Courts have "repeatedly" described this requirement as a "heavy burden" for employers.\(^{186}\) In education cases, the closely related concept of "educational necessity" has developed to describe the burden on the party defending the practice at issue.\(^{187}\)

In both employment and education testing analysis, "necessity" has been interpreted to incorporate the concept of validity. As the Supreme Court explained in *Albemarle*, "discriminatory tests are impermissible unless shown, by professionally accepted methods, to be 'predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs' [in question]."\(^{188}\) The Title IX regulations explicitly adopt a validity standard.\(^{189}\)

Even if an employer successfully met the burden of demonstrating business necessity, a plaintiff could still prevail by showing that other employment practices that did not have a discriminatory effect would also serve the employer's legitimate interest in securing employees who could perform the job.\(^{190}\) This concept of the less discriminatory alternative has also been applied in the judicial analysis of discriminatory test uses in education\(^{191}\) as well as in the Title IX regulatory scheme.\(^{192}\)

---

184 See *Sharif*, 709 F Supp at 361 (court held that “[u]nder this scheme, plaintiffs first must show that a facially neutral practice has a disproportionate effect”). See also, for example, 34 CFR § 106.21(b)(2); 34 CFR § 106.36(b).

185 *Griggs*, 401 US at 431.


187 As the Ninth Circuit held in *Larry P.*, which challenged the use of IQ tests to place students in classes for the educable mentally retarded, once disparate impact on the basis of race was established, "[t]he burden... shifted to the defendants to demonstrate that the IQ tests which resulted in the disproportionate placement of black children were required by educational necessity." 793 F2d at 983. See also *Sharif*, 709 F Supp at 361 ("[i]n educational testing cases, instead of requiring defendants to demonstrate a ‘business necessity,’ courts have required defendants to show an ‘educational necessity’") (citing *Branches of NAACP v State of Ga.*, 775 F2d 1403 (11th Cir 1985) and *Bd. of Educ v Harris*, 444 US 130 (1979)).

188 *Albemarle*, 422 US at 431. See also *Guardians*, 630 F2d at 88 ("[t]he real issue in this case, therefore, is whether the defendants have rebutted the plaintiffs' prima facie case by proving that its test was job-related: that the test accurately selected applicants who would be better police officers"); and *Sharif*, 709 F Supp at 361.

189 For example, the admissions regulations prohibit the use of admission tests which have an adverse effect on the basis of sex unless "the use of such test[s]... is shown to predict validity success in the education program or activity in question..." 34 CFR § 106.52 (employment). The Guidelines for Eliminating Discrimination in Vocational Education place an even heavier burden on test users. Once disparate impact has been shown, the recipient must "demonstrate that such criteria have been validated as essential to participation in a given program." 34 CFR Part 100, App B(IV)(K) (emphasis added).


191 See, for example, *Branches of NAACP v State of Ga.*, 775 F2d 1403, 1417 (11th Cir 1985) (holding under Title VI that plaintiff "may ultimately prevail by proffering an equally effective
b. The Uniform Guidelines on Employee Selection

While the foregoing analysis describes the general framework for analyzing disparate impact discrimination under both Title IX and Title VII, a specific and detailed jurisprudence regarding testing discrimination in employment has also developed under Title VII. This law is based principally on the Uniform Guidelines on Employee Selection Procedures ("Uniform Guidelines"), promulgated by the Equal Employment Opportunity Commission, and the cases which have interpreted the Uniform Guidelines principles. The Uniform Guidelines are a particularly important resource for the analysis of discrimination in educational testing because while the Title IX case law and regulatory scheme incorporate the same principles, the agencies charged with enforcing Title IX have not developed an education-specific counterpart to the Uniform Guidelines.

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." Generally, the proper comparison is between those who are ultimately selected and the pool of those who were qualified for selection for determining whether there has been a disparate impact, although in certain circumstances the pool is drawn

alternative practice which results in less racial disproportionality"); and Larry P., 495 F Supp at 973 (holding that under Title VI, "[e]ven if defendants had discharged their burden . . . plaintiffs would still be entitled to prevail if they could show that alternative devices for placement exist and would serve defendants' legitimate interests without the same discriminatory effect").

See, for example, 34 CFR § 106.36(b) and (c) (counseling regulation which implies that less discriminatory alternatives must be instituted when discrimination is found to arise from the use of counseling materials); 34 CFR § 106.21 (b)(2) (admissions); and 34 CFR § 106.52(b) (employment). See notes 236-41 and accompanying text for a discussion of the relationship between the Title IX and Title VII formulations of this construct.


Although the Uniform Guidelines do not independently have the force of law, they are entitled to deference by reviewing courts. In Griggs, the Court articulated a "great deference" standard, 401 US at 433-34; accord United States v Chicago, 549 F2d 415, 430 (7th Cir 1977) (guidelines should be complied with unless some cogent reason exists for non-compliance). The current state of the law, which has retreated from the full strength of the Griggs articulation, is well summarized by the Second Circuit's observation in Guardians, 630 F2d 79 ("[t]he [Supreme] Court appears to have applied the Guidelines only to the extent that they are useful, in the particular setting of the case under consideration, for advancing the basic purposes of Title VII . . . . Thus, the Guidelines should always be considered, but they should not be regarded as conclusive unless reason and statutory interpretation support their conclusions" (citations omitted). Id at 91).

Courts addressing claims of discriminatory test uses in education have been very aware of this lack of guidance. For example, in 1979, the district court in Larry P. observed that "[t]o date . . . there are no cases applying validation criteria to tests used for EMR [Educable Mentally Retarded] placement," and expressed concern about "[t]he problem of [the] lack of authority." 495 F Supp at 969. Similarly, in Debra P., in evaluating whether a high school competency exam was discriminatory, the Eleventh Circuit observed that "[t]he experts conceded that there are no accepted educational standards for determining whether a test is instructionally valid." 730 F2d 1405, 1412 (11th Cir 1984).


See, for example, Richardson v Lamar County Bd. of Educ., 729 F Supp 806, 815 (MD Ala 1989), aff'd Richardson v Ala. State Bd. of Educ., 935 F2d 1240 (11th Cir 1991).
more broadly.\textsuperscript{198}

The Uniform Guidelines incorporate what is known as the “four-fifths” rule, under which a selection rate for any protected group of less than four-fifths of the rate of the group with the highest selection rate is regarded as evidence of adverse impact.\textsuperscript{199} However, they do not preclude the use of other statistical analyses.\textsuperscript{200} Moreover, where there are smaller differences in the selection rate than at the four-fifths rate and/or where small numbers are involved,\textsuperscript{201} the Uniform Guidelines consider practical as well as statistical significance and permit looking to the use of the selection device over periods of time and in analogous circumstances to establish the requisite impact.\textsuperscript{202}

In education cases, the adverse impact of the test use may be both readily apparent and statistically significant. For example, in \textit{Sharif} the plaintiffs established disparate impact by showing that New York State’s exclusive reliance on SAT scores to award state-sponsored scholarships resulted in the award of seventy-two percent of Empire State Scholarships and fifty-seven percent of Regents Scholarships to males although they were only forty-seven percent of the scholarship competitors.\textsuperscript{203} These represented 15.8 standard deviations from the mean and 31.7 standard deviations from the mean, respectively.\textsuperscript{204} The court found that the plaintiffs proved their 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.”\textsuperscript{205}

\textsuperscript{198} For example, the Title IX counseling regulations take a broader view of the appropriate pool. Under 34 CFR § 106.36(b) and (c), impact is established where there is a substantially disproportionate number of members of one sex in a particular course of study, classification, or class. In a similar vein, in an investigation of discriminatory selection practices at the Chicago Board of Education’s Washburne Trade School, the Office for Civil Rights determined that the relevant pool for comparison was females who were “potentially available for training,” rather than those who had actually applied to the school or taken a particular test. Letter from OCR Regional Director Kenneth A. Mines to Dr. Manford Byrd, Jr., Case No 05-85-1008 at 2 (Mar 28, 1986).

\textsuperscript{199} See \textit{Clady v County of Los Angeles}, 770 F2d 1421, 1428 (9th Cir 1985) (noting that the “four-fifths” rule has been criticized by courts and commentators).

\textsuperscript{200} See \textit{Rivera v Wichita Falls}, 665 F2d 531, 536, 536 n 7 (5th Cir 1982) (“[a] difference of more than two or three standard deviations is generally considered to raise a compelling inference of discrimination” (citing \textit{Castaneda v Partida}, 430 US 482, 496-97 n 17 (1977))).

\textsuperscript{201} “[S]mall sample size may, of course, detract from the value of [statistical] evidence. . . .” \textit{Teamsters v United States}, 431 US 324, 339-40 n 20 (1977), and thus impede a showing of disparate impact. As the court in \textit{United States v Lansdowne Swim Club}, 713 F Supp 785 (ED Pa 1989), explained, “[t]he danger posed by small samples is that they may produce short-term results that would not hold over the long run, and thus erroneously may be attributed to discriminatory practices rather than to chance.” Id at 809. However, “small numbers are not per se useless, especially if the disparity shown is egregious.” \textit{Valentino v United States Postal Service}, 674 F2d 56, 72 (DC Cir 1982). Accord \textit{Rivera v Wichita Falls}, 665 F2d at 536-37 n 7 (sample size of thirty-five will not defeat a showing of disparate impact where disparity is great).

\textsuperscript{202} See 34 CFR § 106.4(D).

\textsuperscript{203} \textit{Sharif}, 709 F Supp at 355.

\textsuperscript{204} Id.

\textsuperscript{205} Id at 362.
Similar analyses can be conducted to compare, for example, percentages of females in the following categories: those who take the PSAT/NMSQT, qualify for semi-finalist status in the National Merit Scholarship competition and are awarded scholarships, those who as seventh graders take the SAT and qualify for admission to the various enrichment programs for gifted and talented adolescents, and those who take the ASVAB or DAT and are admitted to various vocational education courses or programs.

Where statistical significance is difficult to establish because of small sample size, the Uniform Guidelines' recognition of practical significance and mechanisms to expand sample size become relevant. An example of where this alternate analysis may be useful is the consideration of adverse impact in certain vocational education test uses, especially for claims brought by minority women, where relatively small numbers of test-takers and allocated benefits are implicated in any one setting or at any one time. Under the Uniform Guidelines, results from different test administrations may be aggregated to determine whether discriminatory impact is present. Similarly, these provisions may come into play in the analysis of whether disparate impact results from tests used for admissions to small programs or scholarship awards made by relatively small scholarship programs.

Once impact has been established, the analysis turns to whether the test use can be defended under the necessity standard. There are two basic and interrelated questions to be answered. The first is typically framed in terms of test validity: Does the test do what it purports to do? As the district court observed in a case challenging the use of a high school competency test which had a racially adverse impact, "[p]ut simply, the task assigned to this Court by the Court of Appeals was to find out if Florida is teaching what it is testing." The second question is whether the test addresses critical or important skills and abilities necessary for the performance of the job or success in the educational pursuit.

Although formal validation studies are not an absolute requirement under Title VII, the Uniform Guidelines and the case law note the

---

206 Recent studies show that although females represent slightly more than half of the test takers, they only receive between 33% and 40% of National Merit Scholarships. Figures are not publicly available regarding the gender breakdown of semi-finalists although it is believed to resemble the gender breakdown of scholarship winners. See discussion at notes 22-23 and accompanying text.

207 As discussed earlier, the pool of students taking the test is evenly distributed by gender although the math programs enroll twice as many boys as girls. See discussion at notes 28-30 and accompanying text.

208 Debra P., 730 F2d at 1409 (citing with approval Debra P., 564 F Supp at 180). See also Larry P., 495 F Supp at 968 ("[v]alidation' is the determination of whether the placement tests or other evaluation materials are suited for the purposes for which they are used").

209 Watson v Ft. Worth Bank & Trust, 487 US 977, 997 (1988). However, the examples given in Watson in support of the proposition that formal validation studies are not required as a matter of law are limited in scope. They provide little guidance to an educational test user who
importance of establishing validity in accordance with established practice. There are three basic strategies for establishing validity: content validation, criterion-related validation, and construct validation. To the extent that validity studies have been undertaken, criterion-related validity, which analyzes a test's ability to predict accurately important elements of job performance, has been the principal strategy used in connection with education-related testing. For example, the College Board has used criterion analysis to validate the SAT for college admissions purposes. Content validity, which focuses on whether the content of the test is representative of important aspects of the job, and construct validity, which attempts to measure abstract traits (such as judgment) which are deemed important to performance on the job, will likely have a more limited application to education-related testing.

seeks to defend a test use in the absence of a validity study. For example, N. Y. Transit Authority v Beazer, 440 US 568 (1979), involved the blanket exclusion of methadone users from employment by the New York City Transit Authority. The case did not involve an abilities test and the decision turned on broad policy considerations. Id at 590-93. Furthermore, in Washington v Davis, 426 US 229, 235, 250 (1976), the narrow question presented was whether a written entrance examination for a police recruit training course was required to be validated as a predictor of both recruit training success and future job-performance. The Court held that validation as a predictor of recruit training success was sufficient to uphold the use of the test. The Court did not address the issue of whether validity studies were required in general. Moreover, Washington v Davis was a Fifth Amendment rather than a Title VII case. Id at 233. In any event, where the Uniform Guidelines are not followed the test user has a heavier burden in establishing validity and defending against a claim of disparate impact. United States v Chicago, 573 F2d at 427. See also Craig v County of Los Angeles, 626 F2d 659, 665 (9th Cir 1980) (stating that “noncompliance with the EEOC guidelines diminishes the probative value of the defendant's validation study. But it is not necessarily fatal”).


211 29 CFR § 1607.5(B) (“[n]ew strategies for showing the validity of selection procedures will be evaluated as the become accepted by the psychological profession”). See also Washington v Davis, 426 US at 247, n 13.


213 See discussion at notes 77-80 and accompanying text. The Uniform Guidelines caution that “[c]riterion measures consisting of paper and pencil tests will be closely reviewed for job relevance.” 29 CFR § 1607.14(B)(3). See Craig v County of Los Angeles, 626 F2d 659. The same principle applies to the validation of education tests. They must be relevant to the purpose for which they are being used.

214 A classic example of a content-validated test is a typing test for the position of typist. Content validity is generally not suited to the educational context where, by definition, the candidate is being assessed regarding his or her ability to develop skills and not whether he or she already possesses them. See 29 CFR § 1607.14(C)(1) (1991). An exception may be the validation of achievement tests used for placement purposes. See, for example, Larry P., 495 F Supp 926, 970 n 84.

215 29 CFR § 1607.14(D). Construct validity is also less likely to be used because it “is both an extensive and arduous effort involving a series of research studies, which include criterion related validity studies and which may include content validity studies.” 29 CFR § 1607.14(D)(1). Construct validity did come into play in determining the validity of IQ tests for placing Black children in classes for the educable mentally retarded, Larry P., 495 F Supp at 970 n 84, although the court also observed that the “definitions [of the various validity strategies] are sometimes confused . . . and the important concern is the practical one of establishing the relationship. . . .” Id.
Regardless of the validation strategy, any study must start with a careful analysis of the purpose for which the test is being used. In the employment context, such an analysis must determine "critical or important job duties, work behaviors or work outcomes" with a particular emphasis in avoiding bias both in the selection of the measures and their applications. Similarly, in education testing, test-users should identify "critical or important" skills and abilities which the test is designed to elicit. This requirement applies whether the test is being used to select candidates for an undergraduate baccalaureate program or an enrichment program for gifted adolescents, to award scholarships, or to counsel students in their choice of a vocational education program.

The next step is to determine the relationship between performance on the test and performance in the skills and abilities which are being measured. This relationship is often measured in terms of a correlation coefficient. A positive number denotes a positive relationship, with 1.0 indicating a perfect predictive relationship between test scores and job success, while a negative number demonstrates that the better one does on the test, the more unsuited one is to the job in question. In Title VII case law, "courts have commonly not accepted as valid a test having a correlation coefficient of under .30." Validity can also be measured in terms of statistical significance. Under the Uniform Guidelines, a selection procedure is generally considered valid when "the relationship between performance on the [test] and performance on the criterion measure is statistically significant at the 0.05 level of significance, which means that it is sufficiently high as to have a probability of no more than one (1) in twenty (20) to have occurred by chance." Even when a test has been validated for certain purposes, it by no means follows that all applications of the test are valid.

216 The Uniform Guidelines make it clear that in the employment context, "[a]ny validity study should be based upon a review of information about the job for which the selection procedure is to be used." 29 CFR § 1607.14(A). See also 29 CFR § 1607.14(B)(2) (job analysis in criterion-related validity study); 29 CFR § 1607.14(c)(2) (job analysis in content-related validity study); 29 CFR § 1607.14(D)(2) (job analysis for construct-related validity study).

217 29 CFR § 1607.14(B)(2).

218 Id. See also 29 CFR §§ 1607.14(C)(2) and (D)(2).


220 Id at 129, 129 n 131.

221 29 CFR § 1607.14(B)(5). The Uniform Guidelines go on to note that the absence of a statistically significant relationship is not necessarily dispositive of the test's lack of validity. Id.

222 According to the Uniform Guidelines, "[u]nder no circumstances will the general reputation of a test . . . , its author or its publisher, or casual reports of it's [sic] validity be accepted in lieu of evidence of validity. Specifically ruled out are: assumptions of validity based on a procedure's name or descriptive labels; all forms of promotional literature; data bearing on the frequency of a procedure's usage; testimonial statements and credentials of sellers, users, or consultants; and other nonempirical or anecdotal accounts of selection practices or selection outcomes." 29 CFR § 1607.9(A). The district court in Larry P. relied on this section of the Uniform Guidelines to criticize the use of IQ tests for purposes for which they had not been specifically validated. 495 F Supp at 971 (citing to 29 CFR § 1607.8 (1978)). In fact, the court read this section to require a showing of the validity of the test for each minority group for which it is used. Id.
important concept in the educational setting where tests which may be valid for certain purposes are widely used for many other purposes for which there may be no validity evidence. A prime example is the SAT. While its proffered validity is based on the prediction of first-year college grades for high school juniors and seniors, it is used for purposes as diverse as awarding scholarships and identifying mathematically gifted seventh graders. Moreover, the SAT has no “general” validity for admissions purposes but must be validated separately by each institution which uses the test.\(^{223}\) Similarly, the ASVAB, for which there is validity evidence for certain military jobs, is widely used in a range of civilian high school settings for purposes for which there is no validity evidence. Under the Uniform Guidelines, each use must be independently validated.

Moreover, when test-users rely on a cutoff score, such as in an admissions or scholarship program, they must have independent validity evidence for the use of that cutoff score, even where validity has been demonstrated for the test generally.\(^{224}\) The danger of setting an arbitrary cutoff score is that such a practice “may well lead to the rejection of applicants who were fully capable of performing the job.”\(^{225}\) Title VII law does not require a test-user to perform a separate validity study to justify a cutoff score;\(^{226}\) however, there must be a professionally established basis to justify the cutoff point.\(^{227}\) Accordingly, for example, institutions which award scholarships to National Merit finalists must be able to point to validity evidence to support the National Merit cutoff scores, including the use of different cutoff scores by state. Similarly, programs for academically talented youth must be able to justify both their reliance on the SAT and their reliance on the particular cutoff scores they have chosen. The same is true for institutions which use SAT cutoff scores in

\(^{223}\) The predictive value of the SAT is affected by a number of factors including gender. See discussion at notes 80-85 and accompanying text.

\(^{224}\) Guardians, 630 F2d at 105.

\(^{225}\) Id. The Guardians court continued, “This does not mean that every person who fails a test by a single point necessarily has a claim for legal redress. . . . But when an exam produces disparate racial results, a cutoff score requires adequate justification and cannot be used at a point where its unreliability has such an extensive impact. . . .” Id at 106. The court accordingly invalidated a cutoff score where the employer merely determined how many vacancies it needed to fill and selected as many applicants as it needed, moving down the list. Id at 105. See also Thomas v Evanston, 610 F Supp 422, 430-31 (ND Ill 1985); and Burney v Pawtucket, 559 F Supp 1089, 1103 (D RI 1983).

\(^{226}\) See Uniform Guidelines, 29 CFR § 1607.5(H). “Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force.” Id.

\(^{227}\) The Guardians court suggested that an employer could validate a cutoff score “by using a professional estimate of the requisite ability levels,” or by finding a logical “‘break-point’ in the distribution of [the] scores.” Guardians, 630 F2d at 105. See also Gillespie v Wix, 771 F2d 1035, 1041-42 (7th Cir 1985) (court accepted professional estimate of the minimum abilities needed to perform the job in question); Bridgeport Guardians v Bridgeport Police Dept., 431 F Supp 931, 939-40 (D Conn 1977) (to be adequate, a passing grade must distinguish between those who are qualified for the job and those who are not).
making admissions or placement decisions.\textsuperscript{228}

Finally, tests must be fair. That is, they must measure abilities and predict performance in the same manner for all test-takers, without regard to sex, race, or national origin.\textsuperscript{229} The Uniform Guidelines incorporate a requirement of test fairness\textsuperscript{230} which is directly relevant to the analysis of education-related tests such as the SAT, which predict differently by gender. Indeed, the SAT’s lack of “fairness” in precisely this sense of the term was a major underpinning of the decision in \textit{Sharif}.\textsuperscript{231}

Similarly, the Ninth Circuit held in \textit{Larry P.} that a successful defense of the use of IQ tests for the disproportionate placement of Black children in classes for the educationally mentally retarded would have to show that “the tests predict specifically that black elementary schoolchildren (as opposed to white elementary schoolchildren) who score at or below 70 on the IQ tests are mentally retarded and incapable of learning the regular school curriculum.”\textsuperscript{232}

The Uniform Guidelines define “unfairness” as follows:

When 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, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores.\textsuperscript{233}

\textsuperscript{228} Similar concerns are present regarding the practice of basing selection decisions on the rank-ordering of scores, although this practice has not been identified as widespread in the educational arena. Under the Uniform Guidelines, to justify rank-ordering an employer must show that “a higher score . . . is likely to result in better job performance.” 29 CFR \textsection{}1607.14(C)(9) (in reference to content validation).

\textsuperscript{229} Legislative history to the Civil Rights Act of 1991, 105 Stat 1071, underscores the vitality of the legal requirement of fairness. In addressing \textsection{}106 of the Act, which prohibits certain score adjustments in employment related tests, 105 Stat at 1074, Rep Don Edwards, D-Cal, one of the chief House sponsors of the legislation, stated that “this section does not alter existing legal requirements with respect to demonstrating that a test operates as fairly with respect to one gender or race as with respect to another.” He continued, “[a] test which does not provide the same opportunity for selection to men and women, or blacks and whites, or Hispanics and Anglos who perform equally well on the job, or which predicts job performances differently because of race or gender, would not be a fair test and would not be ‘job-related for the position in question and consistent with business necessity.’” 137 Cong Rec H9529 (daily ed Nov 7, 1991).

\textsuperscript{230} 29 CFR \textsection{}1607.14(B)(8) (1991). The Uniform Guidelines characterize fairness as a “developing concept,” and caution that fairness studies are generally only technically feasible where there are large samples involved. Id. This requirement has not been widely applied in the employment context. See, for example, \textit{Clady v County of Los Angeles}, 770 F2d 1421, 1431 (9th Cir 1985). The earlier formulation of the Uniform Guidelines’ provision regarding fairness, which was adopted and applied by the courts, required that “differential validity” be established for minority and nonminority groups wherever technically feasible. See, for example, \textit{Albemarle Paper Co. v Moody}, 422 US 405, 435 (1975); and \textit{United States v Ga. Power Co.}, 474 F2d 906, 914 (5th Cir 1973).

\textsuperscript{231} See \textit{Sharif}, 709 F Supp at 353-54 (“while the SAT will predict college success as well for males within the universe of males as for females within the universe of females, when predictions are within the combined universe of males and females, the SAT underpredicts academic performance of females in their freshman year of college, and overpredicts such academic performance for males” (emphasis in original) (citations omitted)).

\textsuperscript{232} \textit{Larry P.}, 793 F2d at 980.

\textsuperscript{233} 29 CFR \textsection{}1607.14(B)(6)(a). Test-users must generally investigate the issue of unfairness
When unfairness is shown, the test-user must revise or replace the selection instrument. Appropriate revisions include those which "assure compatibility between the probability of successful job performance and the probability of being selected." In other words, scoring and/or test usage may be adjusted to assure that the test is used in a non-discriminatory manner.

As is clear from the review of the empirical literature, there is either very little or no validity evidence for a wide range of education test uses. Consequently, many of the difficult and technical questions which can arise under the Uniform Guidelines and the case law are simply not relevant here. Instead, the far more straightforward analysis of whether baseline requirements for establishing validity in accordance with professionally accepted standards controls. Too often these requirements have not been addressed in any serious fashion at all, much less adequately met.

Even where validity studies may exist, fairness remains a major concern in the analysis of gender discriminatory test uses. For example, users who do not separately validate their use of SAT scores by gender when such use results in a disparate impact take a substantial risk. This risk particularly applies where sufficiently large samples are involved to make a fairness investigation technically feasible. Under the doctrine in the Uniform Guidelines, reflected in both Sharif and Larry P., test-users may well not be able to demonstrate validity, and thus educational necessity within the meaning of the law, if their test use has not considered and does not adjust for the fact that the SAT underpredicts female performance and overpredicts male performance.

If after a showing of adverse impact a test use is nonetheless determined to be valid, a complaining party can still prevail if he or she shows that the employer or educational institution can achieve the same end with a less discriminatory alternative to the test use at issue. Indeed, the consideration of less discriminatory alternatives is an integral part of the Uniform Guidelines scheme which requires that alternative selection procedures be considered as part of any validation study. The Uniform Guidelines also require that test-users investigate alternative selection procedures which have "evidence of less adverse impact and substantial evidence of validity for the same job in similar circumstances."

Title IX adopts this principle generally, and the Title IX voca-

---

235 See discussion of Kirkland v N.Y. State Dept. of Correctional Services, 628 F2d 796 (2d Cir 1980), at notes 373-76 and accompanying text.
236 29 CFR § 1607.3(B).
237 Id.
238 See, for example, 34 CFR § 106.21(b)(2) (admissions).
tional education guidelines and counseling regulations take the principle further. Once disparate impact has been established, the vocational education guidelines explicitly put the burden on the test-user to demonstrate that less discriminatory alternatives are not available.\(^{239}\) The counseling regulation takes a slightly different approach but reaches the same end, since it requires that upon a showing of disparate impact a test-user must take such action as is necessary to assure itself that the impact is not the result of discrimination.\(^{240}\) While the regulation clearly requires the test-user to evaluate the validity of the instrument, it necessarily also requires the test-user to assure itself that less discriminatory alternatives are not available.

Sound policy considerations support these deviations from the Title VII model. As discussed above, while an employer may have a legitimate interest in "weeding out" unqualified applicants, education is designed to create these qualifications in the first place. Accordingly, it is appropriate to put a higher burden on a test-user to justify a test with a disparate impact in education, as opposed to employment, uses.\(^{241}\)

c. Implications of the Civil Rights Act of 1991

The language and history of Title IX and its regulations thus support the application of an analysis at least as stringent as the Griggs-Uniform Guidelines analysis developed under Title VII to sex discrimination in education. However, since 1989 there have been two dramatic changes in the framework for analyzing disparate impact discrimination under Title VII. First, in \textit{Wards Cove Packing Co. v Atonio},\(^{242}\) the Supreme Court reversed \textit{Griggs} with respect to both the allocation and the nature of the burdens of proof on the respective parties. Subsequently, with the enactment of the Civil Rights Act of 1991, Congress rejected \textit{Wards Cove} and codified a \textit{Griggs}-based disparate impact cause of action as part of Title VII.

While the statutory language and controlling legislative history of the Civil Rights Act of 1991 reflect Congress' clear intent to reject \textit{Wards Cove} and restore \textit{Griggs}, it is likely that questions regarding the precise interpretation of the new disparate impact provision will be aggressively litigated. The better view, as set out below, is that the Civil Rights Act of 1991 will be construed to restore \textit{Griggs} fully. As such, it will provide

\(^{239}\) 34 CFR, Part 100, App B(IV)(K).
\(^{240}\) 34 CFR §§ 106.36(b) and (c).
\(^{241}\) See discussion of the Title IX disparate impact analysis at notes 161-70 and accompanying text.
\(^{242}\) 490 US 642 (1989). Non-white cannery workers at Wards Cove's Alaskan salmon canneries challenged hiring and promotion practices, alleging that these practices were responsible for the extreme racial stratification of the workforce. The Supreme Court rejected their claims. Id.
strong support for extending a Griggs-based standard to Title IX disparate impact analysis as well. However, insofar as the courts ultimately adopt an interpretation of the Civil Rights Act of 1991 which does not squarely cohere with Griggs, that interpretation should not extend to Title IX.\textsuperscript{243}

The Civil Rights Act of 1991 was enacted to reverse a series of Supreme Court decisions narrowly interpreting the law of employment discrimination\textsuperscript{244} and to provide, for the first time, a monetary damages remedy in Title VII cases.\textsuperscript{245} One of the prime targets of the legislation was the Supreme Court’s decision in \textit{Wards Cove}. The Court in \textit{Wards Cove} held that any legitimate business reason will constitute business necessity and justify a practice giving rise to a disparate impact. Furthermore, the Court shifted the burden of persuasion from the defendant to the plaintiff, requiring the plaintiff to establish the lack of business necessity, although the defendant retained the burden of producing evidence of justification for the practice.\textsuperscript{246} Finally, the Court addressed the plaintiff’s rebuttal showing of a less discriminatory alternative, making it clear that cost and administrative convenience were fully appropriate employer considerations in rejecting an alternative approach.\textsuperscript{247}

The Civil Rights Act of 1991 rejected \textit{Wards Cove} and restored a Griggs-based analysis to Title VII disparate impact discrimination. First, it put the burden of proving business necessity back on the defendant once the plaintiff has established a disparate impact.\textsuperscript{248} Second, the Act

\textsuperscript{243} See discussion at notes 171-76 and accompanying text.

\textsuperscript{244} In addition to \textit{Wards Cove}, the legislation also reversed in part the Supreme Court’s decisions in, inter alia, \textit{Patterson v McLean Credit Union}, 491 US 164 (1989); \textit{Price Waterhouse v Hopkins}, 490 US 228 (1989); \textit{Martin v Wilks}, 490 US 755 (1989); and \textit{Lorance v AT&T Technologies}, 490 US 900 (1989).

\textsuperscript{245} Until the passage of the Civil Rights Act of 1991, Title VII provided only equitable remedies. See 42 USC § 2000e-5(g). The Civil Rights Act of 1991 creates a monetary damages remedy including both punitive and compensatory damages for intentional violations of Title VII and the Americans with Disabilities Act through a new statutory section codified at 42 USC § 1981A. The standard for the award of damages is based on 42 USC § 1981, which provides damages in cases of intentional employment discrimination on the basis of race or national origin. However, unlike § 1981, which does not limit the amount of damages which may be awarded, the new § 1981A limits available damages through a sliding scale of caps based on the number of employees of the employer and ranging from $30,000 to $300,000. 42 USC § 1981A(b)(1)(3). Legislation was introduced in both the House and the Senate in late 1991 to remove the caps. HR 3975, 102d Cong, 1st Sess (1991); S 2062, 102d Cong, 1st Sess (1991).

\textsuperscript{246} \textit{Wards Cove}, 490 US at 659-60. While the \textit{Wards Cove} court insisted that its holding was consistent with preexisting law with regard to other parts of the decision, it acknowledged with regard to the burden of proof question that “some of our earlier decisions can be read as suggesting otherwise.” Id at 660.

\textsuperscript{247} The Court also addressed the nature of the plaintiff’s burden in making out a prima facie case, introducing a requirement that plaintiffs demonstrate “that specific elements of the... hiring process have a significantly disparate impact...” Id at 658 (emphasis added). While not affecting challenges focusing solely on the giving of tests, which is viewed as a single employment practice, see, for example, \textit{Allen v Seidman}, 881 F2d 375 (7th Cir 1989), this “disaggregation” requirement was exceedingly burdensome in disparate impact cases involving a series of employment practices which could include testing.

\textsuperscript{248} Regarding the prima facie case, the Act maintains a general disaggregation requirement for Title VII cases but provides that, “if the complaining party can demonstrate to the court that
provides that in order to rebut the plaintiff's prima facie case, the employer must demonstrate that the discriminatory practice is "job related for the position in question and consistent with business necessity." While the Act does not define these terms, they are squarely grounded in Griggs. The linkage of business necessity and the requirement that the practice must be job-related for the position in question provides a forthright endorsement of the Griggs doctrine that job-relatedness and business necessity are simply two sides of the same coin.

Moreover, the actual language in the Civil Rights Act of 1991 is taken verbatim from the section of the Americans With Disabilities Act ("ADA") which addresses defenses in disparate impact cases based on disability discrimination and which in turn incorporated Griggs. In his floor statement regarding the Civil Rights Act of 1991, Congressman Don Edwards, D-Cal, one of the Act's principal House sponsors, explained this connection and went on to set out the meaning of the ADA language:

As explained in the legislative history and subsequent regulations issued under that Act [the ADA], this language clearly requires proof by an employer of a close connection between a challenged practice with disparate impact and the ability to actually perform the job in question. See, e.g., Report on the ADA by House Committee on Education and Labor at 343-44, 345; Report on the ADA by the House Committee on the Judiciary at 482. If the employer satisfies its burden, the plaintiff may still prevail by demonstrating that another less discriminatory employment practice is

the elements of a respondent's decisionmaking process are not capable of separation for analysis, the decisionmaking process may be analyzed as one employment practice." Civil Rights Act of 1991, § 105(k)(1)(B)(i), 105 Stat at 1074. Tests will thus continue to be analyzed as a single employment practice. Furthermore, certain Title IX regulations specifically diverge from Title VII in this area. The Title IX requirement should, of course, apply. An example is found in the Title IX counseling regulation, 34 CFR § 106.36(c) (where there is a substantial gender disproportionate enrollment in a particular class, the recipient must "assure itself" that such disproportion is not the result of discrimination in counseling or appraisal materials). Under this regulation, Title IX plaintiffs need only show a disproportionate enrollment and there is no requirement at all that they disaggregate the causes of such disproportion.

The Civil Rights Act of 1991 states in pertinent part:

An unlawful employment practice based on disparate impact is established under this title only if — (i) a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity.


See, for example, Griggs, 401 US at 431: "The touchstone is business necessity. If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited."

42 USCA § 12113 (1990).

available and that the defendant refuses to adopt it. In an unusual statutory provision, the Act states that the standard for this demonstration shall be in accordance with the law as it existed on June 4, 1989, with respect to the concept of ‘alternative employment practice.’ June 4, 1989 was the day before Wards Cove was decided. The Civil Rights Act of 1991 thus rejects the Wards Cove formulation of alternative employment practices and restores the pre-Wards Cove standard, most clearly presented in Albemarle Paper Co. v Moody.

Legislative history confirms this analysis of the disparate impact section. The Civil Rights Act of 1991 specifically endorses exclusive legislative history for interpreting “any provision of this Act that relates to Wards Cove-Business Necessity/cumulation/alternative business practice.” Regarding the defendant’s burden, a Congressional interpretive memorandum states that “[t]he terms ‘business necessity’ and ‘job related’ are intended to reflect the concepts enunciated by the Supreme Court in Griggs v Duke Power Co., 401 US 424 (1971) and in the other Supreme Court decisions prior to Wards Cove Packing Co. v Atonio, 490 US 642 (1989).” Congress has thus left no doubt that it has rejected Wards Cove and restored Griggs. The memorandum does not address the alternative business practices question, leaving the statutory language to speak for itself. The only possible construction of that language is that the Albemarle standard, the standard in effect before Wards Cove, governs.

Notwithstanding Congress’ clear intent to restore the Griggs standard, as evident through both statutory language and supporting legislatively endorsed history, defendants in disparate impact cases may argue, based on a statement introduced into the legislative record by Senator Robert Dole, that the Wards Cove standard is still good law. Indeed,

---

256 Sec. 105(b) of the Act states: “No statements other than the interpretive memorandum appearing at Vol. 137 Congressional Record S15276 (daily ed Oct 25, 1991) shall be considered legislative history of, or relied upon in any way as legislative history in construing or applying, any provision of this Act that relates to Wards Cove-Business Necessity/cumulation/alternative business practice.” Civil Rights Act of 1991, 105 Stat at 1075.
258 Id. Regarding disaggregation, the interpretive memorandum states, “[w]hen a decision-making process includes particular, functionally-integrated practices which are components of the same criterion, standard, method of administration, or test, such as the height and weight requirements designed to measure strength in Dothard v Rawlinson, 433 US 321 (1977), the particular, functionally-integrated practices may be analyzed as one employment practice.” 137 Cong Rec S15276 (daily ed Oct 25, 1991). While tests clearly will be analyzed as a single practice, this explanation still leaves in place a disaggregation requirement which is substantially more restrictive than that under Title IX. In light of such a conflict, the Title IX rule should govern. See discussion at notes 161-70 and accompanying text.
259 137 Cong Rec S15473-76 (daily ed Oct 30, 1991) (statement of Sen Robert Dole, R-Kan). Dole’s statement was joined by several Republican senators, all of whom had opposed the Civil Rights Act until the very last minute. See also 137 Cong Rec H9543-46 (daily ed Nov 7, 1991) (statement of Rep Henry Hyde, R-III).
in his signing statement, President Bush referred to Senator Dole's statement as the definitive history of the Civil Rights Act of 1991 and directed federal agencies to follow that interpretation. However, the Dole statement is wholly irrelevant to the construction of the disparate impact provisions of the Act for the simple reason that it is not part of the exclusive, statutorily approved legislative history. Moreover, Senator Dole had been an opponent of the legislation until the very last moment, and not a chief sponsor. However, given the President's effort to bootstrap it into the controlling interpretation, it is useful to address briefly why it is not persuasive in any event.

Dole argues that instead of rejecting *Wards Cove*, the Civil Rights Act of 1991 actually embraces that decision, with the sole exception of the burden of proof question. To reach this conclusion, Dole first reviews a number of versions of the disparate impact standard which had been advanced in Congress prior to the one which was ultimately adopted, all of which included definitions of business necessity. He observes that there is no definition of business necessity in the enacted version and assumes, without explanation, that, in the absence of a statutory definition, the Purposes section of the Act controls the meaning of both "job-related for the position in question" and "business necessity." This section includes the same language as that in the statutorily approved legislative history, referring to Congress' intent to adopt the principles set forth in *Griggs* and other Supreme Court decisions prior to *Wards Cove*. Dole then argues that post-Griggs cases, principally including *New York Transit Authority v Beazer* and *Watson v Fort Worth Bank & Trust*, incorporate the same standard as *Wards Cove*.

260 Referring to the disparate impact provision of the Civil Rights Act of 1991, the President stated, "[t]hese highly technical matters are addressed in detail in the analyses of S1745 [the Civil Rights Act as passed by Congress] introduced by Senator Dole on behalf of himself and several other Senators and of the Administration (137 Cong Rec S15472-S15478 [daily ed Oct 30, 1991]; 137 Cong Rec S15953 [daily ed Nov 5, 1991]). These documents will be treated as authoritative interpretive guidance by all officials in the executive branch with respect to the law of disparate impact as well as the other matters covered in the documents." Statement on signing the Civil Rights Act of 1991, 27 Weekly Comp Pres Doc 1701 (Nov 21, 1991).

261 For statutory approval of legislative history, see Civil Rights Act of 1991, § 105(b), 105 Stat at 1075.

262 It is a basic rule of statutory construction that statements by members, as distinct from sponsors or committee members, during legislative debate are considered only "where they show a common agreement in the legislature about the meaning of an ambiguous provision," Norman J. Singer, *Sutherland Statutes and Statutory Construction* § 48.13 (4th ed, Callaghan, 1984), or if they are consistent with statutory language and other legislative history which justify reliance upon them as evidence of legislative intent. Id. As a minority view in direct opposition to both statutory language and the statements of the chief sponsors, Dole's views are entitled to no weight.

263 According to Sen Dole, "the bill embodies longstanding concepts of job-relatedness and business necessity and rejects proposed innovations. In short, it represents an affirmation of existing law, including *Wards Cove*." 137 Cong Rec S15474 (daily ed Oct 30, 1991).


265 See notes 256-58 and accompanying text.


Therefore, he concludes, the Purposes section demonstrates that the Civil Rights Act of 1991 actually codifies *Wards Cove*, and thus all that an employer need establish to defeat a disparate impact claim is a relationship between the discriminatory practice and some legitimate business purpose.

This argument is not supported. First, since the formulation of the employer's burden is taken straight out of the ADA — in addition to incorporating language which is firmly based in *Griggs* — it cannot reasonably be argued that it is devoid of intrinsic meaning or in need of statutory definition at all. Moreover, the pre-*Wards Cove* cases that Dole relies on do not contain holdings which square with *Wards Cove*. There is no majority decision in *Watson* and, given the reference in the Purposes section to "decisions," Congress demonstrated no intent to adopt Justice O'Connor's plurality opinion in that case. Furthermore, the holding in *Beazer* is a classic application of the *Griggs* job-relatedness standard. In *Beazer*, the Court held that methadone users could be excluded from certain positions with the New York City Transit Authority because they could not reliably perform the specific job requirements of these "safety sensitive" positions.268

In sum, Dole's interpretation turns the plain meaning of the statutory language on its head as he argues that the terms "job-related for the position in question" and "consistent with business necessity" do not mean that at all, but only mean that there must be some general connection to the employer's legitimate business concerns. To the contrary, by setting out a job-relatedness standard, Congress has spoken clearly: practices which result in a disparate impact on a protected class must be related to the job in question and must be justified by business necessity. In short, they must satisfy the test set forth by the Supreme Court in *Griggs*.

An analysis of the Civil Rights Act of 1991 thus confirms Congress' intent that strong standards should be applied in analyzing claims of disparate impact discrimination. In Title IX challenges to testing discrimination, educational test-users will similarly have the heavy burden of demonstrating that the test is justified by educational necessity and is related — that is, valid — for the particular use to which it is put.

4. The Particular Legal Issues Confronting Women of Color: The Intersection of Title IX and Title VI

The final question in the development of the federal statutory framework for analyzing discrimination in educational testing concerns the

268 440 US at 587 n 31. See also *Conn. v Teal*, 457 US 440, 446 (1982) (describing *Griggs* as holding tests invalid because "they had a disparate impact and were not shown to be related to job performance" (emphasis added)).
treatment of claims by females of color. While there has been little research done regarding the extent and nature of the test-based discrimination experienced by this population, the empirical literature which does exist makes it clear that minority females suffer a very real double jeopardy based on both their sex and their race. They are, for example, consistently the lowest scorers on tests ranging from the SAT\textsuperscript{269} to the ASVAB.\textsuperscript{270} Moreover, their injury is not neatly attributable to either their racial or gender status; rather, it flows from a complex, and by no means fully understood, interaction of the two.

On their face the two relevant statutes, Title IX and Title VI, do not recognize the multidimensional nature of this discrimination. They separately prohibit sex discrimination and race or national origin discrimination, but not their combined impact. However, the problems presented by forcing a claim brought by women of color to be tried as either a sex discrimination case or a race/national origin discrimination case can be substantial. Where defenses are available against race and sex discrimination claims when analyzed separately, a bona fide victim of multiple discrimination may well be left without any remedy at all. For example, in a challenge to the discriminatory impact of the ASVAB by a female of color, the defendant could argue that gender-linked score differentials, such as the lower scores of women on the mechanical and electronic sections of the test, are irrelevant to a race claim. Similarly, racial differentials could be factored out of a sex discrimination claim. The remaining sex or race discrimination could then be insufficient to support a finding of a violation of the law. This result could occur despite the fact that females of color score lower on the ASVAB than do members of any other group and suffer the greatest injury from the use of ASVAB scores.

However, a jurisprudence weaving together the well-established principles prohibiting sex and race/national origin discrimination is beginning to develop which could address satisfactorily the unique problems faced by women of color. The analysis, which is found in legal commentary\textsuperscript{271} as well as in several cases, is based in the doctrine of

\begin{itemize}
\item \textsuperscript{269} From highest scoring group to lowest, the 1991 SAT average combined scores by gender and ethnicity were as follows: Asian American males, white males, Asian American females, white females, Native American males, Latin American males, Mexican American males, Native American females, Puerto Rican males, Latin American females, Mexican American females, Black males, Puerto Rican females, and Black females. College Board, 1991 National Ethnic/Sex Data (College Entrance Examination Board, 1991) (cited in note 20).
\item \textsuperscript{270} The ASVAB reports academic and occupational scores. Both the academic and occupational scores of females of color are particularly low. On the academic composite, race differences predominate; the scores of white students, both male and female, are approximately twice as high as those of Hispanic and Black students. By contrast, on the occupational scores, gender differences are particularly significant, with males outsoring females in their racial/ethnic group on both the mechanical and the electrical sections of the test. See note 43.
\item \textsuperscript{271} See, for example, Cathy Scarborough, Conceptualizing Black Women's Employment Experiences, 98 Yale L J 1457 (1989); Judy Scales-Trent, Black Women and the Constitution: Finding Our Place, Asserting Our Rights, 24 Harv CR-CL L Rev 9 (1989); Kimberlé Crenshaw, Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidis-
“sex-plus discrimination.” “Sex-plus discrimination” has been applied to cases in which an employer singles out a certain subset of women for discrimination. It was first recognized by the Supreme Court in Phillips v Martin Marietta Corp., one of the early important sex discrimination cases decided under Title VII. The Court found a prima facie showing of discrimination where the employer refused to hire females with preschool children although it hired males with preschool children. This finding was in spite of the fact that only some, and not all, women were affected by the policy.

The Fifth Circuit adopted a combined gender and race approach by analogy to the sex-plus theory in Jefferies v Harris County Community Action Association, a Title VII case. The plaintiff, a Black woman, challenged her employer’s failure to promote her to the position of Field Representative. Because the person promoted into the job was a Black man, she could not prove race discrimination. In addition, although the lower court’s findings were not dispositive, statistical evidence appeared to make the sex claim difficult to prove, as it showed that one of the previous Field Representatives was a woman and that women held approximately half of the employer’s supervisory positions. However, the plaintiff did show that every position she applied for had been filled by either a man or a white woman.

In considering the combined claim, the court held that Black women are a protected class under Title VII. “We agree that discrimination against Black females can exist even in the absence of discrimination against Black men or white women.” In reaching this conclusion the Fifth Circuit relied on Phillips v Martin Marietta Corp. and the cases applying the sex-plus theory. It explained:

an employer may not apply different standards of treatment to women with young children, to married women, or to women who are single and pregnant. It is beyond belief that, while an employer may not discriminate against these subclasses of women, he could be allowed to discriminate

272 400 US 542 (1971).
273 Id at 544. See also In re Consolidated Pretrial Proceedings, 582 F2d 1142, 1145 (7th Cir 1978) (invalidating policy requiring that female cabin attendants with children accept ground duty positions); Jacobs v Martin Sweets Co., 550 F2d 364, 371 (6th Cir 1977) (finding Title VII violation where company fired single women who became pregnant); and Sprogis v United Air Lines, Inc., 444 F2d 1194, 1198 (7th Cir 1971) (finding that no-marriage rule for female flight attendants violated Title VII). The decision in Phillips v Martin Marietta, 400 US 542, was cited with approval in UAW v Johnson Controls, Inc., 111 S Ct 1196 (1991).
274 615 F2d 1025 (5th Cir 1980).
275 Id at 1029-30.
276 Id at 1030-31.
277 Id at 1029.
278 Id at 1032.
279 Id at 1033. See cases cited in note 273.
against black females as a class.\textsuperscript{280}

The Tenth Circuit relied on \textit{Jefferies} to reach a similar conclusion in \textit{Hicks v Gates Rubber Co.}\textsuperscript{281} In \textit{Hicks}, the Tenth Circuit remanded a Title VII case so that the district court could aggregate the evidence of racial hostility with the evidence of sexual hostility to determine if the plaintiff, a Black female, had been subjected to a hostile working environment.\textsuperscript{282} Other cases applying a combined race and sex analysis include \textit{Judge v Marsh},\textsuperscript{283} which involved the Title VII claim of a Black woman in the Army, and \textit{Graham v Bendix Corp.},\textsuperscript{284} a Title VII discharge claim brought by a Black woman. Further, in \textit{Chambers v Omaha Girls Club, Inc.},\textsuperscript{285} the Eighth Circuit accepted the district court's combined race and sex analysis of a Title VII claim brought by a Black single woman who became pregnant and was discharged from her position at the Omaha Girls Club.\textsuperscript{286}

Although developed in the Title VII context, this framework provides strong precedential and conceptual support for combining Title VI and Title IX claims where a minority female is subjected to discrimination in education. Whether the name for the claim would be "race-plus," "sex-plus," or "race and sex" discrimination, the important fact is that the multidimensional effect of race and sex discrimination could be considered within the confines of one claim. The fact that two statutes are involved — unlike the analysis under Title VII where one statute addresses both race and sex discrimination — should not be of any import. The legal standards implicated under Title IX and Title VI, along with the underlying purpose of prohibiting discrimination, are identical. Moreover, in \textit{Phillips v Martin Marietta} and its immediate progeny, the "plus" factor was not a statutorily prohibited criterion. Accordingly, under a strict application of that principle, a Title IX claim could be made on behalf of virtually any subset of women, or a Title VI

\textsuperscript{280} Id at 1034.
\textsuperscript{281} 833 F2d 1406 (10th Cir 1987).
\textsuperscript{282} Id at 1416-17.
\textsuperscript{283} 649 F Supp 770 (D DC 1986). The \textit{Judge} court added an additional constraint to the analysis: a plaintiff in a Title VII case can claim only one "plus," otherwise, the statute would "be splintered beyond use and recognition." Id at 780.
\textsuperscript{284} 585 F Supp 1036 (ND Ind 1984). The court held, "[u]nder Title VII, the plaintiff as a black woman is protected against discrimination on the double grounds of race and sex, and an employer who singles out black females for less favorable treatment does not defeat plaintiff's case by showing that white females or black males are not so unfavorably treated." Id at 1047.
\textsuperscript{285} 834 F2d 697 (8th Cir 1987).
\textsuperscript{286} Id at 700 n 9, 701, 701 n 12. Ms. Chambers' claim was ultimately rejected on other grounds. However, the decision put to rest any question regarding the Eighth Circuit's position on this issue arising out of \textit{DeGraffenreid v General Motors}, 413 F Supp 142 (ED Mo 1976), aff'd in part, rev'd in part, 558 F2d 480 (8th Cir 1977). The district court in \textit{DeGraffenreid} had refused to find that Black women are a protected class under Title VII and had required the plaintiffs' race and sex claims to be tried separately. It proceeded to reject the plaintiffs' Title VII claims. The Eighth Circuit affirmed the dismissal on other grounds, noting that "we do not subscribe entirely to the district court's reasoning in rejecting appellants' claim of race and sex discrimination under Title VII." 558 F2d at 484.
STANDARDIZED TESTING

claim could be made on behalf of any subset of members of a racial or national minority group. That being the case, the observation in *Jefferies* that it would be "beyond belief" that discrimination against these groups would not be tolerated but discrimination against minority females could be, is fully appropriate.

In the context of testing discrimination there are broad potential applications of a combined Title IX and Title VI "sex-plus" or "race-plus" theory. These applications include the analysis of the full set of problems confronted by young women of color in connection with uses of the ASVAB and the SAT. The impediment is not so much the lack of availability of a cause of action as it is the lack of information regarding the effects of test usage on minority females. As more is learned about the impact of testing discrimination on girls and women of color, this theory will surely have wide-ranging impact.287

B. Federal Constitutional Analysis

In evaluating the legal constraints concerning discriminatory test uses in education, Fourteenth Amendment equal protection analysis is also a key part of the equation. It applies to actions taken by public schools and universities as well as other public entities.288 Two basic models of equal protection analysis are potentially applicable: "heightened scrutiny" where the practices at issue discriminate on the basis of gender on their face, or where an invidious intent to discriminate on the basis of gender is present; and "rational basis" analysis in other circumstances. Because the discriminatory test uses we have identified are rarely gender-discriminatory on their face but instead present examples of facially neutral practices which nonetheless adversely affect girls and women, a showing of intentional discrimination will often be a necessary element of a claim invoking the heightened scrutiny standard. However, even where invidious intent cannot be demonstrated, rational basis

287 The precedent from Title VII supporting a combination race and sex discrimination claim could also support a combined claim of race and sex discrimination under the Constitution. The constitutional analysis presents additional considerations because, unlike the statutory analysis, race and sex discrimination receive different levels of judicial scrutiny under the Constitution. Nonetheless, a plaintiff pursuing a "sex-plus" or "race-plus" constitutionally based analysis could choose whether to bring her claim under a race or sex theory and might well opt for the race alternative because of the higher scrutiny afforded. At least one commentator would avoid these problems altogether with the argument that combined race and sex claims are entitled to the most rigorous scrutiny. See Scales-Trent, 24 Harv CR-CL L Rev at 23 (cited in note 271) (constitutional claims of Black women may be entitled to more rigorous scrutiny than those of Black men or white women because of their social and historical status).

288 The protections of the Fourteenth Amendment have been read to limit only state action. See, for example, *Moose Lodge No. 107 v Irvis*, 407 US 163 (1972). But, of course, the Fourteenth Amendment limits not only the state government but public education officials as well. See, for example, *Brown v Bd. of Educ.*, 347 US 483 (1954) (finding segregation of public school system in violation of Equal Protection Clause).
promises to be a useful tool in the effort to eradicate gender discrimination in education-related test uses.

1. Heightened Scrutiny

Challenges to gender-based discrimination under the Equal Protection Clause are entitled to a heightened scrutiny standard of review. This standard falls between the rational relationship test, which ordinarily applies in the absence of a protected classification, and the strict scrutiny accorded in cases of discrimination based on race, alienage, national origin, and religion. Under heightened scrutiny, the party relying on the gender-based classification has the burden of showing an "exceedingly persuasive justification" for the classification. Kirchberg v Feenstra, 450 US 455, 461 (1981); Personnel Administrator of Mass. v Feeney, 442 US 256, 273 (1979). The burden is met only by showing at least that the classification serves "important governmental objectives and that the discriminatory means employed" are "substantially related to the achievement of those objectives." Moreover, judicial review must be free of fixed or stereotypical notions concerning the roles and abilities of men and women. Under this standard, many forms of gender discrimination have been found to violate the Fourteenth Amendment's requirement of equal protection.
In the case of facial discrimination, the intent to discriminate — a necessary element of a race- or gender-based equal protection claim — is presumed. Laws or practices which are neutral on their face but which have a disproportionately adverse effect upon women or racial minorities may also violate equal protection under a heightened or strict scrutiny analysis. However, to support this claim, the plaintiff must establish an invidious intent to discriminate. In determining whether the requisite intent exists, discriminatory impact "provides an 'important starting point' [Village of Arlington Heights v Metropolitan Housing Dev. Corp., 429 US 252, 266 (1977)] but purposeful discrimination is 'the condition that offends the Constitution.' [Swann v Charlotte-Mecklenburg Bd. of Educ., 402 US 1, 16 (1971)]."

In order to prove intent, it is not necessary to present the proverbial "smoking gun." In addition to the existence of a disparate impact, other relevant factors include: "the historical background of the decision . . . particularly if it reveals a series of official actions taken for invidious purposes;" departures from the normal procedural or substantive sequence of events which may indicate that improper purposes are playing a role; and the legislative or administrative history of the decision, including contemporaneous statements made by members of the decisionmaking body. In addition, the foreseeability of an adverse impact on a protected group may give rise to an inference of invidious intent.

Nonetheless, as a practical matter, it has been difficult for litigants to show the requisite invidious intent to discriminate in constitutionally based disparate impact cases. This has also been true in the few challenges to educational test uses which have addressed the issue, all of which have been in the context of racial or national origin discrimina-

---

294 Washington v Davis, 426 US at 242, 246, 256; Feeney, 442 US at 272.
295 Feeney, 442 US at 274.
296 As the Supreme Court explained in Washington v Davis, "[n]ecessarily, an invidious discriminatory purpose may often be inferred from the totality of the relevant facts, including the fact, if it is true, that the law bears more heavily on one race than another." 426 US at 242.
298 Id at 267-68. See also De La Cruz v Tormey, 582 F2d 45, 58-59 (9th Cir 1978) (citing same factors).
299 For example, the Court in Feeney held that "when the adverse consequences of a law upon an identifiable group are as inevitable as the gender-based consequences of [Massachusetts' veterans' preference statute at issue], a strong inference that the adverse effects were desired can reasonably be drawn." Feeney, 442 US at 279 n 25. See also Columbus Bd. of Educ. v Penick, 443 US 449, 464 (1979) ("actions having foreseeable and anticipated disparate impact are relevant evidence to prove the ultimate fact, forbidden purpose").
300 For example, in Washington v Davis, 426 US 229, the Court upheld the use of a test administered to applicants for positions as police officers with the District of Columbia, despite the fact that four times as many Blacks as whites failed the test. In Feeney, 442 US 256, the Court similarly rejected a challenge on grounds of gender discrimination to the application of a veterans' preference law which virtually excluded women from many categories of civil service jobs. But see Rogers v Lodge, 458 US 613, 622 (1982) (Supreme Court upheld district court's finding that at-large electoral system in Georgia "was being maintained for the invidious purpose of diluting the voting strength of the black population").
tion. The only cases which have prohibited test uses on these constitutional grounds have been those where there was a history of de jure segregation which continued to affect current students.

2. Rational Relationship

In the absence of a facial classification or the showing of an invidious intent to discriminate, a classification created by a test use will be evaluated under the rational basis test. Under this test, "[t]he State may not rely on a classification whose relationship to an asserted goal is so attenuated as to render the distinction arbitrary or irrational." While the rational basis test certainly provides deference to governmental classifications, it is not a proxy for automatic approval of such classifications. For example, in *City of Cleburne v Cleburne Living Center*, the Court invalidated a zoning ordinance requiring a home for mentally retarded individuals to seek a use permit not required of other multiple dwelling facilities. The Court found that there was no legitimate basis for the ordinance and that it was based merely on an "irrational prejudice against the mentally retarded." Of direct relevance, the *Sharif* court

---

301 In *United States v State of S.C.*, 445 F Supp 1094 (D SC 1977), aff'd as Nat'l Educ. Ass'n v S.C., 434 US 1026 (1978), the Supreme Court summarily affirmed the three-judge district court's conclusion that the state's use of the National Teacher Examination to certify teachers and set salaries did not violate equal protection although the test disqualified 83% of Black applicants as compared to 17.5% of white applicants, 434 US at 1027 (White, J., dissenting), and reproduced the significant pay differentials that had existed between Black and white teachers when the state had a dual segregated system of schools, 445 F Supp at 1104-07. See also *United States v LULAC*, 793 F2d 636, 646 (5th Cir 1986) (no violation in use of "pre-professional skills" test without showing of invidious intent despite a significant disparate impact, the state's awareness that disparate impact was likely before adopting test, a history of racial discrimination in Texas schools, and the failure of the Board to offer remediation courses or take other action to reduce the anticipated disparate effect); *Anderson v Banks*, 520 F Supp 472, 499, 486 (SD Ga 1981) (no invidious intent in use of high school exit examination despite "overwhelming and essentially uncontested" evidence of disparate impact on Black students); and *Larry P.*, 793 F2d 969, 984 ("pervasiveness of discriminatory effect [will not], without more, be equated with... discriminatory intent.

302 In *Anderson v Banks* the court held that the exit examination requirement could not be imposed until those students who were exposed to a segregated educational system had graduated. 520 F Supp at 500-03. Similarly, in *Debra P.*, 474 F Supp at 244, the court enjoined for a period of four years Florida's requirement that students pass a "functional literacy examination" because of the state's history of de jure school segregation. Compare *Vaughn v Bd. of Educ.*, 758 F2d 983, 991-92 (4th Cir 1985) (where Black enrollment was disproportionately high in special education programs and low in gifted programs and district had not attained unitary status, plaintiffs entitled to presumption that disparities were causally related to prior segregation, and burden shifted to defendant to prove otherwise).

303 *Sharif*, 709 F Supp at 364 (quoting *City of Cleburne v Cleburne Living Center*, 473 US 432, 446 (1985)).


found that New York State's use of SAT scores to award scholarships based on high school achievement failed the rational basis test because the SAT was not designed to measure such achievement.\textsuperscript{306}

While \textit{Sharif} is the only case which has applied rational basis analysis to a claim of a gender discriminatory test use, its analysis is surely correct.\textsuperscript{307} Where a test is designed and validated for a purpose unrelated to the use at issue — or it has not been validated at all — there cannot be a rational basis for the demonstrated discriminatory impact which results from that use. Because so many uses to which tests are put in the educational context have not been validated, the rational basis test has a significant potential applicability in this area.

In addition to the facts presented in \textit{Sharif}, a prime example of a practice which would be vulnerable under rational basis analysis is the widespread use of the ASVAB for counseling high school students in connection with vocational education courses. As discussed earlier, the test has not been validated for this purpose. Moreover, scores on particular skill batteries — which reflect substantial gender differentials — have not been correlated with performance in civilian occupations in those areas.\textsuperscript{308} But it is precisely these scores which are used for counseling. This use of the test is not rational. Similarly, the programs for gifted and talented adolescents, which include programs at public universities and which rely on SAT scores to identify candidates for admission, point to no hard evidence to support their use of the SAT. Yet the selection device they have chosen results in nearly twice as many males as females identified as mathematically gifted and offered the opportunity to participate in an enrichment program. Again, the rational basis for this test use is highly questionable.

In sum, while heightened scrutiny likely has a limited applicability to the matters at hand, a number of test uses may well be vulnerable under rational basis analysis.

C. State Constitutional Guarantees and Laws Prohibiting Gender Discrimination in Education

The legal analysis of gender bias in educational test uses is by no means limited to the federal constitutional and statutory provisions already discussed. State equal rights amendments (ERAs), the equal protection clauses of state constitutions, and statutory and regulatory provisions may well also prohibit gender-discriminatory testing.

\textsuperscript{306} \textit{Sharif}, 709 F Supp at 364.
\textsuperscript{307} Compare \textit{Student Doe v Pa.}, 593 F Supp 54, 57 (ED Pa 1984) (with no discussion of the test at issue, the court concluded that a testing procedure for admitting students into a gifted class "may not be perfect, indeed it may not be the best method available, but the Court is unable to conclude that it is a method that cannot reasonably be used").
\textsuperscript{308} See discussion at notes 95-99 and accompanying text.
Although there is virtually no case law applying these provisions to educational testing, their strong facial guarantees of equity provide an important source of law to complement and expand upon Title IX and federal equal protection analysis.

1. State Constitutional Guarantees

a. Equal Rights Amendments

State equal rights amendments (ERAs) have important implications for the analysis of gender-discriminatory test uses. Prompted by action towards a federal ERA, fourteen states adopted ERAs as part of their state constitutions in the 1970s. Two other states — Utah and Wyoming — adopted ERAs in the 1890s to support those states' progressive guarantees of women's suffrage. Most of these provisions are similar to the proposed federal Equal Rights Amendments. Others also prohibit discrimination on the basis of race, color, or national origin. Most of the state ERAs incorporate some state action requirement, although the natures of the requirements differ. Some states do not impose the rigorous requirements for a finding of state action which exist under the federal Constitution.

The interpretation of state ERAs varies widely among the different states, with some states rigorously enforcing the prohibition of gender discrimination.

---

309 After being introduced in every Congress since 1923, in 1972 a federal ERA was approved by overwhelming majorities in both the House of Representatives and the Senate and submitted to the states for ratification. Beth Gammie, *State ERAs: Problems and Possibilities*, 1989 U Ill L Rev 1123, 1124 (1989). The first section of the Amendment provides: "Equality of rights under the law shall not be denied or abridged by the United States or by any State on account of sex." HR J Res 208, 92d Cong, 1st Sess (1971); S J Res 8, 92d Cong 1st Sess (1971). The ERA had to be ratified by 38 states in order to be adopted as a Constitutional Amendment. In June of 1982 the deadline for ratification passed with only 35 states ratifying. Gammie, 1989 U Ill L Rev at 1124.

310 The states are: Alaska, Alaska Const Art I, § 3 (1972); Colorado, Colo Const Art II, § 29 (1972); Connecticut, Conn Const Art I, § 20 (1974); Hawaii, Hawaii Const Art I, § 3 (1972); Illinois, Ill Const Art I, § 18 (1971); Maryland, Md Const Art 46 (1972); Massachusetts, Mass Const part I, Art I (1976); Montana, Mont Const Art II, § 4 (1972); New Hampshire, NH Const part I, Art II (1974); New Mexico, NM Const Art II, § 18 (1973); Pennsylvania, Pa Const Art I, § 28 (1971); Texas, Tex Const Art I, § 3(a) (1972); Virginia, Va Const Art I, § 11 (1971); and Washington, Wash Const Art 31, § 1 (1972).

311 Utah Const Art IV, § 1 (1896), and Wyo Const Art 1, §§ 2, 3 and Art 6, § 1 (1890).


313 See, for example, the ERAs of Colorado, Hawaii, Illinois, Maryland, New Mexico, Pennsylvania, and Washington (cited in note 310).

314 See, for example, the ERAs of Alaska, Connecticut, Massachusetts, New Hampshire, Montana, Texas, and Virginia (cited in note 310).

315 See, for example, *Hartford Accident & Indem. Co. v Ins. Comm'n of the Commonwealth of Pa.*, 505 Pa 571, 586, 482 A2d 542, 549 (1984) ("[t]he rationale underlying the 'state action' doctrine is irrelevant to the interpretation of the scope of the Pennsylvania Equal Rights Amendment, a state constitutional amendment adopted by the Commonwealth as part of its own organic law").
Most courts impose at least “strict scrutiny” analysis to gender classifications. This analysis is more stringent than the lower, intermediate level of scrutiny afforded gender classifications under the federal Constitution. Under strict scrutiny analysis, a gender-based classification can only be sustained upon the showing of a compelling state interest. Experience with the strict scrutiny test under the Fourteenth Amendment teaches that this is an exceedingly difficult burden to sustain.

Some states, including Pennsylvania, Washington, Colorado, and Maryland, have adopted a “strict scrutiny plus” or “absolute standard” which tolerates virtually no gender-based discrimination at all. A few states have simply followed the federal constitutional standard — or applied a lower level of scrutiny than the federal standard — thus nullifying the value of any extant ERA in these states. State ERAs are an important source of increased protection against discrimination for women in those states where they are interpreted to impose more than an intermediate level of scrutiny.

While most state ERA cases focus on explicitly sex-based classifications, a small number have addressed the appropriate analysis of facially neutral policies with a disproportionate impact on one sex. For example, in Massachusetts and in Maryland, where courts rigorously review claims of sex discrimination under the states’ ERAs, courts have held that official actions which are neutral on their face can violate the ERA if they are discriminatory as applied without an additional showing of

---

317 See, for example, Attorney Gen. v Mass. Interscholastic Athletic Ass’n, 378 Mass 342, 393 NE2d 284, 291 (1979); People v Ellis, 57 Ill 2d 127, 311 NE2d 98, 101 (1974); and Mercer v Bd. of Trust, North Forest Indep. School Dist., 538 SW2d 201, 204-05 (Tex Civ App 1976).
318 See, for example, City of Cleburne, 473 US at 440 (laws that employ a suspect classification so seldom incorporate compelling state interests that they are “deemed to reflect prejudice and antipathy” such that they are subject to strict scrutiny).
320 See, for example, Archer & Johnson v Mayes, 213 Va 633, 194 SE2d 707, 710 (1973) (applying a rational basis standard to the state ERA); Stanton v Stanton, 30 Utah 2d 315, 517 P2d 1010, 1012 (1974) (applying rational basis analysis to federal equal protection-type issue), rev’d on federal equal protection grounds, 421 US 7 (1975); and Dynyn v Dept. of Liquor Control, 12 Conn App 455, 531 A2d 170, 175 (1987) (applying an intermediate standard of review).
321 See, for example, Hopkins v Blanco, 457 Pa 90, 320 A2d 139 (1974) (common law rule that a wife had no right to recover for loss of her husband’s consortium violates Pennsylvania ERA); Opinion of the Justices to House of Rep., 374 Mass 836, 371 NE2d 426, 429-30 (1977) (bill prohibiting women from participating in contact sports with men violates Massachusetts ERA); Kline v Ansell, 287 Md 585, 414 A2d 929, 933 (1980) (common law rule that only a man could sue or be sued for criminal conversation violates Maryland ERA); Darrin, 540 P2d at 893 (athletic association rule forbidding girls to play on all-male high school football teams violates Washington ERA); and Colo. Civ. Rights Comm’n v Travelers Ins. Co., 759 P2d 1358, 1361 (Colo 1988) (group health insurance policy which excluded from coverage medical expenses associated with normal pregnancy violates Colorado ERA).
The implications for the analysis of testing discrimination are substantial. Test uses which result in disparate impacts on the basis of sex—including admissions to post-secondary institutions and special programs, criteria for granting scholarships, and uses of tests in counseling or admissions to vocational education programs—would have to be justified, at a minimum, as serving a compelling state interest.

b. Equal Protection

In addition to the state ERAs, at least two state constitutional guarantees of equal protection have been held to provide stronger protections against gender discrimination than are found in the Fourteenth Amendment. Courts in Oregon and California have clearly found that sex-based classifications are suspect under their state constitutions and are entitled to strict scrutiny. Courts in several other states appear to have adopted this conclusion, although their holdings are not as definitive.

Furthermore, in California, the strict scrutiny standard has supported holdings that sex-based disparate impact discrimination is actionable without the proof of invidious intent. This result has been alternatively framed under the analysis that intent to discriminate may be inferred from the legislature's awareness of the obvious consequences of legislation. Under either analysis, however, public policies which

---

323 See Buchanan v Dir. of Div. of Empl. Sec., 393 Mass 329, 471 NE2d 345, 348-49 (1984) (citing School of Braintree v Mass Comm'n Against Discrimination, 377 Mass 424, 386 NE2d 1251, 1255-56 (1979) (court held that disparate impact discrimination was actionable, but rejected plaintiff's claim because the record was devoid of any factual findings of disparate impact)); Burning Tree Club, Inc. v Bainum, 305 Md 53, 501 A2d 817 (1985) (striking down under the Maryland ERA an exception to a law prohibiting sex discrimination that permitted clubs whose facilities operate with the primary purpose of serving or benefiting members of a particular sex, despite the alleged neutrality of the provision).

324 See, for example, Hewitt v State Accid. Ins. Fund Corp., 294 Or 33, 653 P2d 970, 975-79 (1982) (gender classification of workers' compensation statute invalidated under section of state constitution prohibiting laws granting privileges to any citizen not belonging to all citizens); Sail'er Inn, Inc. v Kirby, 5 Cal 3d 1, 95 Cal Rptr 329 (1971) (invalidating under state equal protection clause a law excluding most women from bartending).

325 See, for example, Moffett v Zitvogel, 1990 WL 123068 (Del Super Ct); Hanson v Williams County, 389 NW2d 319, 323 n 9 (ND 1986). See also S.W. Wash Ch., Nat'l Elec. Cont Ass'n v Pierce County, 100 Wash 2d 109, 128 n 3, 667 P2d 1092, 1102 n 3 (Wash 1983) (en banc) (while Washington constitution has in the past been construed to impose a strict scrutiny test for sex-based classifications, the state ERA alone now governs review of such classifications).

326 For example, in Hardy v Strumpf, 37 Cal App 3d 958, 112 Cal Rptr 739 (1974), the plaintiffs challenged the Oakland Police Department's facially neutral height and weight requirements for the position of patrol officer. The court held that because sex-based classifications are suspect, "a seemingly neutral job requirement which has the effect of disqualifying a disproportionate number of one sex is discriminatory and must be viewed under the strict scrutiny test." 112 Cal Rptr at 743. Relying on statistics that showed that over 80% of all American women were effectively excluded by the height and weight requirements, the court found disparate impact. 112 Cal Rptr at 743-44. The court further found that the police department failed to show that the requirements were "demonstrably related to job performance." 112 Cal Rptr at 745. Accordingly, the court struck down the requirements, holding that "[i]t is not necessary to conclude that these standards were adopted with intent to discriminate. . . ."

112 Cal Rptr at 743.

STANDARDIZED TESTING

rly on classifications giving rise to gender-based disparate impact are actionable without the onerous burden of proving intent as contemplated under federal constitutional law.328

As was just discussed in connection with the state ERAs, broad constructions of state equal protection guarantees may well lead to viable discriminatory test uses. Given the fact that gender differentials in scoring are well-established in a number of standardized tests ranging from the ASVAB to the SAT, the use of such scores to allocate benefits would easily lead to an inference of intentional discrimination under the California analysis. Furthermore, it would likely be difficult for a test-user to successfully defend its use by demonstrating the requisite compelling state interest under a strict scrutiny analysis. Indeed, the strict scrutiny standard would almost surely exceed the Title IX standard of "educational necessity" under either a Griggs or a Wards Cove formulation.

2. State Educational Equity Laws

In addition to state ERAs and equal protection guarantees, state educational equity laws also provide an important resource for challenges to discriminatory test use. A number of states — including Alaska, California, Florida, Hawaii, Iowa, Maine, Massachusetts, New Jersey, Oregon, Rhode Island, Washington, and Wisconsin — have enacted laws modeled after Title IX which prohibit sex discrimination in educational programs or institutions receiving state or county financial assistance.329 Other states have human or civil rights laws prohibiting sex discrimination in educational institutions or in public accommodations, defined broadly to include educational institutions.330

328 Feeney, 442 US at 272. Other states also interpret their constitutions in a less rigid and formalistic manner than the Supreme Court has interpreted the Fourteenth Amendment. See, for example, Colo. Civ. Rights Comm'n v Travelers Ins. Co., 759 P2d 1358 (Colo 1988) (rejecting the Supreme Court's holding in Geduldig v Aiello, 417 US 484 (1974), that discrimination on the basis of pregnancy does not constitute unconstitutional sex discrimination).


Some of these laws, and the regulations promulgated pursuant to them, are more explicit and far-reaching than is Title IX.\textsuperscript{331} Thus, in Washington, the educational equity regulations specifically warn school districts that compliance with the Title IX alone may not constitute compliance with state regulations which extend beyond Title IX.\textsuperscript{332} For example, the counseling and guidance regulations require "reasonable efforts" to encourage students to consider and explore "nontraditional" occupations\textsuperscript{333} and mandate training sessions on eliminating sex bias.\textsuperscript{334} In the area of physical education, where the use of a particular standard has an adverse effect on the basis of sex, the school district must immediately replace the standard with one that does not have such an effect, even if the standard with an adverse effect has been validated.\textsuperscript{335} These requirements all go beyond those mandated by Title IX and its regulations.

Other examples of such laws are found in Oregon and Massachusetts. In Oregon, "discrimination" is explicitly defined in the statute to include "any act that is fair in form but discriminatory in operation."\textsuperscript{336} This closely tracks the Supreme Court's formulation of disparate impact discrimination in \textit{Griggs}. In Massachusetts, admissions standards are closely and explicitly regulated.\textsuperscript{337} For example, a prerequisite requirement must be "essential to success in a given program."\textsuperscript{338} And if access mechanisms have limited the opportunities of a class of students to participate in the prerequisite, then the students must be allowed to enter the program without the prerequisite or must be admitted to the prerequisite.\textsuperscript{339}

The enforcement mechanisms of the different educational equity laws and civil and human rights statutes vary, but may generally allow for the termination of state funds to a school that is not in compliance,\textsuperscript{340}

\begin{itemize}
\item \textsuperscript{331} These examples are illustrative only and are not a comprehensive list of all of the aspects of state law that may be more favorable for a plaintiff than Title IX or the federal constitution.
\item \textsuperscript{332} Wash Admin Code § 392-190-005 (1986).
\item \textsuperscript{333} Id at § 392-190-015. Similarly, California requires guidance counselors to "affirmatively explore" with pupils the possibility of courses leading to nontraditional careers. Cal Educ Code § 40(d) (West Supp 1991). Alaska also encourages nontraditional career counseling and requires regular training in recognizing and overcoming the effects of gender bias. 4 Alaska Admin Code § 06.530(a), (b) (July 1988).
\item \textsuperscript{334} Wash Admin Code § 392-190-020 (1986).
\item \textsuperscript{335} Wash Admin Code § 392-190-030 (1986).
\item \textsuperscript{336} Or Rev Stat § 659.150(1) (1989).
\item \textsuperscript{337} See, for example, 603 Mass Admin Code § 26.02(4), (5) (school admissions); 26.03(1) (course admissions); and 26.04(3) (guidance materials) (1986).
\item \textsuperscript{338} Id at § 26.03(1) (emphasis added).
\item \textsuperscript{339} Id. See also Cal Educ Code § 212.5, 230 (West Supp 1991), which specifically define sexual harassment and prohibit it in all aspects of the educational process.
\item \textsuperscript{340} See, for example, Wash Admin Code § 392-190-080(a) (1986).
\end{itemize}
the filing of administrative complaints, and private suits for equitable relief or damages.

These state provisions can provide a strong basis for challenges to discriminatory test use.

D. Remedies for Gender-Discriminatory Uses of Educationally Related Tests

Both governmental and private remedies are available to enforce legal prohibitions against gender discrimination in educational testing. Governmental remedies, which come into play under Title IX and certain state statutes, include defunding recipients of federal and/or state financial assistance and administrative or judicial proceedings to enforce the law through compliance actions and injunctive or declaratory relief. Private remedies include injunctions prohibiting or restricting the use of invalid tests, requiring that tests be validated, or mandating that new selection devices or mixes of devices be implemented. Prospective affirmative relief may also be available to eliminate the discriminatory effects of an invalid test use. Finally, attorneys fees are available to prevailing parties and, in certain circumstances, monetary damages may be awarded.

1. Government Enforcement and Remedies

Title IX is enforced publicly, primarily through the Office for Civil Rights (OCR) in the Department of Education, as well as through a private right of action. In the event that the Department of Education finds a violation of Title IX which the recipient fails to correct, it has the statutory authority to defund a recipient of federal financial assistance. The Department of Education has not, to date, invoked the defunding remedy in a Title IX case. Nonetheless, the defunding remedy remains statutorily available and there is no reason to believe that it would never be invoked, especially in the case of an egregious violation.

---

341 See, for example, 603 Mass Admin Code § 26.09 (1986).
342 See, for example, 603 Mass Admin Code § 26.10 (1986). Many civil and human rights laws also allow for damage actions.
343 OCR also has responsibility for enforcing, inter alia, Title VI, § 504 of the Rehabilitation Act, and the Age Discrimination Act. In addition to OCR, other federal agencies have the responsibility for enforcing Title IX requirements among their grantees. However, because Title IX is limited to discrimination in education, primary enforcement is through the Department of Education.
345 20 USC § 1682 (1991) states, “Compliance with any requirement adopted pursuant to this section may be effected (1) by termination of or refusal to grant or to continue to grant or to continue assistance under such program or activity . . . or (2) by any other means authorized by law.”
346 In Storey v Bd. of Regents of U of Wis. Sys., 604 F Supp 1200, 1202 (WD Wis 1985) the court denied the defunding remedy as a cumbersome, costly, and extreme remedy, only to be used as a last resort where no other relief is available.
which the fund recipient refuses to correct.\textsuperscript{347} The Title IX regulations adopt and incorporate the procedural provisions governing the enforcement of Title VI to regulate defunding proceedings.\textsuperscript{348}

In addition to the ultimate remedy of defunding, the Department of Education through OCR has access to a number of other mechanisms to enforce Title IX.\textsuperscript{349} OCR has the authority, among other things, to conduct compliance reviews and investigate complaints, to make findings of non-compliance with the law, to conciliate claims, and to refer cases to the Department of Justice for judicial enforcement.\textsuperscript{350} Further, the Title IX regulations give OCR the specific regulatory authority to require a recipient of federal funding to take "such remedial action as the Assistant Secretary deems necessary to overcome the effects of . . . [gender] discrimination."\textsuperscript{351} OCR has relied on these mechanisms to varying degrees.\textsuperscript{352}

While OCR has broad authority to act to eradicate sex discrimination in education, serious questions regarding the effectiveness of its enforcement activities have been raised.\textsuperscript{353} Moreover, long-standing judicially imposed requirements designed to assure the timely resolution of complaints and compliance reviews have been vacated.\textsuperscript{354} In addition, neither OCR nor any other federal entity has issued regulations or guidelines or developed other policy guidance to address the issue of discrimination in educational testing, even though OCR has been presented with testing questions in the course of complaints and compliance reviews, most notably in connection with vocational education programs.\textsuperscript{355}

\textsuperscript{347} The Department of Education recently cut off the funds of the DeKalb County School District after the school district refused OCR access to investigate complaints about the district's policy for educating disabled students. \textit{Freeman v Cavazos}, 923 F2d 1434, 1435-36 (11th Cir 1991). A challenge by the school district to the Department's action is pending in the Eleventh Circuit. Id at 1436-37. In \textit{Freeman}, the school district's emergency motion for stay of the Department's action was denied. Id at 1441.

\textsuperscript{348} 34 CFR § 106.71.

\textsuperscript{349} Again, the Title IX regulations incorporate the Title VI procedures. Id.

\textsuperscript{350} 34 CFR §§ 100.7, 100.8. Complaints may be filed by third parties in addition to being filed by injured parties. 34 CFR § 100.7(c).

\textsuperscript{351} 34 CFR § 106.3(a).

\textsuperscript{352} See note 349.

\textsuperscript{353} See, for example, Report on Investigation of the Civil Rights Enforcement Activities of the Office for Civil Rights, US Department of Education, Majority Staff of House Committee on Education and Labor, 100th Cong, 2d Sess (Comm Print 1988). This report concluded, "Since 1981, the Office for Civil Rights of the Department of Education has been stymied by an administration which actively opposed the laws which were entrusted to it and took efforts to minimize the agency's potential impact. As a consequence, the OCR has been beset with confused policy directives, administrative mismanagement, numerous changes in leadership, and severe reductions in resources." Id at 6.

\textsuperscript{354} \textit{WEAL v Cavazos}, 906 F2d 742 (DC Cir 1990).

\textsuperscript{355} The only relevant document produced by OCR in response to a request under the Freedom of Information Act for documents relating to these matters was an OCR memorandum, dated Apr 16, 1982, regarding "Methodology for Analyzing Admissions Programs in Institutions of Higher Education." The memo peripherally addressed issues of validation, but has been withdrawn and designated an "historical document." Antonio J. Califà, \textit{Methodology for Analyzing Admissions Programs in Institutions of Higher Education} (Dept. of Educ., 1982).
result, the government lacks any coherent policy or enforcement activities regarding testing discrimination.

A prime example of this lack of coherent policy and enforcement has been OCR’s treatment of the ASVAB. In the course of the investigations of several complaints and compliance reviews, OCR determined that secondary schools which administered the ASVAB were using scores that were separately normed by gender. That is, the scores of young women were determined by comparison of their performance only to the performance of other female test-takers, and young men’s test scores represented a comparison only to the scores of other male test-takers. Without addressing the underlying validity questions presented in the widespread use of the ASVAB and without exploring the link of the demonstrated score differentials in the ASVAB to the widespread gender segregation in vocational education programs, OCR concentrated exclusively on the separate norming question. Regarding this narrow question, OCR first found that separate norming constituted a violation of Title IX. However, after consultation with the Department of Defense, which publishes and distributes the ASVAB, OCR concluded that, in fact, separate norming was permissible.

An additional concern is raised by OCR’s failure to treat the question of the double discrimination faced by minority women. Although OCR has jurisdiction over both Title IX and Title VI, and in spite of the development of a legal analysis regarding this question, it has taken no steps to explore their interplay.

In sum, while federal administrative enforcement is technically available under Title IX, the administrative forum currently appears to offer little likelihood of meaningful practical relief to victims of gender bias in testing.

2. Private Enforcement and Remedies

In addition to governmental enforcement, the federal legal rights

---

356 See Memorandum from Richard D. Komer, Deputy Assistant Secretary for Policy, US Dept. of Educ., to OCR senior staff (Jun 5, 1990) (regarding ASVAB separate norming) (on file with the National Women’s Law Center) (“Komer Memorandum”).

357 See discussion at notes 91-107 and accompanying text.

358 See discussion at notes 36-53 and accompanying text.

359 See Komer Memorandum (cited in note 356). In another example, in a compliance review of vocational counseling/testing services that was triggered by disproportionate enrollment patterns at a community college, OCR concluded that although interest inventories were being used and were of little value, there was no violation because the tests were not required for admission. See Letter from Office of the Regional Director, Region VII to Dr. Howard Fryett, President, Flathead Valley Community College 6 (Mar 1, 1988) (on file with the National Women’s Law Center).

360 See discussion at notes 271-87 and accompanying text.

361 As discussed at notes 329-42 and accompanying text, state administrative remedies also provide a variety of enforcement mechanisms, including defunding. Actual enforcement practices vary by state and should be considered carefully by potential claimants.
outlined above are privately enforceable. Only Sharif has directly addressed the question of the appropriate remedy to be granted in a case involving gender-based testing discrimination. However, based on remedies issued in relevant non-testing cases, as well as remedies issued in testing cases brought pursuant to Title VI and Title VII, private litigants who prove discrimination may look to a range of remedies. These remedies principally include injunctions prohibiting or restricting the use of an invalid test or requiring that a test be validated or that a valid selection procedure be developed. In addition, depending on the facts and circumstances, courts may order affirmative relief to eliminate the discriminatory effects of an invalid test use. Monetary relief may be granted when certain criteria are met. Finally, attorneys fees are available to prevailing plaintiffs.

a. Injunctions Barring or Restricting the Use of Discriminatory Tests or Requiring Adoption of Valid Selection Procedures

The most straightforward remedy, and the one with the most consistent applicability to the questions under consideration here, is to enjoin the use of test scores which result in discrimination or to impose restrictions on their use in order to eliminate the discrimination. This approach was taken in Sharif. Upon finding that New York State's reliance on SAT scores as the sole criterion for awarding scholarships was in violation of Title IX, the court enjoined the further use of the SAT scores in this fashion. Instead, it required that the state use a combination of SAT scores and grade point averages to determine the scholarship winners, reasoning that the SAT scores would compensate for the differences in grading policies of different schools.

The Sharif injunction is well grounded in both Title VI and Title VII testing law. For example, in Larry P., which was decided under Title VI, the Ninth Circuit upheld the district court's injunction barring the non-validated use of IQ tests to place children in classes for the educable mentally retarded where the test use resulted in a significant discriminatory impact on the basis of race. Similar examples are common in Title VII cases. In Guardians, for example, the court enjoined the further use of an entry level exam for police officers which discriminated impermissibly against minorities, except on an interim

---

362 The Supreme Court has held that Title IX contains an implied right of action without any administrative exhaustion requirement. Cannon, 441 US at 709. Further, Fourteenth Amendment rights are privately enforced through the Civil Rights Act, 42 USC § 1983, which also contains no exhaustion requirement, Patsy v Bd. of Regents of State of Fla., 457 US 496 (1982).
363 709 F Supp 345.
364 Id at 363. The district court did not limit the alternatives to SAT scores the state could use in the future, including the development of a valid statewide achievement test. Id at 354-65.
365 793 F2d 969.
366 Id at 984.
STANDARDIZED TESTING

basis with adjustments that would avoid the test's disparate impact.367

In employment cases, courts have also ordered that tests be validated or replaced with valid selection procedures, often in conjunction with the issuance of prohibitory injunctions.368 A case in point is Berkman v City of New York, where the Second Circuit affirmed the district court's order that a new properly validated physical portion of the firefighter selection exam with the least adverse impact on women be prepared to replace the exam held to be invalid.369 These orders are valuable not only because they eliminate a discriminatory selection practice but also because they assure, through the court's continued jurisdiction, that a valid and nondiscriminatory practice will be instituted in its place.370

In addressing the establishment of valid selection procedures, at least one court has recognized that a test may have a different predictive validity for different population groups and that eliminating this differential in the continued use of the test is an appropriate subject for relief.371 In Kirkland v New York State Dept. of Correctional Services372 the court approved the addition of 250 points to the raw scores of minority test-takers where it was shown that the test had a different predictive value for minority and non-minority test-takers.373 Because minority test-tak-

367 630 F2d 79, 109 (2d Cir 1980). See also Berkman v City of N.Y., 705 F2d 584, 586 (2d Cir 1983) ("Berkman") (use of eligibility list derived from invalid fire department entrance exam which discriminated against women enjoined except for cases of "compelling necessity"); Easley v Anheuer-Busch, Inc., 758 F2d 251, 273 (8th Cir 1985) (discriminatory test-based hiring practices enjoined); and Vulcan Pioneers, Inc. v N.J. Dept. of Civ. Service, 832 F2d 811 (3d Cir 1987) (use of eligibility lists based on invalid exam enjoined).

368 See, for example, Guardians, 630 F2d at 108 ("an appropriate compliance remedy . . . may properly assure the establishment of a lawful new procedure").

369 Berkman, 705 F2d at 588. See also Firefighters Inst for Racial Equality v St. Louis, 616 F2d 350 (8th Cir 1980) (new exam for promoting firefighters to hire captains must be developed unless, on remand, further evidence of current test's validity was shown); and Morrow v Crisler, 491 F2d 1053 (5th Cir 1974) (en banc) (tests used in connection with hiring state highway patrol officers must be validated).

370 Courts have reached different results on the question of how specific they may be in requiring that tests meet particular standards of validity. Compare, for example, Guardians, 630 F2d at 110 (district court went too far in requiring that newly developed tests comply with EEOC's Uniform Guidelines and the American Psychological Association Standards) with Firefighters Inst. for Racial Equality, 616 F2d at 363 (upheld order that new test must be consistent with EEOC Guidelines).

371 As the Guardians court explained, "[o]nce an exam has been adjudicated to be in violation of Title VII, it is a reasonable remedy to require that any subsequent exam or other selection device receive court approval prior to use." 630 F2d at 109.

372 See notes 80-85 and accompanying text for discussion of differential validity.

373 628 F2d 796 (2d Cir 1980).

374 Id at 798. The Civil Rights Act of 1991 makes it an unlawful employment practice under Title VII "for a respondent, in connection with the selection or referral of applicants or candidates for employment or promotion to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment related tests on the basis of race, color, religion, sex, or national origin." Civil Rights Act of 1991, 105 Stat at 1075. The floor statement of Senators John C. Danforth, R-Mo and Ted Kennedy, D-Mass, the chief Senate sponsors, explains that "[b]y its terms, the provision applies only to those tests that are 'employment related.' Therefore, this section has no effect in disparate impact suits that raise the issue of whether or not a test is, in fact, employment related. The prohibitions of this section only become appli-
ers with lower scores performed on the job just as well as non-minority test-takers with higher scores, the point addition served to make the test race-neutral.\textsuperscript{375} This approach comports with the holding in \textit{Larry P.} that the test use at issue would have to be validated separately for Black students.\textsuperscript{376}

Other cases have considered the revision of scoring methods in order to eliminate discriminatory impact without compromising the validity of the test use. The remedy in \textit{Sharif}, which required a combined use of SAT scores and grades, is a prime example. Most other applications of remedial score adjustments have been in connection with rank-ordering\textsuperscript{377} which, as discussed above, has only a limited applicability to education testing. There are implications for adjustments to cutoff scores as well. For example, in \textit{Guardians}, after determining the particular cutoff score to be invalid, the court held that “the City may use a cutoff score somewhat lower” than what it had been using to accomplish the proper interim hiring.\textsuperscript{378} However, where an invalid cutoff score is symptomatic of broader problems with the test, courts have structured broader relief.\textsuperscript{379}

cable once a test is determined to be employment related.” 113 Cong Rec S15484 (Oct 30, 1991). The same language appears in the statement of Rep Don Edwards, D-Cal, a chief House sponsor, at 137 Cong Rec H9529 (Nov 7, 1991). As such, score adjustments will continue to be permitted under Title VII as a remedy where tests are found to be in violation of the law. Because the limitation is Title VII-specific, it should not affect remedies under Title IX or any other statute.

\textsuperscript{375} 628 F2d at 798.

\textsuperscript{376} \textit{Larry P.}, 793 F2d at 980 (but note that this analysis contributed to finding of liability rather than formulation of relief).

\textsuperscript{377} In \textit{Kirkland v N.Y. State Dept. of Correctional Services}, 711 F2d 1117 (2d Cir 1983), an interim remedy calling for the adjustment of a rank-ordered eligibility list into a tiered zone system, in which all candidates in a particular scoring zone had the same rank, was approved. See also \textit{Reid v State of N.Y.}, 370 F Supp 1003 (SD NY 1973). But see \textit{Berkman v City of N.Y.}, 812 F2d 52 (2d Cir 1987) (“Berkman I”), where the court rejected the lower court’s ordered changes to the order of a new entry level firefighters physical test which had resulted in only two women scoring in the top 6,500 applicants. The court found that the proposed changes which included collapsing a seven-band scoring system into a three-band system and renorming the test neither added to test validity nor added any women to the eligibility list, id at 60, and served to lessen the differentiating power of the test, id at 61. Moreover, the court rejected the use of a “compensation ratio” for women on the grounds that it was inappropriate affirmative relief. Id at 62. See notes 343-61 and accompanying text for discussion of affirmative remedies.

\textsuperscript{378} \textit{Guardians}, 630 F2d at 113.

\textsuperscript{379} See, for example, \textit{Ass’n Against Discrimination In Employment, Inc. v City of Bridgeport}, 647 F2d 256 (2d Cir 1983), aff’d, 710 F2d 69 (2d Cir 1983). The Second Circuit approved the district court’s finding that the lowering of an invalid cutoff score was not enough to cure an otherwise invalid test use. Instead, the court ordered the city to prepare a list of minority persons to be offered positions as firefighters and to actively recruit minority firefighters. See also \textit{Burney v Pawtucket}, 559 F Supp 1089, 1104-05 (D RI 1983) (court directed the city to prepare new physical screening procedures and to delete graduation requirement which incorporated previous physical test, after holding the cutoff scores and the physical test invalid); and \textit{San Francisco Police Officers Ass’n v San Francisco}, 812 F2d 1125 (9th Cir 1987), vacated as moot, 842 F2d 1126, 1132 (9th Cir 1988), rev’d, 869 F2d 1182 (9th Cir 1988) (Ninth Circuit struck down a revision of the cutoff and weighing procedures for a police department promotions exam to improve minority scores on the grounds that the procedures unnecessarily trammeled the rights of non-minorities; court held that a less burdensome alternative,
Injunctive remedies of the types described have a broad applicability to the gender-discriminatory use of test scores in educational settings. As in *Sharif* and *Larry P.*, the continued use of tests which result in a disparate impact on the basis of gender and have not been shown to be valid for the use at issue should not be tolerated. The test use must either cease or be revised so that it is no longer discriminatory. For example, if it is proven that the ASVAB does not validly predict success in or aptitude for vocational education programs or careers, an injunction prohibiting its use in selection for these programs is fully appropriate. Similarly, a use of the SAT which is not supported by validity evidence and which has a disparate impact on the basis of gender should be enjoined.

Orders requiring the substitution of valid selection methods, including both tests and other selection devices for those which are determined to be discriminatory, also have a direct application. Just as in employment cases, the continued jurisdiction of the court will assure that non-discriminatory test uses are implemented in an effective and lawful fashion.

Remedies designed to cure demonstrated discrimination by modifying the weighting or scoring of a test, such as the relief granted in *Sharif* where the state was ordered to combine grade point averages with test scores, or in *Kirkland* where points were added to the scores of minority test-takers to eliminate racial bias, also provide relevant models. The *Kirkland* framework, in particular, has interesting ramifications for education-related tests which predict differently by gender, such as the SAT. If such differential prediction is proven in the context of a discrimination case, an invalid use of the SAT or other test could be remedied by adjusting the scores of female and male test-takers so that they bear the same predictive relationship to the criterion at issue. This purpose would most likely be accomplished simply by adjusting upward the scores of female test-takers so they achieved the same correlation as the male scores.

b. Remedies Addressing the Continuing Effects of Past Discrimination

In addition to enjoining the use of discriminatory tests and requiring the development of valid selection devices, courts have also addressed the question of structuring relief to remedy the effects of past discrimination in testing.\(^3\) For example, in *Larry P.*, the Ninth Circuit affirmed the

\(^3\) As Justice Brennan wrote in the plurality decision in *Local 28 of Sheet Metal Workers' Int'l Ass'n v EEOC*, 478 US 421, 445 (1986), Title VII does not “prohibit a court from ordering, in appropriate circumstances, affirmative race-conscious relief as a remedy for past discrimination. Specifically, we hold that such relief may be appropriate where an employer or a labor
district court's order enjoining the use of IQ tests which were not valid for this use, ordering the re-evaluation of every current Black educable mentally retarded (EMR) student, requiring that every school district with a racially disproportionate EMR program devise a three-year remedial plan, and ordering school districts to bring to the court's attention any disparities which persisted at the end of this period. As another example, in a case dealing with long-standing discriminatory testing which excluded many women who sought to become firefighters, New York City was required to reserve entry-level firefighter positions for up to 45 qualified female firefighters. Similarly, in the face of a history of racial discrimination, the city of St. Louis was ordered to promote eight qualified Black firefighters to the position of captain pending the development of a valid promotion exam. Generally, affirmative remedies in Title VII testing cases have been limited to the time period between the finding that a test or other selection procedure is invalid and the institution of a valid procedure, although longer-term remedies have also been imposed.

Because the availability of affirmative relief is fact-specific, it is difficult to predict where such relief may come into play in gender bias in educational testing cases. However, where a discriminatory test use is long-standing and/or has inflicted a serious injury on female test-takers, such relief should be seriously considered. The nature of the relief could include the Larry P. model of devising remedial plans to eliminate the effects of the bias. It could also include, for example, requirements that, at least over the short term, a certain number of spaces in a particular college or other program or a certain number of scholarships in a competitive scholarship program should be reserved for qualified females.

c. Voluntary Remedies

Courts have also been called upon to consider voluntary plans and

union has engaged in persistent or egregious discrimination, or where necessary to dissipate the lingering effects of pervasive discrimination." This principle is equally applicable to cases of gender discrimination and in discrimination claims modeled on Title VII analysis.

381 Larry P., 793 F2d at 984.
382 Berkman, 705 F2d at 588-89, 594-97.
383 Firefighters Inst. for Racial Equality v St. Louis, 616 F2d at 362-63. See also Ass'n Against Discrimination in Employment, Inc. v Bridgeport, 647 F2d 256 (2d Cir 1981) (defendant city of Bridgeport ordered to compile a list of minority candidates for city fire department to whom offers of employment were required to be made).
384 For example, in Guardians the Second Circuit struck down a long-term goal that minority hiring be comparable to the minority proportion of the relevant labor force. 630 F2d at 113. Further, in Berkman IV the Second Circuit rejected the district court's remedy affording women on a fire department's eligibility list an opportunity to be hired ahead of equally ranked males to compensate for past discrimination. 812 F2d at 61-62.
385 See, for example, EEOC v Local 638, 565 F2d 31, 34 (2d Cir 1977) (upholding an order which replaced the examining board that administered a discriminatory test to assure a more impartial composition, and which imposed an examination bypass mechanism).
actions undertaken by test-users to remedy discrimination. Although Title IX regulations permit federal funding recipients to undertake such affirmative action, there is no case law interpreting the scope of the section. By analogy to Title VII, voluntary plans may be permissible where they meet certain criteria even where a court may not, itself, have authority to issue such relief.

Nonetheless, the United States Supreme Court has made it clear that test-users cannot remedy a non-valid test use which has a discriminatory impact simply by resort to a non-discriminatory "bottom line." *Connecticut v Teal* concerned a promotion eligibility test for the position of Connecticut state welfare eligibility supervisor. The test was invalid and had an adverse effect on Blacks. However, in making promotions from the eligibility lists generated by the exams, the state promoted a greater percentage of eligible Black candidates than white candidates; it argued that this "affirmative action program" should be a complete defense to the suit.

The Court rejected this argument, holding that the discrimination against the minority members who were denied the opportunity to advance based on the invalid test scores could not be justified on the basis that other minority applicants received favorable treatment. The Court emphasized that there must be equal opportunity for each applicant:

Title VII does not permit the victim of a facially discriminatory policy to be told that he has not been wronged because other persons of his or her race or sex were hired. That answer is no more satisfactory when it is given to victims of a policy that is facially neutral but practically discriminatory. Accordingly, *Teal* makes it clear that a user of a discriminatory test must address the test itself and assure that it is valid for the use or uses to which it is being put. A manipulation of the test-generated selection list,
even if voluntarily undertaken and achieving an equitable "bottom line,"
will not defeat a finding of discrimination.

In the context of education-related tests, this principle means that,
for example, a scholarship-granting organization which relies on a non-
valid test use with a discriminatory impact cannot defend against a claim
of discrimination by awarding at the "bottom line" the same number of
scholarships to males and females. Thus, if the SAT is not valid for
determining math ability in young adolescents, a Johns Hopkins or simi-
lar program for academically talented youth could not compensate for
that problem by manipulating admissions rates. However, Teal should
not cause problems for a voluntary remedy following the Kirkland model
in cases where a test has a differential predictive validity by gender.
Since a gender-based score adjustment would actually enhance validity, it
is readily distinguishable from the practice prohibited in Teal.

d. Monetary Damages

It is well established that monetary damages are available under 42
USC § 1983, the statutory cause of action under which equal protection
claims are typically brought. Moreover, in Franklin v Gwinnett
County Public Schools, the Supreme Court recently held that monetary
damages are available for intentional violations of Title IX. Resolving a
split in the Circuits, the Court explained that "absent clear direction
to the contrary by Congress, the federal courts have the power to award
any appropriate relief in a cognizable cause of action brought pursuant to
a federal statute." Because Congress had in no way limited the relief
available pursuant to Title IX's private right of action and, in fact, had
reinforced its view that Title IX should be read broadly through the pas-
sage of the Civil Rights Restoration Act and the Civil Rights Remes-
dies Equalization Act, the Court concluded that damages are available
where the statute is intentionally violated. Although Franklin addressed
compensatory damages, the Court's reasoning extends equally to the
availability of punitive damages.

393 See, for example, Carey v Piphus, 435 US 247, 254-57 (1978).
394 60 USLW 4167 (1992).
395 The Eleventh Circuit had held that damages were not available under Title IX, Franklin v
Gwinnett County Public Schools, 911 F2d 617 (11th Cir 1990), while the Third Circuit had
found that damages were available under Title IX. Pfeiffer v Marion Center Area School Dist.,
917 F2d 779 (3rd Cir 1990).
396 60 USLW at 4171.
399 Although Franklin was a Title IX case, the analysis should apply equally to claims brought
under Title VI of the Civil Rights Act of 1964 and § 504 of the Rehabilitation Act because the
three statutes are interpreted in the same fashion. See notes 151-54 and accompanying text.
e. Attorneys Fees

Finally, attorneys fees are available under the Civil Rights Attorneys Fees Awards Act to plaintiffs prevailing under either a 42 USC § 1983-based equal protection claim or a Title IX claim.\footnote{400 42 USC § 1988 (1991)}

In sum, a broad range of remedies is available where the gender-discriminatory use of standardized tests in education is established. A number of government agencies, including the Department of Education's Office for Civil Rights, have the power to defund recipients which discriminate in violation of relevant statutes. Moreover, some agencies, including OCR, have a wide variety of other enforcement mechanisms at their disposal. Private litigants may seek injunctive relief and, in certain cases, affirmative relief to redress the effects of past discrimination may be available. Attorneys fees are available to prevailing plaintiffs in federal litigation. Finally, damages may be awarded pursuant to 42 USC § 1983 and Title IX.

IV. Conclusions

The examination of the empirical literature on gender bias in standardized testing demonstrates that a wide range of standardized tests reflect gender differentials in scoring. These differentials work to the concrete disadvantage of females in general, and females of color in particular. Many of these differentials are neither fully understood nor explained in the literature. Further, serious validity and fairness concerns are raised in connection with many uses of standardized test scores in education. An analysis of the applicable federal and state constitutional and statutory provisions demonstrates that the current practice which condones the broad use of standardized test scores that reflect gender differentials on only the slightest, if any, evidence of validity and fairness, leaves many education test-users at legal risk.

First, Title IX prohibits a broad range of discriminatory uses of standardized test scores in education. Moreover, it applies to test uses which, although neutral on their face, adversely affect girls and women. Such disparate impact discrimination is at least subject to the same standards which governed employment discrimination under the Title VII analysis of Griggs v Duke Power Co. and continue to govern under the EEOC Uniform Guidelines for Employee Selection. Moreover, there is a well-based argument that discrimination in education should be subject to more stringent standards than those developed in the employment context.

Under Title IX, once a showing of disparate impact is made, the education test-user must show that the test use at issue is justified by
educational necessity. Educational necessity includes a showing that the test use is valid, as well as that the criterion or criteria which have been invoked are important to the selection at issue. In addition, test uses must be fair; that is, their predictive value must not be affected by the gender or race of the test-taker. Even if educational necessity is demonstrated, a challenger to a test use may still prevail by showing that a less discriminatory alternative is available to achieve the same goal.

Second, there is promising support for the proposition that girls and women of color who are victims of discrimination based on both their race and sex may bring claims based on a theory of race and sex discrimination. Combining these claims, either through a joint Title IX and Title VI claim or a statutory or constitutional claim based on a "sex-plus" or "race-plus" theory, has a developing foundation in discrimination law. Such an approach assures that the full and unique nature of the discrimination suffered by women of color is appropriately recognized by the courts.

Third, pursuant to federal constitutional analysis, a test use which is facially or otherwise intentionally discriminatory is subject to a heightened scrutiny analysis. In other circumstances the rational basis test will apply. Rational basis analysis will provide a meaningful standard in many testing discrimination cases since so many test uses are unsupported by validity evidence. As the court concluded in Sharif, where a test which has never been validated for a particular use results in a discriminatory impact, it will be extremely difficult to demonstrate a rational basis for that test use.

Fourth, a number of state constitutional and statutory provisions prohibit gender-discriminatory uses of standardized tests. Some state constitutional provisions afford stronger protections to discrimination victims than appear in federal constitutional analysis.

Fifth, a full range of both government and private remedies are available to victims of testing discrimination. Government remedies include defunding as well as conciliation and enforcement processes. Private remedies extend to injunctive relief, affirmative relief to address the effects of past discrimination in appropriate cases, and attorneys fees. Finally, damages may be awarded pursuant to 42 USC § 1983 and Title IX.

Standardized testing currently plays a major role throughout our system of education. Moreover, there are significant pressures to expand that role. Many important and profound concerns are raised regarding the proper role of testing in education. One major concern is raised by the fact that many standardized test uses discriminate on the basis of gender and the combined bases of gender and race, resulting in very concrete losses to women and girls. Such continued reliance on these tests
should be untenable as a matter of educational policy. It is untenable as a matter of law.