Vanderbilt Law Review

Volume 55 | Issue 4 Article 2

5-2002

Killing the Messenger: The Misuse of Disparate Impact Theory to Challenge High-Stakes Educational Tests

Jennifer C. Braceras

Killing the Messenger: The Misuse of Disparate Impact Theory to Challenge High-Stakes Educational Tests

Jennifer C. Braceras[*]

I. INTRODUCTION 1113
II. EXIT EXAMS AND THE TEST SCORE GAP 1121
A. *Background* 1121
B. *Purposes of High-Stakes Assessments* 1125
C. *Public Opinion* 1127
D. *The Education Reform Landscape and the Test-Score Gap* 1128
1. Texas 1130
2. Massachusetts 1133
E. *Political Opposition to High-Stakes Tests* 1137
III. THEORIES OF DISCRIMINATION 1140
A. *Two Models of Addressing Discrimination: Disparate Treatment Versus Disparate Impact* 1140
B. *Theoretical Justifications for the Disparate Impact Model* 1143
C. *Title VII as the Disparate Impact Prototype* 1147
IV. CLAIMS OF DISCRIMINATION IN EDUCATIONAL TESTING 1150
A. *Title VI's Prohibition on Purposeful Discrimination in Education* 1150
B. *The Use of the Disparate Impact Doctrine to Challenge High-Stakes Educational Tests* 1157

---

* John M. Olin Fellow in Law, Harvard Law School; Commissioner, United States Commission on Civil Rights. J.D., Harvard Law School; B.A., University of Massachusetts-Amherst. This Article was made possible by a grant from the Harvard Law School under the auspices of the Charles Hamilton Houston Fellowship, which the author held during the 2000-2001 academic year. The author is grateful for the advice and/or comments on prior drafts of this Article provided by Patricia L. Bellia, Roberto M. Braceras, Roger Clegg, Robert N. Driscoll, Charles Fried, Mary Ann Glendon, Christine Jolls, Randall L. Kennedy, Gary Lawson, Micbael W. Steinberg, Kate Stith, Abigail Thernstrom, and Stephan Thernstrom. The views expressed in this Article are solely those of the author, and should not be interpreted as the opinions of Harvard University, the U.S. Commission on Civil Rights, or any otber organization with which the author is affiliated.

1. The Resource Guide: Importing Title VII's Disparate Impact Regime to Title VI.....1159
2. Case Law 1160
3. *GI Forum v. Texas Education Agency*.....1162
4. *Alexander v. Sandoval*.....1164
V. ARGUMENTS REGARDING THE APPLICATION OF DISPARATE IMPACT THEORY TO EDUCATIONAL TESTING 1166
A. Rationales for Applying the Disparate Impact Model to High-Stakes Tests 1167
1. Disparate Impact as Objective Evidence of Purposeful Discrimination 1167
2. The Bias Rationale 1170
a. Bias Defined 1171
b. Test Bias as Rationale for Applying the Disparate Impact Model to High-Stakes Educational Tests 1177
c. Reformulating the Disparate Impact Model to Conform to the Bias Rationale 1179
3. The Social Impact/Inequality Rationale 1180
4. Other Rationales 1185
a. Statistical Discrimination.....1186
b. Functional Equivalence.....1188
c. Strong Theories of Disparate Impact.....1189
B. Normative Arguments Against Applying the Disparate Impact Model to Education 1190
1. The Failure of the Educational Necessity Standard 1191
2. Recent Efforts to Refine and Codify the Definition of "Educational Necessity".....1193
3. Adversarial Versus Cooperationist Models.....1197
4. Rewarding Willful Blindness 1200
VI. CONCLUSION.....1201

You don't solve the problem by assaulting the messenger . . . who brings bad news; you don't destroy the thermometer that tells you . . . a . . . fever exists; you don't cure malnutrition by throwing out the scales that identify . . . underweight babies.[1]

I. INTRODUCTION

Educators and policymakers have long been concerned about low levels of American academic achievement.[2] And with good reason. In recent studies by the U.S. Department of Education, twenty-three percent of American twelfth graders were unable to read at even the most "basic" level;[3] more than one-third of all high school seniors lacked basic competency in mathematics;[4] and fifty-seven percent of high school seniors lacked basic knowledge of American history.[5]

---

1. Ronald L. Flaugher, *The Many Definitions of Test Bias*, 33 AM. PSYCHOLOGIST 671, 672 (1978).

2. *See* NAT'L COMM'N ON EXCELLENCE IN EDUC., A NATION AT RISK: THE IMPERATIVE FOR EDUCATIONAL REFORM 5 (1983) [hereinafter A NATION AT RISK] (stating that the "educational foundations of our society are . . . being eroded by a rising tide of mediocrity"); CTR. FOR EDUC. REFORM, A NATION STILL AT RISK: AN EDUCATION MANIFESTO 1 (1998) (noting that the low levels of achievement by American students have remained steady since 1983 when *A Nation at Risk* was released), *available at* http://edreform.com/pubs/mani-fest.htm (last visited April 10, 2001); DIANE RAVITCH, NATIONAL STANDARDS IN AMERICAN EDUCATION: A CITIZEN'S GUIDE, at xxi, 52-53, 59-97 (1995); TWENTIETH CENTURY FUND, MAKING THE GRADE: REPORT OF THE TWENTIETH CENTURY FUND TASK FORCE ON FEDERAL ELEMENTARY AND SECONDARY EDUCATION POLICY (1983); William J. Bennett, *The Case for Education Reform*, *at* http://www.empoweramerica.org/ea/servlet/dispatcher/Articlewebcmd (June 28, 1999) (arguing that education reform is necessary to raise levels of academic achievement).

3. NAT'L CTR. FOR EDUC. STATISTICS, U.S. DEP'T OF EDUC., *The Condition of Education 2000: Reading Performance of Students in Grades 4, 8, and 12* (2000) (summarizing data from the 1998 National Assessment of Educational Progress), *available at* http://nces.ed.gov/pubs2000/coe2000/section2/s_table13_2.html (last visited July 5, 2001); *see also U.S. Teens "Average" on Multination Exam*, CNN.com, *at* http://www.cnn.com/2001/fyi/teachers.ednews/12/04/students.tests.ap/index.html (Dec. 4, 2001) (reporting that U.S. fifteen-year-olds rank fifteenth among the world's thirty-two most developed nations in overall literacy). Diane Ravitch has noted that in high-poverty schools the number of high school seniors unable to read at a basic level tops seventy percent. Diane Ravitch, *Education: See All the Spin*, WASH. POST, Mar. 23, 1999, at A17.

4. James S. Braswell et al., Nat'l Ctr. for Educ. Statistics, The Nation's Report Card: Mathematics 2000 26 (2001), *available at* http://nces.ed.gov/nationsreportcard/pdf/main-2000/2001517a.pdf; Roger Doyle, *Can't Read, Can't Count: Up to One-Third of American High School Seniors Aren't Ready for the Real World*, SCI. AM., Oct. 2001, at 24, 24 (citing data from the 2000 mathematics portion of the National Assessment of Educational Progress). American students also fare poorly in mathematics and science as compared to students in other nations. Thus, the Third International Mathematics and Science Study ("TIMSS") ranked American students nineteenth out of twenty-one nations in math and sixteenth out of twenty-one nations in science. Office of Educ. Research & Improvement, U.S. Dep't of Education, *Highlights from the Third International Math and Science Study* (1999), *available at* http://nces.ed.gov/pubs99/1999081.pdf; RAVITCH, *supra* note 2, at xxii, 84-85 ("[I]nternational

Perhaps even more troubling than the generally poor state of academic achievement among all American high school students is the persistent educational gap between black and Latino[6] students, on the one hand, and white students, on the other.[7] On a wide variety of measures (including test scores, grade point averages, and high school

tests reveal beyond doubt that American students have not learned what their peers in other countries have learned."); *U.S. Teens "Average" on Multination Exam*, *supra* note 3.

5. NAEP, 1994 U.S. HISTORY REPORT CARD: FINDINGS FROM THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (1996), *available at* http://nces.ed.gov/nationsreported/ushistory/findbasic.asp (last visited Apr. 5, 2001); CHESTER E. FINN, JR. & DIANE RAVITCH, EDUCATION REFORM 1995-1996, at 7 (1996) (discussing the NAEP data).

American college students also demonstrate high levels of historical illiteracy. A study of college seniors at America's top colleges in 2000 found that forty percent could not identify the half century in which the American Civil War was fought; two-thirds could not identify George Washington as the American General at Yorktown; only twenty-nine percent could correctly identify the term "Reconstruction" as referring to the readmission of the Confederate states to the union and the protection of the rights of black citizens; and only twenty-three percent could correctly identify James Madison as the "Father of our Constitution" (as compared to ninety-nine percent who could correctly identify "Beavis and Butthead" and ninety-eight percent who could correctly identify "Snoop Doggy Dog"). *See* Center for Survey Research and Analysis at the University of Connecticut, *Elite College History Survey*, *at* http://www.csra.uconn.edu/reports/history.pdf and http://www.csra.uconn.edu/reports/hist_qstnr.htm (last visited May 15, 2001) (reporting the results by a survey conducted for the American Council of Trustees and Alumni); *see also* David S. Broder, *Neglecting History*, WASH. POST, July 2, 2000, at B7 (describing the results of the survey).

6. In this Article, I use the terms "black" and "African-American" interchangeably. I also use the terms "Latino" and "Hispanic" interchangeably to refer to those U.S. residents of Mexican, Cuban, Puerto Rican, Dominican, or Central or South American ancestry. *See generally* Latino Officers Ass'n v. City of New York, 196 F.3d 458, 460 n.1 (2d Cir. 1999) (Cabranes, J.) (explaining the origins of the terms "Hispanic" and "Latino"). Although Hispanics can be of any, or mixed, race, *see* UNITED STATES CENSUS, THE HISPANIC POPULATION IN THE UNITED STATES: POPULATION CHARACTERISTICS (2001), *available at* http://www.census.gov/population/www/socdemo/hispanic, for purposes of this Article they are treated as a separate classification. Accordingly, I use the term "white" in this piece to refer only to those whites who are not of Hispanic origin. I use the term "minorities" here to refer to African-Americans and Latinos collectively. Although Asian-Americans are also considered minorities for most purposes, I do not include them in the term in this Article because, unlike African-Americans and Latinos, Asian-Americans do not, on average, underperform academically vis-à-vis their white peers. For purposes of this Article, then, I use the term "minority" as short-hand for "non-Asian minority" and the term "white" as shorthand for "non-Hispanic white," in order to avoid the awkwardness of those phrases.

7. In recent studies by the U.S. Department of Education's National Assessment of Educational Progress, 43% of black high school seniors and 36% of Latino high school seniors were unahle to read at an even hasic level, compared to 17% of white high school seniors. *See* Doyle, *supra* note 4. Sixty-nine percent of black high school seniors and 56% of Hispanic high school seniors lacked basic proficiency in mathematics, compared to 26% of white high school seniors. *Id.*; *see also* RAVITCH, *supra* note 2, at 63, 72 (noting that the racial achievement gap has narrowed somewhat over the past several decades, but is still substantial); ABIGAIL THERNSTROM & STEPHAN THERNSTROM, GETTING THE ANSWERS RIGHT: RACE, CLASS, AND ACADEMIC ACHIEVEMENT (forthcoming 2003) (discussing the achievement gap); Felicia R. Lee, *Testing Gap: Black, White and Gray Matter*, N.Y. TIMES, Nov. 14, 1999, at B1 (noting the "persistent" achievement gap between non-Asian minority students, on the one hand, and white and Asian students on the other).

and college graduation rates), black and Latino students consistently lag behind white students at the same educational level.[8]

In an attempt to address the problem of low academic achievement in general—and the minority achievement gap in particular—politicians[9] and education reformers[10] have pushed for uniform academic standards and increased accountability. The movement in favor of educational standards and regular academic assessments seeks to hold students, teachers, and administrators accountable by rewarding academic achievement and by exposing academic failure to public criticism as well as to corrective and remedial action. Reformers argue that only by testing all students, and by attaching consequences to the results of such tests, can we encourage teachers to focus on a uniform core curriculum, motivate student learning, and raise levels of student performance.[11] Proponents of this new accountability also contend that any serious attempt to close the minority achievement gap must necessarily include a testing regime to provide the public and policymakers with comparative data and to provide incentives to raising achievement levels in even the most disadvantaged communities.[12]

---

8. Jodi Wilgoren, *Report Calls for New Focus on Aid for Minority Students*, N.Y. TIMES, Oct. 17, 1999, at A30 (reporting that, for decades, African-American and Latino students have trailed white students and Asian-American students academically and that the gap in achievement is evident in grades, test scores, graduation rates, and levels of academic attainment); *see also* STEPHAN THERNSTROM & ABIGAIL THERNSTROM, AMERICA IN BLACK AND WHITE: ONE NATION INDIVISIBLE 360 (1997) (discussing the black-white academic gap).

9. During the 2000 presidential campaign, both Al Gore and George W. Bush pressed the need for accountability through testing, with Bush proposing to make improvement on statewide tests a condition of federal funding. *See* Andrew Goldstein, *Who's the Education President?*, TIME, Nov. 6, 2000, at 82, 82; Patrick Healy, *Candidates Take on Education Goals May Be the Same But the Methods Are Different*, BOSTON GLOBE, Nov. 4, 2000, at A12.

10. *See generally* RAVITCH, *supra* note 2; *High Standards: Giving All Students a Fair Shot*, ACHIEVE POLICY BRIEF (Achieve, Inc., Washington, D.C.), Fall 2000, at 2 (arguing that uniform educational standards and tests with consequences will improve the academic achievement of all students and help to level the playing field between educationally rich and educationally poor communities), *at* http://www.achive.org/dstore.nsf/lookup/Higb%20standards%20Policy%20-Brief%/$file/High%20Standards%20Policy%20Brief.pdf; Ctr. for Educ. Reform, *at* http://www.ed-reform.com (last visited Mar. 1, 2001); Thomas B. Fordham Foundation, *Standards, Testing & Accountability*, *at* http://www.edexcellence.net/topics/standards.html (last visited Mar. 1, 2001).

11. Achieve, Inc., *supra* note 10, at 2 ("[T]ying consequences to results creates incentives for schools to raise performance and encourages students to work hard."); Tbomas B. Fordham Foundation, *supra* note 10.

12. *Testing: Setting the Record Straight*, ACHIEVE POLICY BRIEF (Achieve, Inc., Washington, D.C.), Summer 2000, at 5 ("Tests based on common standards . . . do a *better* job than the current system in fostering equity by providing common expectations for all students. . . . Without consequences tied to test results, there would be little incentive for schools to deal with current gaps in achievement."), *at* bttp://www.achive.org/dstore.nsf/Lookup/Testing-Setting%20tbe-%20Record%20Straight/$file/Testing-Setting%20the%20Record%20Straight.pdf; William L. Taylor, *Standards, Tests, and Civil Rights*, EDUC. WK., Nov. 15, 2000, at 56 ("Standards and

Generally speaking, the new accountability takes the form of academic assessments that a student must pass in order to advance to the next grade level or receive a high school diploma. Typically, students who fail these assessments are provided remedial help and additional chances to pass the exam.[13] Despite the good intentions behind such reform efforts, in the short term a disproportionately large number of African-American and Hispanic students have failed such tests, both on the first try and after multiple attempts.[14]

Because of the high failure rates of minority students on standardized academic assessments, some testing critics argue that requiring students to pass a mandatory exam in order to move to the next educational level will entrench current inequalities.[15] Others go further, arguing that the tests are themselves "biased" against African-American and Latino students and that the tests, therefore, fail to measure accurately what students from these communities actually know.[16] For one or both of these reasons, many groups that

---

accountability expose the sham that passes for education in many heavily minority schools and provide measurements and pressure to prod schools to target resources where they are needed most.").

13. *See infra* notes 76-78 and 104-07 and accompanying text.

14. *See, e.g.*, Abigail Thernstrom, *Testing and Its Enemies: At the Schoolhouse Barricades*, NAT'L REV., Sept. 11, 2000, at 38, 40 (noting that in one administration of the Massachusetts Comprehensive Assessment System "[eighty] percent of non-Asian minority students failed the tenth-grade math assessment [and that] [f]ifty-seven percent flunked English"); *see also* GI Forum v. Tex. Educ. Agency, 87 F. Supp. 2d 667, 679 (W.D. Tex. 2000) (explaining that on the Texas Assessment of Academic Skills, the gap between the passing rate of Hispanic and black students, on the one hand, and white students, on the other, shrinks when cumulative pass rates are examined but that, nevertheless, the gap on cumulative administrations of the exam still reveal "significant statistical differences" sufficient to satisfy a legal showing of adverse impact).

15. *See, e.g.*, NAT'L RESEARCH COUNCIL, HIGH STAKES: TESTING FOR TRACKING, PROMOTION, AND GRADUATION 4 (Jay P. Heubert & Robert M. Hauser eds., 1999) (arguing that the use of high-stakes tests to make determinations regarding promotion or graduation serves only to reinforce current educational inequalities); Linda M. McNeil, *Creating New Inequalities: Contradictions of Reform*, 81 PHI DELTA KAPPAN 728, 732 (2000) (arguing that standardized educational tests mask current inequalities and do nothing to level the playing field between educationally rich and educationally poor communities); *see also* Richard Whitmire, *Controversy Sharpens over High-stakes Testing*, GANNETT NEWS SERVICE, May 22, 2000, (describing arguments that high-stakes tests create a racist roadblock to social advancement).

16. *See, e.g.*, David M. White, *Culturally Biased Testing and Predictive Validity: Putting Them on the Record*, 14 HARV. C.R.-C.L. L. REV. 89, 90 (1979) (arguing that the minority test score gap is caused by "cultural bias" in the tests which masks minority test-takers' true knowledge and abilities); Hagit Elul, Note, *Making the Grade, Public Education Reform: The Use of Standardized Testing to Retain Students and Deny Diplomas*, 30 COLUM. HUM. RTS. L. REV. 495, 517 (1999) (arguing that racial and ethnic disparities in test results are due to test measurement errors, not actual achievement gaps); Lawrence Feinberg, *Florida Tests Literacy: 37% of Juniors Fail; One Junior in Three Fails Fla. High School Literacy Test*, WASH. POST, Apr. 2, 1978, at A1 (describing charges by the Southern Christian Leadership Conference and the NAACP that Florida's statewide graduation exam is "culturally biased" against minorities); David Gonzalez, *Testing Worth, and Patience, of Teachers*, N.Y. TIMES, Aug. 29, 1998, at B1

purport to represent the interests of minorities and organizations that oppose standardized testing argue that tests which produce disproportionate racial or ethnic results are discriminatory.[17]

As a logical outgrowth of this view, a number of commentators and special interest groups have advocated using disparate impact theory—which has its roots in the law of employment discrimination[18]—to mount legal challenges to a variety of facially neutral testing regimes with disproportionate demographic outcomes.[19] The legal hook for such challenges is Title VI of the Civil

---

(reporting claims that New York's teacher exam is "culturally biased" against black and Hispanic teachers); Peggy Walsh-Sarnecki, *Race Plays Big Role in MEAP Scores; Expert Finds that White, Black Students Give Different Answers,* DETROIT FREE PRESS, May 18, 2001, at 1A (describing claims that Michigan's statewide academic assessment is biased against minorities); J. Steven Warner, Letter to the Editor, PALM BEACH POST, Feb. 9, 2001, at 17A (arguing that a passage on the reading comprehension portion of the Florida statewide assessment was "culturally biased" because it described a two-parent nuclear family).

17. *See infra* notes 115-28 and accompanying text (describing claims by representatives of the NAACP and various Latino advocacy groups that high-stakes tests are discriminatory); *see also* Georgia N. Alexakis, *Test Prep: What Bush Can Learn from a Tryout of School Reform in Massachusetts,* WASH. MONTHLY, Mar. 1, 2001, at 29, 34-36 (describing claims that mandatory testing is discriminatory); Fair Test Examiner, *Parents File OCR Complaint* (Winter 1999-2000) (applauding a claim filed with the U.S. Department of Education by Chicago parents alleging that Chicago's policy of requiring students to pass a test before being promoted discriminates against black and Latino students), *at* http://www.fairtest.org/examarts/winter00/Parents_File_OCR_Complaint.html (last visited Aug. 3, 2001).

18. For a thorough discussion of the principles underlying "disparate impact" law in the employment context, see Pamela L. Perry, *Two Faces of Disparate Impact Discrimination,* 59 FORDHAM L. REV. 523 (1991), and Steven L. Willborn, *The Disparate Impact Model of Discrimination: Theory and Limits,* 34 AM. U. L. REV. 799 (1985).

19. *See generally* Arthur L. Coleman, *Excellence and Equity in Education: High Standards for High-Stakes Tests,* 6 VA. J. SOC. POL'Y & L. 81 (1998) (arguing that the availability of the disparate impact model for addressing claims of discrimination in testing promotes the creation and implementation of fair and equitable tests); Preston C. Green III, *Can Title VI Prevent Law Schools from Adopting Admissions Practices that Discriminate Against African-Americans?,* 24 S.U. L. REV. 237 (1997) (promoting the use of disparate impact lawsuits to enjoin educational institutions from relying on test scores with disproportionate racial results); William C. Kidder, *The Rise of the Testocracy: An Essay on the LSAT, Conventional Wisdom, and the Dismantling of Diversity,* 9 TEX. J. WOMEN & L. 167 (2000) (favoring the disparate impact model as a way to root out so-called biased tests); James S. Wrona, *Eradicating Sex Discrimination in Education: Extending Disparate Impact Analysis to Title IX Litigation,* 21 PEPP. L. REV. 1 (1994) (advocating disparate impact lawsuits against tests with different average outcomes for men and women); Elul, *supra* note 16, at 506 (arguing in favor of disparate impact litigation to challenge high-stakes educational assessments); Daniel J. Losen, Note, *Silent Segregation in Our Nation's Schools,* 34 HARV. C.R.-C.L. L. REV. 517 (1999) (arguing in favor of disparate impact litigation to eliminate tests used to "track" students by ability); Andrea L. Silverstein, Note, *Standardized Tests: The Continuation of Gender Bias in Higher Education,* 29 HOFSTRA L. REV. 669 (2000) (arguing that disparate impact litigation should be mounted against the use of tests on which the average male test-taker outperforms the average female test-taker); *see also* GI Forum v. Tex. Educ. Agency, 87 F. Supp. 2d 667 (W.D. Tex. 2000) (concerning unsuccessful lawsuit brought by the Mexican American Legal Defense and Educational Fund against the Texas Education Agency, arguing that the Texas Assessment of Academic Skills has an unlawful disparate impact

Rights Act of 1964,[20] which prohibits recipients of federal funding from discriminating on the basis of race or ethnicity.[21] Although Title VI does not by its terms provide for disparate impact litigation, regulations promulgated to implement the statute explicitly forbid federally funded entities from using "criteria or methods of administration which have the *effect* of subjecting individuals to discrimination."[22] In other words, Title VI's implementing regulations prohibit institutions that receive federal assistance—including most American schools and colleges—from implementing policies that have a disproportionate adverse impact on minorities.

In practice, this means that the federal government may withdraw funding from any educational institution that employs tests with disproportionate demographic outcomes. Until recently, it also meant that private parties could bring federal lawsuits against schools that employed such tests.[23]

---

on black and Hispanic students); Ronald Brownstein, *Call for Academic Standards Could Face Test from Civil Rights Law*, L.A. TIMES, Aug. 11, 1997, at A5 (describing a 1994 investigation by the Clinton Administration's Department of Education into the State of Ohio's graduation exam); Kenneth J. Cooper, *Standardized Exam Faces Test in Texas: Trial Gauges Discriminatory Effect*, WASH. POST, Sept. 22, 1999, at A19 (describing the *GI Forum* case in Texas, and citing complaints filed with the U.S. Department of Education against similar exams in North Carolina and Nevada).

20. 42 U.S.C. § 2000d (1994).

21. Title VI provides that "[n]o person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance." *Id.*

22. 34 C.F.R. § 100.3(b)(2) (2001) (emphasis added).

23. Disparate impact lawsuits against educational tests have met with mixed results. *Compare* Larry P. v. Riles, 793 F.2d 969, 980, 983 (9th Cir. 1984) (holding unlawful under Title VI the use of so-called IQ tests to identify students to be placed in "dead end" classes for the educable, mentally retarded on the ground that the use of the test resulted in a disproportionately high number of minority students being placed in these classes), Cureton v. NCAA, 37 F. Supp. 2d 687 (E.D. Pa.) (holding unlawful an NCAA rule requiring that incoming student athletes have achieved a minimum score on the SAT [820] or ACT [68] in order to compete in intercollegiate athletics on the ground that it results in an unjustified disparate impact on African-Americans), *rev'd on other grounds*, 198 F.3d 107 (3d Cir. 1999), Groves v. Ala. State Bd. of Educ., 776 F. Supp. 1518, 1526-33 (M.D. Ala. 1991) (holding unlawful Alabama's practice of requiring a minimum score on the American College Testing Program's ACT examination for admission to undergraduate teacher training programs where reliance on a particular cutoff score had a disparate impact on black students), *and* Sharif v. N.Y. State Educ. Dep't, 709 F. Supp. 345 (S.D.N.Y. 1989) (holding that New York's reliance upon SAT scores to award scholarships violated Title IX's prohibition against sex discrimination because of the test's disproportionate impact on female students), *with* Debra P. v. Turlington, 730 F.2d 1405 (11th Cir. 1984) (allowing the State of Florida beginning in 1983 to deny diplomas to students who fail to pass a state exam, irrespective of the disparate impact on black students); *GI Forum*, 87 F. Supp. 2d at 667 (upholding the State of Texas's high school graduation exam from a disparate impact challenge).

In April 2001, the U.S. Supreme Court dealt a blow to private disparate impact lawsuits against educational tests. In *Alexander v. Sandoval*—a case that on its face had nothing to do with education policy—the Court rejected a disparate impact challenge to a state driver's license exam, offered only in English, holding that Title VI does not provide a private right of action for claims of disparate impact.[24]

Nevertheless, the disparate impact regulations promulgated under Title VI remain in force, thus allowing the federal government to withdraw funding from institutions (like school districts) that utilize tests or other policies that result in disproportionate racial or ethnic outcomes. Moreover, supporters of the disparate impact approach argue that private lawsuits based on the disparate impact theory may still proceed under 42 U.S.C. § 1983. In addition, supporters of the disparate impact approach have called upon Congress to amend Title VI to permit explicitly private lawsuits against federally funded institutions that maintain policies or practices that affect some racial or ethnic groups differently than others (including schools that use standardized graduation exams).[25]

The fundamental normative question raised by these developments is whether the disparate impact model of discrimination is appropriately applied in the educational testing context. This Article answers that question in the negative, arguing that doctrines born of employment law should not be transported reflexively to other antidiscrimination statutes.[26] In particular, I argue that it is improper to use the disparate impact model to determine whether primary and secondary educational assessments are discriminatory. Because standardized tests used by primary and secondary schools can just as easily be understood as a tool for *remedying* educational inequality as

---

24. 532 U.S. 275 (2001). Prior to *Sandoval*, it had been widely assumed that private parties could bring disparate impact lawsuits for injunctive relief pursuant to Title VI's regulations. *See* Guardians Ass'n v. Civil Serv. Comm'n, 463 U.S. 582 (1983) (White, J.) (Marshall, J., dissenting) (Brennan, Blackmun & Stevens, JJ., dissenting).

25. *See, e.g.*, David Dante Troutt, *Behind the Court's Civil Rights Ruling*, N.Y. TIMES, Apr. 29, 2001, at A4; Press Release, ACLU, Two Supreme Court Rulings Expand Police Powers and Limit Civil Rights Enforcement (Apr. 24, 2001) (quoting Steven Shapiro of the ACLU as stating that "Congress must act once again to ensure that victims of discrimination can have their day in court"), *available at* http://www.aclu.org/news/2001/n042401c.html; Telephone Interview with Chris Anders, ACLU (May 3, 2001) ("I anticipate that there will eventually be legislative efforts to amend [Title VI in light of *Sandoval*]. . . . [T]he ACLU would certainly support such efforts.").

26. *Cf.* Latimore v. Citibank Fed. Sav. Bank, 151 F.3d 712, 712-14 (7th Cir. 1998) (Posner, J.) (rejecting the application of Title VII's burden-shifting framework to claims of lending discrimination, and noting that the "wholesale transportation" of discrimination theories and standards of proof from one statutory context to another "display[s] insensitivity to the thinking behind the standard").

a tool for *exacerbating* inequality, educational policymakers should be permitted to use their discretion in deciding whether and how to employ these tests absent intentional discrimination. In other words, because testing regimes necessarily impose both costs *and* benefits on minority test-takers and communities, I argue that they are inappropriate targets of the disparate impact doctrine.

To be sure, concerns regarding a test's disproportionate burden on certain communities may be socially and politically relevant. In determining whether or how to implement a testing regime, politicians and policymakers may reasonably consider the consequences that such a regime will have on various constituencies and whether a policy that results, at least initially, in a large number of minority students failing to receive diplomas is politically or socially advisable. Ultimately, however, the issue should remain a policy determination, and schools that utilize academic assessments with consequences should not be penalized for doing so in the absence of evidence that the tests were imposed for a discriminatory purpose.

In short, the federal government should not pursue policies to withdraw funding from schools that employ high-stakes educational assessments absent a showing of intentional discrimination, and the law should not be interpreted or amended to permit private, disparate impact lawsuits challenging educational assessment policies.[27]

The remainder of this Article proceeds as follows: Part II discusses the movement in favor of statewide educational assessments and, in particular, describes the tests employed by Texas and Massachusetts to determine whether students receive a high school diploma. Part II also examines the test score gap between African-American and Latino students, on the one hand, and white students, on the other, and describes how this test score gap has fueled charges that such tests "discriminate" against minority students. Part III of

27. Thomas A. Lambert, who has written against the use of private litigation to enforce Title VI's disparate impact regulations, has argued that disparate impact claims may properly be adjudicated by the federal agencies charged with enforcing Title VI. *See* Thomas A. Lambert, *The Case Against Private Disparate Impact Suits*, 34 GA. L. REV. 1155, 1186 (2000) (arguing "in favor of an enforcement approach that permits 'agency nullification' of the disparate impact regulations in order to avoid overdeterrence of disparity-causing decisions"). These agencies, Lambert argues, are well suited to weighing the costs and benefits of a given policy and will be able to screen out claims of disparate impact that also have an overall beneficial outcome. *See id.* (arguing that if enforcement of the disparate impact regulations is limited to agency action, then "administrative agencies could perform a screening function, picking out the bad disparity-causing actions for prosecution but allowing the good ones . . . to slide by"). I do not share Lambert's faith in Washington bureaucrats. To the contrary, I argue here that, at least as far as the education context is concerned, the costs and benefits of educational reforms should be balanced at the local level by local educators, policymakers, and parents through the regular democratic process.

this Article discusses various theories of discrimination and outlines the standards of proof that apply in a disparate impact case brought by an employee against an employer under Title VII of the Civil Rights Act of 1964. Part IV considers claims of discrimination in standardized educational testing—in particular, Title VI's prohibition on *intentional* discrimination by educational institutions and the legal framework for challenging the educational tests on grounds of *disparate impact*. Part V reviews the arguments for and against the use of disparate impact theory in the educational testing context. Part VI concludes that, absent proof of an intent to discriminate, educational institutions should not be punished for implementing strict accountability programs and should not have to bear the risk of losing federal funding or have to battle costly lawsuits simply because they administer high-stakes tests that reveal a minority performance gap.

II. EXIT EXAMS AND THE TEST SCORE GAP

In this section, I briefly describe the accountability and testing movement in the United States and outline one of the major critiques of standardized achievement tests: the argument that standardized tests are racially and ethnically discriminatory. Part II.A sketches the history of the American testing movement, while Part II.B explains the argument in favor of requiring students to pass an exam as a condition of promotion or graduation. Part II.C describes the current testing landscape and discusses the gap between the scores of the average black or Hispanic student and the average white student, using the Texas and Massachusetts systems as illustrations of the national phenomenon. Part II.D discusses the level of public support for high-stakes graduation tests. Finally, Part II.E describes how the test score gap has fueled charges that state accountability programs—and high school graduation exams in particular—are "culturally biased" and "discriminatory."

A. Background

For more than a century,[28] standardized tests have been used as an objective measurement of academic performance. The tests that

28. The State of New York first offered a high school Regents Exam in 1879 as a standard for colleges to use in evaluating high school graduates. Raymond Hernandez, *Pataki Defends Tougher Graduation Tests*, N.Y. TIMES, May 13, 1999, at B5. Students bound for college took college prepatory exams in various subjects, and the state awarded Regents Diplomas to those who passed eight or more such tests. Somni Sengupta, *Tough New Regents Exams Are Unveiled*

are the focus of this Article—those tests administered in primary and secondary schools for the purpose of making promotion and graduation decisions—are commonly referred to as "achievement tests" or "academic assessments." Unlike so-called "aptitude" tests (such as most employment tests) that seek to predict an individual's future performance or chance of success in a particular setting,[29] achievement tests are intended to measure what test-takers have learned after completing a particular course of study.[30]

In the 1970s, education reformers popularized the idea of using standardized achievement tests to make promotion and graduation decisions.[31] Then, as now, the idea behind requiring all students to

---

in New York, N.Y. TIMES, June 25, 1998, at B5. In the 1970s, New York began allowing students to take basic competency exams in order to earn local high school diplomas. James Dao, *Passing of Regents' Exams to Be Required for Diploma*, N.Y. TIMES, Apr. 24, 1996, at A1. Currently, all students must take the Regents Exam in order to earn a diploma. *Id.* Students who do not pass the Regents Exam can earn an equivalency diploma hy passing a less stringent test of basic skills. *Id.*

29. The SAT is an example of an academic aptitude test. *See generally* NICHOLAS LEMANN, THE BIG TEST: THE SECRET HISTORY OF THE AMERICAN MERITOCRACY (1999) (describing the history and development of the SAT).

30. *See* HOWARD B. LYMAN, TEST SCORES AND WHAT THEY MEAN, 22-23, 150-51 (1998). Standardized tests can be scored in two basic ways: by reference to norms or by reference to some outside criterion. CTR. FOR EDUC. REFORM, ACTION PAPER: THE NEW GENERATION OF STANDARDIZED TESTING (2000), *at* http://edreform.com/pubs/testing.htm. Norm-referenced tests measure the *relative* performance of a particular school or cbild. *Id.* These tests are the most common standardized tests and include tbe Iowa Test of Basic Skills ("ITBS"), the SAT, the California Test of Basic Skills ("CTBS"), the California Achievement Test ("CAT"), and the Terra Nova. *Id.* Because norm-referenced tests are designed for the purpose of making relative comparisons, they do not directly provide information about a test-taker's actual mastery of the material tested. *Id.* For example, a score in the 90th percentile on a norm-referenced test tells us only that the student performed better than ninety percent of those tested, not tbat the child has actually mastered the material tested. *Id.*

Criterion-referenced tests, by contrast, are those tests which are linked to particular standards. *Id.* These tests are designed to measure wbether (and how well) the test-taker has mastered a particular skill or area of knowledge. *Id.* Examples of criterion-referenced exams include state bar examinations, medical licensing exams, airline pilot tests, and high school exit exams. *Id.* The publishers of these types of tests must identify successful and unsuccessful levels of performance and then establish a cutoff score that defines a passing grade. *Id.*

31. NAT'L GOVERNORS ASS'N, HIGH SCHOOL EXIT EXAMS: SETTING HIGH EXPECTATIONS 1 (1998); RAVITCH, *supra* note 2, at 47-50; Rachel F. Moran, *Sorting and Reforming: High-stakes Testing in the Public Schools*, 34 AKRON L. REV. 107, 111 (2000).

There has been a debate in recent years among education reformers as to whether national or statewide standards and tests are best. In 1991, the U.S. Department of Education developed voluntary national reading and math tests to be administered in the fourth, eighth, and twelfth grades. *See, e.g.*, Kenneth J. Cooper, *Exams Opposed over Potential Harm to Minorities*, WASH. POST, June 12, 1991, at A21. Some political conservatives objected to tbe proposal on the ground that *national* tests would erode local control over schools. RAVITCH, *supra* note 2, at xvi-xvii (describing conservative claims that national standards and tests represented "a dangerous step toward federal control of education"); Cooper, *supra*. Political liberals, meanwhile, also opposed the plan based on suspicions that such tests would reveal disparate racial and ethnic results. *See* Cooper, *supra*; *see also* RAVITCH, *supra* note 2, at 19 (describing liberals' fears that meaningful

pass a uniform exam was to: (1) ensure that all students receive a firm foundation in the basics (particularly reading, writing, and mathematics); (2) raise student achievement levels; and (3) permit the public to hold students, teachers, and schools accountable for their performance.[32]

The movement in favor of exams with consequences—or "high-stakes tests" as they have come to be called[33]—enjoyed modest success in the early 1970s.[34] But the testing movement soon encountered resistance from those who argued such testing would unfairly penalize students who were educated under *de jure* segregation or who were otherwise victims of intentional discrimination.[35] Such concerns

---

standards would cause poor and minority children to fail or drop out of school). In 1998, a coalition of liberal and conservative groups convinced the House of Representatives to pass a bill prohibiting the U.S. Department of Education from using appropriations to develop national standards tests. *For the Record*, WASH. POST, Feh. 12, 1998, at M10 (providing roll call of vote on bill passed in House of Representatives concerning national testing), 1998 WL 2467206.

Currently, the only federal achievement test is the National Assessment of Educational Progress ("NAEP"), which is used exclusively to generate research data and is not tied to any consequences. Moran, *supra* note 31, at 112. Throughout the 1990s, the campaign for standardized testing with consequences focused on the states. *Id.* at 113; *see also* Thomas B. Fordham Foundation, *supra* note 10 (noting that, although standards can be set at the local, state, or national level, in the late 1990s and the early part of the twety-first century, the standards movement has been spearheaded at the state level).

On January 8, 2002, President George W. Bush signed into law the "No Child Left Behind Act of 2001" which requires all states to set high standards of achievement and create a system of accountability to measure results. The law requires that, at a minimum, states test every child in grades three through eight in reading and math. No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

32. *See* CTR. FOR EDUC. REFORM, *supra* note 30; Achieve, Inc., *supra* note 12, at 5; Thomas B. Fordham Foundation, *supra* note 10.

33. As defined by Professor Perry A. Zirkel, "high-stakes tests" are standardized examinations that have "dramatic and direct [educational] consequences" for the test-taker. Perry A. Zirkel, *Tabular Analysis of the Case Law Concerning High Stakes Testing*, 143 EDUC. REP. 697 (2000); *see also* GREGORY J. CIZEK, FILLING IN THE BLANKS: PUTTING STANDARDIZED TESTS TO THE TEST 10 (1998) (describing the difference between high-stakes tests and low-stakes tests), *available at* http://www.edexcellence.net/library/cizek.pdf; NAT'L RESEARCH COUNCIL, *supra* note 15, at 51 n.1 (defining "high-stakes tests" as those which can lead to adverse consequences for the individuals taking the exams).

34. *See* Jay P. Heubert, *High-Stakes Testing: Opportunities and Risks for Students of Color, English-Language Learners, and Students with Disabilities*, *in* THE CONTINUING CHALLENGE: MOVING THE YOUTH AGENDA FORWARD (M. Pines ed.) (forthcoming, n.d.) (noting that state minimum competency tests gained popularity during the 1970s), *available at* http://www.cast.org/ncac/index.cfm?i=920 (last visited Apr. 5, 2001). Florida, for example, instituted a graduation exam requirement during the 1977-1978 school year. *See* NAT'L GOVERNORS ASS'N, *supra* note 31, at 1.

35. *See, e.g.*, Anderson v. Banks, 520 F. Supp. 472, 500 (S.D. Ga. 1981) (holding that a school diploma policy must be evaluated in light of the past *de jure* segregation in the school district and prohibiting the imposition of the diploma sanction on those who spent their primary years in the dual system), *rev'd in part on reh'g*, 540 F. Supp. 761 (S.D. Ga. 1982), *appeal dismissed sub nom.* Johnson v. Sikes, 730 F.2d 644 (11th Cir. 1984); Debra P. v. Turlington, 474 F. Supp. 244, 269 (M.D. Fla. 1979) (enjoining Florida's testing regime for four years in order to

stalled initial efforts to establish statewide standardized exams with high-stakes consequences.[36]

During the 1980s, however, many state policymakers became concerned that large increases in educational spending had rendered only modest returns.[37] Such concerns were fueled in part by a series of reports published in the early 1980s detailing the decline in American education in general and the gap between white and minority students in particular. In the most widely heralded report, *A Nation at Risk*, the President's National Commission on Excellence in Education warned that the "educational foundations of our society [are] . . . being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people."[38] In support of this conclusion, the Commission pointed to the increase in the number of functional illiterates, particularly in minority communities;[39] a decline in SAT and other achievement test scores among all racial groups; and a significant increase in the demand for remedial education at the college level.[40] In response to these problems, the Commission recommended, among other things, implementation of performance standards and assessments for students and teachers at all levels of primary and secondary school.[41]

"purge the taint of past segregation"), *aff'd in part, vacated and remanded in part*, 644 F.2d 397 (5th Cir. 1981); *see generally* Moran, *supra* note 31, at 111 (explaining that in the 1970s "[t]he United States had only recently embarked on efforts to desegregate the schools and to undo past inequities in education based on race" and that "[a]s a result, denying a diploma to students who could not pass an exit examination seemed to punish the victims of past discrimination for having attended inferior schools").

36. Moran, *supra* note 31, at 111; *see also* NAT'L GOVERNORS ASS'N, *supra* note 31, at 1 (noting that throughout the 1970s, requirements for graduation continued to be controlled by local school districts).

37. The failure of increased spending to improve academic outcomes has been documented by noted education scholar Chester Finn. *See* CHESTER E. FINN, JR., WE MUST TAKE CHARGE: OUR SCHOOLS AND OUR FUTURE 2 (1991) (noting that in 1989 America spent 29 percent more real dollars per pupil in the public schools than it did in 1980, yet academic test scores remained level and graduation rates rose only slightly); *see also* THERNSTROM & THERNSTROM, *supra* note 7 (arguing that the problem of poor academic achievement cannot be explained as a problem of underfunding).

38. A NATION AT RISK, *supra* note 2, at 1.

39. The Commission found that approximately thirteen percent of all seventeen-year-olds were functionally illiterate by the simplest tests of everyday reading, writing, and comprehension and that functional illiteracy among minority youth was as high as forty percent. *Id.* at 8.

40. *Id.* at 8-9.

41.

> Standardized tests of achievement (not to be confused with aptitude tests) should be administered at major transition points from one level of schooling to another and particularly from high school to college or work. The purposes of these tests would be to: (a) certify the student's credentials; (b) identify the need for remedial intervention; and (c) identify the opportunity for advanced or accelerated work. The tests should be

The findings and recommendations of *A Nation at Risk* received widespread publicity and generated a national discussion about education reform. According to Diane Ravitch, the noted historian of American education, *A Nation at Risk* gave birth to hundreds of state-level task forces charged with recommending new ways to raise academic standards, and breathed new life into efforts to enact educational accountability regimes.[42]

B. Purposes of High-Stakes Assessments

Proponents of accountability through high-stakes tests argue that meaningful education reform can only be achieved through a tripartite strategy of standards, assessments, and consequences.[43] According to this school of thought, academic standards should be established to define what basic material students should know at various points in their educational careers.[44] Students should then be tested on the material outlined in the standards so that educators, policymakers, and the public at large may assess whether students and schools are meeting the standards. Finally a system of rewards and punishments—the key aspect of any accountability program—must be established in order to provide the proper incentives for schools and students to commit to teaching and learning the material.[45]

Jay Heubert and Robert Hauser have summarized the three components of standards-based reform:

> [S]tandards-based reform[s] . . . are premised on the idea of setting clear, high standards for what children are supposed to learn and then holding students—and often educators and schools—to those standards. The logic seems clear: Unless we test students' knowledge, how will we know if they have met the standards? And the idea of accountability, which is also central to this theory of school reform, requires that the

---

administered as part of a nationwide (but not Federal) system of State and local standardized tests. This system should include other diagnostic procedures that assist teachers and students to evaluate student progress.

Id. at 28

42. RAVITCH, *supra* note 2, at 53 (citing *A Nation Responds*); *see also* William J. Johnson, EDUCATION ON TRIAL: STRATEGIES FOR THE FUTURE 4 (William J. Johnson ed., 1985) (noting that one year after publication of *A Nation at Risk* no less than 275 task forces had been formed to study the problem of low academic achievement and to present specific proposals for local reform).

43. Thomas B. Fordham Foundation, *supra* note 10.

44. *Id.*

45. *Id.*

> test results have direct and immediate consequences: a student who does not meet the standard should not be promoted, or awarded a high school diploma.[46]

Reformers often argue that requiring all students to master the fundamentals in core subjects will help to equalize educational outcomes and ensure that all of a state's graduates are competent in basic subjects.[47] Only by setting uniform standards, reformers argue, can we ensure that students from diverse racial and socioeconomic backgrounds receive a common education and are afforded equal educational opportunities.[48]

Accountability advocates contend that the second component of reform—standardized tests—provides an important tool for policymakers and the public to evaluate and compare the performance of various students, schools, and districts over time.[49] Standardized tests provide information to teachers and students regarding the state of academic achievement, draw attention to struggling schools and students in need of help, and identify educational programs and pedagogical methods that work.[50] Such tests also help to determine how scarce resources should be spent and can spur targeted spending increases.[51]

Finally, reformers argue that, in order to be effective, academic assessments must be accompanied by both a carrot and a stick. By attaching consequences to test scores—rewards for improvement, remedial assistance for struggling schools and students, and penalties for repeated failure—high-stakes exams motivate teaching and learning. The theory behind this approach is that if students and/or teachers are penalized for repeated failure, school districts and teachers will have a greater incentive to teach the material outlined in

---

46. NAT'L RESEARCH COUNCIL, *supra* note 15, at 13. Although they acknowledge the "logic" of the pro-testing argument, Heubert and Hauser nevertheless oppose the use of standardized tests as the sole basis for making bigh-stakes decisions. *Id.*

47. Achieve, Inc., *supra* note 12, at 5.

48. *See, e.g.*, RAVITCH, *supra* note 2, at xxiii, 26 ("Standards establish the principle that all students should encounter the same educational opportunities and the same performance expectations, regardless of who their parents are or what neighborhood they live in."); Achieve, Inc., *supra* note 12, at 5.

49. CTR. FOR EDUC. REFORM, *supra* note 30 ("Over time, trends in test scores reveal how much progress schools have made in their efforts to maintain high scores or raise inadequate scores."); Elliot W. Eisner, *The Uses and Limits of Performance Assessment*, PHI DELTA KAPPA ONLINE, *at* http://www.pdkintl.org/kappan/keis9905 (last visited Jan. 10, 2001) ("One of the motivations behind the standards movement is the desire to hold schools accountable, and that accountability is facilitated if schools, classrooms, and students can be compared.").

50. CTR. FOR EDUC. REFORM, *supra* note 30.

51. *Id.*; Alexakis, *supra* note 17, at 29 (noting that accountability measures, including high-stakes testing, can increase political support for targeted spending increases, much in the same way that adding a work requirement to welfare payments increased political support for those benefits).

state standards and students will have a stronger incentive to learn it.[52] By shining a spotlight on schools that lag behind, threatening state receivership or potential closure for those that fail to improve, and rewarding schools that demonstrate results, states can ensure that schools no longer ignore or make excuses for deficient academic performance, particularly in poor and minority communities.[53]

C. Public Opinion

High educational standards and high-stakes exams enjoy consistent and widespread public support.[54] Indeed, for several decades, Gallup polls have indicated public support of over seventy percent for standardized tests as determinants of promotion or graduation.[55] Gallup's findings are not an aberration. A 1999 survey sponsored by National Public Radio and the Kennedy School of Government found that ninety-three percent of those polled favored or strongly favored "making students meet adequate academic standards to be promoted or graduated."[56] Similarly, in 1994, Public Agenda found that eighty-eight percent of those polled thought that students should not be allowed to receive a high school diploma unless they had demonstrated the ability to read and write English well, while eighty-two percent endorsed "very clear guidelines on what students should learn."[57]

Parents, in particular, have consistently expressed the view that testing promotes accountability and leads to improvement in the

---

52. RAVITCH, *supra* note 2, at xiv (citing JEAN JOHNSON & STEVE FARKAS, GETTING BY: WHAT AMERICAN TEENAGERS REALLY THINK ABOUT THEIR SCHOOLS 35 (1997)) (noting that, in polls, students state "unequivocally that they would work harder if more were expected of them"); *see also* Heuhert, *supra* note 34 (summarizing the argument that tests will improve incentives).

53. Achieve, Inc., *supra* note 12, at 4 (arguing that without tests, inequities between between rich and poor will likely persist beneath the radar); Alexakis, *supra* note 17, at 29.

54. CIZEK, *supra* note 33, at 1 ("Results of opinion surveys from 30 years ago and those conducted today reveal broad and durable support for *even more* achievement testing. Contrary to assertions that tests are foisted upon the public by self-serving politicians, the evidence is clear that consumers of U.S. education favor testing.").

55. Gallup Organization, *Education: A Vital Issue in Election 2000*, (Oct. 2, 2000), *at* http://www.gallup.com/poll/releases (last visited Mar. 1, 2001).

56. Public Agenda Online, *Education: Major Proposals*, *at* http://www.public-agenda.org/issues/major_proposals_detail2.cfm?issue_type=education&proposal_graphic=mp3.gif (last visited Feb. 3, 2002).

57. *See* RAVITCH, *supra* note 2, at 171 (citing a 1994 Public Agenda survey); *see also* Diane Ravitch, *Editorial, Who Says Parents Oppose Standards-Based Reform?*, RECORD, Feb. 19, 2001, at L03 (noting that teachers endorse standards-based reforms by a margin of seventy-three to nineteen percent and that support for standards-based reform is strongest among teachers in urban schools and among black and Hispanic teachers), 2001 WL 5238721.

educational system.[58] And surveys indicate that support for high-stakes exams transcends race and class. In a 2000 survey conducted by the Business Roundtable, two-thirds of African-American parents and two-thirds of white parents stated that they favored graduation exams; support for such exams was even higher among Hispanic parents, eighty-two percent of whom favored such tests.[59]

D. The Education Reform Landscape and the Test Score Gap

Currently, most states administer some form of standardized educational test on a regular basis,[60] and twenty-eight states either require students to pass a graduation exam before receiving a diploma or are in the process of instituting such a program.[61]

Not surprisingly, accountability regimes and testing practices vary widely from state to state. Some states, such as Florida, adopted exam requirements in the 1970s. Other states, like Alaska, have

---

58. *See* June Kronholz, *State High School Graduation Tests Favored by Most Parents, Says Poll*, WALL ST. J. INTERACTIVE ED., Sept. 13, 2000 (noting that "[m]ost parents think their children should have to pass statewide tests in order to graduate" and that eighty percent of the public supports the use of high-stakes tests in which students are given "several" attempts to pass such exams), *at* http://interactive.-www.wsj.com/archive/retrieve.cgi?id=SB96880-4416727439474.djm; *see also* CIZEK, *supra* note 33, at 1 (discussing the consistent public support for high-stakes tests); PUB. AGENDA, SURVEY FINDS LITTLE SIGN OF BACKLASH AGAINST ACADEMIC STANDARDS OR STANDARDIZED TESTS (2000) (noting that eighty-two percent of parents say their school district's effort toward higher academic standards is "careful and reasonable"), *at* http://www.publicagenda.org/aboutpa/pdf/standards-backlash.pdf.

59. Kronholz, *supra* note 58; *see also Most Latinos Baek TAAS Exit Exams Despite Bias Fears*, FORT WORTH STAR-TELEGRAM, July 24, 2000, at 11. Consistent with these survey results is evidence from some states that the school districts which are most supportive of high-stakes graduation exams are those poorer districts which are most likely to have high short-term failure rates. Alexakis, *supra* note 51, at 29 (reporting that it is the primarily white, affluent Massachusetts suburbs that have complained about the Massachusetts graduation test while blue collar, racially mixed communities tend to support the exam); Rick Klein, *MCAS Criticism Rising in Suburbs Contrast with View from Urban Areas*, BOSTON GLOBE, Nov. 3, 2000, at A1.

60. Every state except Iowa has established academic standards in at least some subjects. Jodi Wilgoren, *State School Chiefs Fret over U.S. Plan to Require Testing*, N.Y. TIMES, July 17, 2001, at A1. Forty-eight states (all but Iowa and Nebraska) have statewide testing programs, most of which are aligned with the academic standards. Achieve, Inc., *supra* note 12, at 3; Thernstrom, *supra* note 14; Wilgoren, *supra*.

61. *See* COUNCIL OF CHIEF STATE SCH. OFFICERS, KEY STATE EDUCATION POLICIES ON K-12 EDUCATION: 2000, at 16 (2000) (listing twenty-eight states that either currently require graduating students to pass an exit exam or that are in the process of developing or instituting exit exams); Margaret E. Goertz & Marc C. Duffy, *Assessment and Accountability Across the Fifty States*, CPRE POLICY BRIEFS, May 2001, at 5, *available at* http://www.cpre.org/publications/rb33.pdf ("[B]y 2008, students in 28 states will have to pass a state examination to graduate"); Scott S. Greenberger, *Standardized Tests Pose Questions for States*, BOSTON GLOBE, Oct. 8, 2000, at B1 (reporting that, as of October 2000, twenty-four states required students to pass graduation exams as a condition for receiving a diploma).

passed statewide accountability and testing programs only recently.[62] While most state graduation exams are tests of basic skills, designed to ensure that all graduates have mastered certain core competencies,[63] some state exams seek to raise the bar by testing students on their mastery of more rigorous academic content.[64]

Despite such differences, however, state accountability and assessment policies share some important characteristics. To begin with, most accountability regimes are, in effect, contractual arrangements between states and localities whereby the state agrees to: (1) provide increased financial resources and technical assistance; (2) relax the state regulatory burdens; and (3) increase local autonomy and flexibility. In exchange for these benefits, localities essentially agree to allow the state to conduct regular testing of all students and regular evaluations (based on test scores and other factors) of each school and/or district.[65] The results of these assessments are made available to the public and are used by the state to reward schools or districts that have improved their performance, and to assist students and schools that are struggling by providing remedial education for the students[66] and targeted financial assistance, peer reviews, and school improvement planning for the schools.[67]

---

62. Florida implemented its first high school exit exam during the 1977-1978 school year. Alaska's testing regime took effect in 2000. NAT'L GOVERNORS ASS'N, *supra* note 31, at 1.

63. *Id.* at 3 (noting that most existing high school graduation tests measure basic competency rather than mastery of more rigorous standards); CTR. FOR EDUC. REFORM, *supra* note 30 ("While it's important to keep in mind that the quality and rigor of each test can vary from state to state, the evidence suggests that these tests are not excessively difficult and primarily measure the acquisition of basic skills and knowledge appropriate to each grade level.").

64. Massachusetts, for example, employs a test, which seeks to measure mastery of rigorous academic subjects as well as core competencies. *See* REACHING HIGHER, (Mass. Insight Educ., Boston, Mass.), Spring 2001, *at* http://www.massinsight.com/meri/pdf-files/Reaching%20Higher%202001.PDF. Some testing advocates argue that the implementation of "tests worth teaching to" will raise standards and improve tbe quality of instruction. *See, e.g.*, Anemona Hartocollis, *New Reading Test Was Worth All the Anxiety, Experts Say*, N.Y. TIMES, Jan. 18, 1999, at B1 (explaining the view of some experts that, by raising the bar for what children should know, New York's reading test is a test worth teaching to); William C. Symonds, *How to Fix America's Schools*, BUS. WK., Mar. 19, 2001, at 67 (explaining that, while not all tests are created equal, the use of sophisticated tests that require students to use complex problem-solving skills can actually improve teaching); *see also* RAVITCH, *supra* note 2, at 25-26 (arguing that national standards improve student achievement by clearly delineating what should be taught and what must be learned to succeed).

65. *See* Susan H. Fuhrman, *The New Accountability*, CRPE POLICY BRIEFS, Jan. 1999, at 1-2, *available at* http://www.cpre.org/Publications/rb27.pdf (last visited June 21, 2001).

66. *See* NAT'L GOVERNORS ASS'N, *supra* note 31, at 4. Tutoring and extended learning time are the most common forms of remediation. *Id.*

67. *See id.*; Fuhrman, *supra* note 65, at 2, 5.

Unlike the older generation of standardized tests, most of the newer tests do not rely exclusively on multiple-choice questions. Instead, they combine a variety of formats, including short-answer questions and essays.[68] Because the goal of the tests is to measure and improve student performance, all of the states that administer high-stakes exit exams allow students multiple opportunities to master the material and pass the exams.[69]

In recent years, the educational accountability systems in place in Massachusetts and Texas have received much public attention and scrutiny. Those exams—and the racial and ethnic academic gap that they reveal—are emblematic of statewide testing regimes throughout the country and provide a window into state testing policies for purposes of analysis.

1. Texas

The Texas accountability system[70] was enacted in response to a series of court challenges to the constitutionality of the financing of the Texas educational system.[71] Under this accountability system, the state establishes academic standards that all students and schools are required to meet. In order to assess performance on state educational standards, Texas tests all students annually in grades three through eight.[72] Texas law requires that all public school students, in addition to completing the required high school curriculum, pass an exit-level examination of academic skills in order to receive a high school diploma.[73]

The current exit-level Texas Assessment of Academic Skills ("TAAS") assesses tenth grade students in reading, writing, and

---

68. *See* Achieve, Inc., *supra* note 12, at 4.

69. NAT'L GOVERNORS ASS'N, *supra* note 30, at 2. In Massachusetts, for example, students are allowed to retake the exit-level exam five times during the eleventh and twelfth grades. *See* Scott S. Greenberger, *MCAS Retest Plan Approved*, BOSTON GLOBE, Jan. 24, 2001, at B4; REACHING HIGHER, *supra* note 64. Some states even allow students to retake the exam after their scheduled graduation date. Georgia, for example, grants certificates of performance to those students who have met all of the requirements for graduation other than passing the exam. These students have unlimited opportunities to take the exam, and, upon passage, are awarded regular diplomas. *See* NAT'L GOVERNORS ASS'N, at 64.

70. *See* TEX. EDUC. CODE ANN. §§ 39.021-39.185 (Vernon Supp. 2002).

71. In 1989, the Texas Supreme Court held that the state's puhlic financing system violated the Texas Constitution. Edgewood Indep. Sch. Dist. v. Kirhy, 777 S.W.2d 391, 397 (Tex. 1989).

72. In 2003, Texas will expand its testing program to include all students in grades three through eleven. *See More Difficult TAKS to Replace TASS in '03*, HOUSTON CHRON., June 20, 2001, at 22 [hereinafter *More Difficult TAKS*], 2001 WL 23608957.

73. TEX. EDUC. CODE ANN. § 39.025 (2000); 19 TEX. ADMIN. CODE § 101.7 (Testing Requirements for Graduation), *available at* http://lamb.sos.state.tx.us/tac/index.html (last visited May 13, 2002).

mathematics.[74] During the 2002-2003 school year, Texas will replace the TAAS exam with the Texas Assessment of Knowledge and Skills ("TAKS") and will begin requiring students in the eleventh grade to pass exit-level tests in English/language arts, mathematics, social studies, and science in order to graduate.[75]

Texas law requires that local school districts provide failing students with remedial help in the specific subject area where the student encountered difficulty.[76] Under the TAAS system, students who do not pass the exit-level exam on the first try are provided seven additional opportunities to pass the exam before the end of their senior year.[77] Under the new TAKS system, a student who does not pass the exit-level exam on the first administration of the test may continue to take the test until he passes. Even those who have been denied diplomas can continue to take the exam and will receive a diploma upon passing.[78]

In addition to measuring the performance of individual students, Texas evaluates institutional performance through a variety of measures, including TAAS scores and dropout rates, as well as attendance rates, rates of improvement in TAAS sores compared to other like institutions, and SAT and ACT participation levels and results.[79] A subset of this information is used to rate individual districts and schools as "Exemplary," "Recognized," "Academically Acceptable," and "Academically Unacceptable" and to reward schools and districts that perform well[80] and to provide assistance (up to and including receivership) to consistently underperforming schools.[81]

---

74. Beginning in 2003, Texas will replace the TAAS with the Texas Assessment of Knowledge and Skills ("TAKS"). The TAKS exam reflects a change in Texas's curriculum standards and will cover more suhjects than the current exam. *See More Difficult TAKS*, *supra* note 72.

75. TEX. EDUC. CODE ANN. § 39.025.

76. *See* 19 TEX. ADMIN CODE § 101.11 (Remediation), *available at* http://lamb.sos.state.tx.us/tac/index.html (last visited May 13, 2002); *see also* GI Forum v. Tex. Educ. Agency, 87 F. Supp. 2d 667, 673 (W.D. Tex. 2000).

77. *GI Forum*, 87 F. Supp.2d at 673.

78. TEX. EDUC. CODE ANN. § 39.025.

79. *See* 2001 TEXAS ACCOUNTABILITY MANUAL, *at* http://www.teas.state.tx.us/perfreport/account/2001/manual (visited June 13, 2001).

80. TEX. EDUC. CODE ANN. §39.072 (Vernon Supp. 2002); 2001 TEXAS ACCOUNTABILITY MANUAL, *supra* note 79, § III. In 2001, for example, a campus was rated "Exemplary" if at least 90% of students overall and 90% of students in each of several demographic subgroups (black, Hispanic, white, and economically disadvantaged) meeting minimum size requirements passed the state's reading, writing, and mathematics exams and if the dropout rate was 1% or less for students overall and within each group. *Id.*

81. Schools that perform poorly receive visits from a peer review team in the following school year and must develop and implement an improvement plan to address areas of poor performance. If the district or campus receives the lowest rating over two or more consecutive

In addition, Texas keeps records on how various demographic groups perform on the exam. If one subgroup fails to meet minimum performance standards, a school or district will receive a low accountability rating.[82] This approach helps schools and districts appreciate tbe academic needs of various demographic subgroups so that they can work toward closing the academic achievement gap. Such record keeping also allows the state to allocate resources where they are most needed in a way that will help eliminate the gap.[83]

The passing score for TAAS was originally set at sixty percent and was later raised to seventy percent.[84] In deciding upon a passing score, the Texas Education Agency ("TEA") was aware of projections that minority students would not fare as well on the exam as their peers.[85] Nevertheless, the TEA determined that the test should be imposed as an objective measurement of mastery in order to eliminate the inconsistent and subjective teacher evaluations of students, which members of the TEA believed harmed minority students by allowing inflated grades to mask gaps in learning.[86]

As predicted, between the first administration of TAAS in 1994 and the spring 2001 administration of the test, white students consistently outperformed black and Latino students on all portions of the exam.[87] For example, on the 2001 TAAS, 89% of white students passed all portions of the exit-level exams administered that year, as compared to 68% of African-American students and 70% of Latino students.[88] This gap was evident not only in overall pass rates on the 2001 exit exam, but also in each of the three core subject

---

years, the level of state intervention increases. *See* TEX. EDUC. CODE ANN. § 39.131 (Vernon Supp. 2002); 2001 TEXAS ACCOUNTABILITY MANUAL, *supra* note 79, § VIII.

82. *See GI Forum*, 87 F. Supp. 2d at 674.

83. Achieve, Inc., *supra* note 12, at 5.

84. *See GI Forum*, 87 F. Supp. 2d at 673.

85. *Id.* ("When it implemented the TAAS test, the TEA projected that, with a 70-percent cut score, at least 73 percent of African Americans and 67 percent of Hispanics would fail the math portion of the test; at least 55 percent of African Americans and 54 percent of Hispanics would fail the reading section; and at least 62 percent of African Americans and 45 percent of Hispanics would fail the writing section. The predictions for white students were 50 percent, 29 percent, and 36 percent, respectively.").

86. *Id.*

87. *Id.* at 675 ("[I]n every administration of the TAAS test since October 1990, Hispanic and African American students have performed significantly worse on all three sections of the exit exam than majority students."); *see also* TEX. EDUC. Agency, STUDENT PERFORMANCE REPORT EXECUTIVE SUMMARY 16-17 (2000) (reporting pass rates by subject and race), *available at* http://www.tea.state.tx.us/student.-assessment/researchers.html (last visited Mar. 2, 2002).

88. STUDENT ASSESSMENT DIV., TEX. EDUC. AGENCY, TEXAS ASSESSMENT OF ACADEMIC SKILLS, PERCENT MEETING MINIMUM EXPECTATIONS SPRING 1994-SPRING 2001, *at* http://www.tea.state.tx.us/student.assessment/reporting/results/swresults/spring/g10all.html (last visited Aug. 5, 2001).

areas—reading, writing, and mathematics. Thus, 96% of white students passed the reading portion of the 2001 exam, as compared to 83% of African-American students and 83% of Latino students. Ninety-four percent of white students passed the 2001 writing exam, as compared to 85% of African-American students and 83% of Latino students. Pass rates on the 2001 mathematics portion of the exam also reflected the racial achievement gap: 94% of white students passed the mathematics exit exam, as compared to 79% of black students and 83% of Hispanic students.[89] Cumulative pass rates also display a similar pattern. The 1998 data show cumulative failure rates of 17.6% for black students and 17.4% for Latino students, as compared to a 6.7% cumulative failure rate for white students.[90]

But there is some good news as well. Although African-American and Hispanic students in Texas still lag behind their white peers, they are beginning to close the achievement gap.[91] In the eight administrations of the exit-level TAAS between 1994 and 2001, the pass rate for black students rose by 40%; during the same time period, the pass rate for Latinos rose by 36%.[92] The black-white gap on the test shrunk from 36 percentage points on a single administration of the test in 1994 to 21 percentage points on a single test administration in 2001; the gap between Hispanic students and white students also shrunk from 36 percentage points in 1994 to 19 percentage points in 2001.[93] Thus, while those black and Latino students who are unable (after seven attempts) to pass the test and obtain a diploma are burdened by the Texas accountability regime, other black and Latino students are clearly aided by the system and the resulting increase in levels of academic success.

2. Massachusetts

In 1993, the Massachusetts Supreme Judicial Court held in *McDuffy v. Secretary of Executive Office of Education*[94] that the

89. *Id.*

90. Heubert, *supra* note 34.

91. *GI Forum*, 87 F. Supp. 2d at 675 (observing that "minority students [in Texas] have continued to narrow the passing rate gap at a rapid rate" and that "these gains are reflected not only on TAAS scores, but also on "other measures of academic progress, such as the National Assessment of Educational Progress").

92. STUDENT ASSESSMENT DIV., *supra* note 88; *see also* TEX. EDUC. AGENCY, *supra* note 87 (noting that percentage of African-American students passing the exit-level TAAS exam went from 28% in 1994 to 67% in 2000, and that the percentage of Latino students passing the exit-level TAAS increased from 34% in 1994 to 70% in 2000).

93. STUDENT ASSESSMENT DIV., *supra* note 88.

94. In *McDuffy*, the Massachusetts Supreme Judicial Court held that "the Commonwealth has a duty to provide an education for *all* its children" and that by allowing poorer school

Commonwealth's educational financing system violated the education provisions of the Massachusetts Constitution.[95] The decision served as the catalyst for passage by the state legislature of the Massachusetts Education Reform Act of 1993 (the "Massachusetts Reform Act"),[96] which not only increased the amount of money Massachusetts spent on education, but also required the Commonwealth to conduct statewide assessments of student competency in relation to standards set by the Massachusetts Board of Education.[97] In addition, the Massachusetts Reform Act made satisfactory performance on the state assessment a condition of receiving a high school diploma.[98]

The Massachusetts Comprehensive Assessment System ("MCAS"), the statewide testing program adopted by the Massachusetts Board of Education pursuant to the Massachusetts Reform Act, is one of the country's most ambitious exams and is aligned with rigorous state academic standards.[99] The MCAS tests Massachusetts students in various grades in four content areas—English/language arts, mathematics, science and technology, and history and social science—with a high school exit exam administered in the tenth grade.[100] Students who take the MCAS receive a scaled numeric score and a designation (based on the numeric score) as "Advanced" (260-280), "Proficient" (240-259), or

districts to operate with significantly less money than middle class school districts, the Commonwealth violated Section 2 in Chapter 5 of Part II of the Massachusetts Constitution. 615 N.E.2d 516, 548, 553-54 (Mass. 1993).

95. *Id.* at 553-54. The Massachusetts Constitution provides, "it shall be the duty of the legislatures and magistrates in all future periods of this Commonwealth, to cherish the . . . public schools and grammar schools in the towns. . . ." MASS. CONST. pt. 2, ch. 5, § 2.

96. MASS. GEN. LAWS ANN. ch. 69, §§ 1-36 (West 1996).

97. Under the Massachusetts Reform Act, the Board of Education is empowered to establish "curriculum frameworks"—that is, "academic standards for the core subjects of mathematics, science and technology, history and social science, English, foreign languages and the arts . . . . [for] grades kindergarten through twelve." MASS. GEN. LAWS ANN. ch. 69, §§ 1D, 1E. These frameworks "set forth the skills, competencies and knowledge expected to be possessed by all students at the conclusion of individual grades or clusters of grades" and are designed "to set high expectations of student performance and to provide clear and specific examples that embody and reflect these high expectations." *Id.*

98. MASS. GEN. LAWS ANN. ch. 69, § 1D.

99. Heubert, *supra* note 34; REACHING HIGHER, *supra* note 64.

100. Massachusetts must annually test students in the fourth, eighth, and tenth grades. In 2001, Massachusetts added tests for third, fifth, sixth, and seventh grades. MASS. DEP'T OF EDUC., OVERVIEW OF THE MCAS 2001 TESTS 19 (2001), *at* http://www.doe.mass.edu/mcas/2001/overview/complete.pdf (last visited May 13, 2002); Massachusetts Department of Education, *Overview of the MCAS: Frequently Asked Questions, at* http://www.doe.mass.edu/mcas/over-view_jag.html#jag2 (last visited Apr. 6, 2002).

"Needs Improvement" (220-239). Prior to the tenth grade, students who score below 220 receive an "Academic Warning."[101]

Beginning with the class of 2003, students in Massachusetts are required to pass both the tenth grade English/language arts section and the tenth grade mathematics portion of the test in order to receive a high school diploma.[102] For the class of 2003, the Board has set the passing score at 220 (the lowest score in the "Needs Improvement" category) for each portion of the test. The Board intends to raise the threshold scaled score required for passage of the exit exams in future years.[103]

Students who fail either the English or the mathematics portion of the exit-level MCAS are provided remedial assistance[104] and may retake the exam *four* additional times before their scheduled graduation date.[105] Unlike the initial tenth grade exit exam, which classifies students by skill level, the MCAS retests are pass/fail exams that focus only on the most essential skills and do not include the toughest examination questions.[106] In January 2002, the Massachusetts Board of Education instituted an appeals process to allow—in compelling cases—students with good grades in school, but failing MCAS scores, to prove through other means that they have

101. Linda Bock, *Public Opinion Still Mixed over MCAS*, WORCESTER TELEGRAM & GAZETTE, Nov. 27, 1999, at A2; Greenberger, *supra* note 61.

102. MASS. GEN. LAWS ANN. ch. 69, § 1D ("Satisfaction of the requirements of the competency determination shall he a condition for high school graduation."); MASS. BD. OF EDUC., *supra* note 100, at 19. In the spring of 2001, all Massachusetts tenth graders (who are scheduled to graduate in 2003) took the exit-level MCAS for the first time. Scott S. Greenberger, *For First Time, Students Take MCAS Test for Real*, BOSTON GLOBE, Apr. 12, 2001, at B5, 2001 WL 3928726.

103. Massachusetts Comprehensive Assessment System and Standards for Competency Determination, MASS. REGS. CODE tit. 603, § 30.03.

104. Massachusetts provides funding so that local districts can provide remedial help for failing students. In 2000, the state legislature appropriated $20 million for additional instruction for failing students. In 2001, the Massachusetts Legislature appropriated $40 million for the same purpose. Local districts are free to spend this money in the manner in which they see fit. *See* Editorial, *The Importance of MCAS*, BOSTON GLOBE, Nov. 15, 2000, at A26; REACHING HIGHER, *supra* note 64; Massachusetts Department of Education, *Fact Sheets on Public Education in Massachusetts*, Apr. 2001), at 1, *at* http:// www.doe.mass.edu/edre-form/erfacts/fact-sheet.01.pdf. Thus, in Boston, students who have not met the standards for their grades participate in the Transition Services Program—a mix of before and after school, Saturday, and summer school remedial classes. (All of these are programs that would not have been implemented without the test). Editorial, *supra*; REACHING HIGHER, *supra* note 64.

105. MASS. DEP'T OF EDUC., *supra* note 100, at 19; Greenberger, *supra* note 61; REACHING HIGHER, *supra* note 64.

106. Greenberger, *supra* note 61.

mastered the material tested by the MCAS, and thereby receive a diploma.[107]

As in the case of Texas, the Massachusetts testing regime was adopted as part of an overall reform package and in response to litigation challenging the constitutionality under state law of the state's public financing system.[108] The Massachusetts Reform Act requires the "Massachusetts Department of Education to evaluate whether schools and districts are improving student performance based on the learning standards contained" in state frameworks.[109] According to state regulations, each school district in which more than twenty percent of the students score below "Proficient" on the MCAS is required to submit an MCAS success plan to the state's Department of Education describing the district's strategies for helping each student to master the skills, competencies, and knowledge mandated by the state frameworks.[110] Schools that fail to improve performance may be deemed "under-performing." These schools are assigned priority status for state assistance, are placed on Academic Warning until the end of the following rating cycle, and are subject to monitoring by a state-appointed fact-finding team. Moreover, "under-performing" schools are required to submit an improvement plan to the Board of Education setting forth specific goals for improvement, specific means for attaining such goals, and a timetable for implementation. If an "under-performing" school fails to demonstrate significant improvement in student performance within twenty-four months after approval of a remedial plan by the Board, the Board may

---

107. Massachusetts Comprehensive Assessment System and Standards for Competency Determination, MASS. REGS. CODE tit. 603, § 30.05 (performance appeals), *available at* http://www.doe.mass.edu/lawsregs/a603cmr30.html#05 (last visited June 24, 2002); Massachusetts Department of Education, Press Release, MCAS Appeals Process Unanimously Approved (Jan. 23, 2002) ("[I]n order to be considered, students will need to have: taken the grade 10 level MCAS at least three times; scored at least a 216 on the exam at least once; maintained a 95 percent attendance level during the previous school year and the year of the appeal; and participated in [remedial] services made available by the school. . . . The appeal will contain evidence of the student's knowledge and skills in the subject area, including teacher recommendations, the student's grades . . ., work samples and scores on the standardized tests in the subject areas."), *at* http://www.doe.mass.edu/news/news.asp?id=488 (last visited Feb. 8, 2002).

108. As in the case of Massachusetts and Texas, other states have implemented accountability systems in response to lawsuits challenging as unconstitutional the states' system for allocating money to the schools. *See, e.g.*, MARC S. TUCKER & JUDY B. CODDING, STANDARDS FOR OUR SCHOOLS: HOW TO SET THEM, MEASURE THEM, AND REACH THEM 230 (1998) (explaining that the Kentucky legislature passed that state's accountability regime in response to a 1989 decision by the Kentucky Supreme Court holding the state educational system unconstitutional).

109. Massachusetts Department of Education, *Background on the MCAS Tests of May 1998*, *at* http://www.doe.mass.edu/mcas/1998/bg/section9.html (last visited May 13, 2002).

110. MASS. GEN. LAWS ANN. ch. 69, § 1I (West 1996); Under-Performing Schools and School Districts, MASS. REGS. CODE, tit. 603, § 2.03 (WESTLAW through Mar. 29, 2002).

declare the school to be chronically "under-performing" and may intervene or take over management of the school. [111]

Consistent with the results in other jurisdictions, black and Latino students in Massachusetts have failed the MCAS at higher rates than students from other ethnic backgrounds.[112] The 2001 exit-level exam was the first administration of the test to "count" vis-à-vis the diploma sanction. On that exam, 40% of African-American tenth graders and 48% of Latino tenth graders failed the English/language arts portion of the exam, as compared to 20% of Asian-American tenth graders and 12% of white tenth graders. Fifty-two percent of African-American tenth graders and 58% of Latino tenth graders failed the mathematics portion of the test, as compared to 15% of Asian-American tenth graders and 18% of white tenth graders.[113] Overall, 77% of white tenth graders earned a competency determination on a single administration of the test in 2001 (meaning that they passed both the English and mathematics portion of the exam), as compared to 29% of Hispanics and 37% of African-Americans.[114]

E. Political Opposition to High-Stakes Tests

Standardized tests have long been attacked by many on the political left who instinctively charge racism whenever facially neutral policies lead to different demographic outcomes. Indeed, as early as 1969, the Association of Black Psychologists issued a statement calling for a complete moratorium on "all testing of Black people" until "more equitable" tests could be developed.[115] In 1980, the NAACP was a signatory to a "Statement to Urge a Ban on Standardized Testing of

---

111. MASS. REGS. CODE tit. 603, § 2.03.

112. *See* Karen Crummy, *Minority Leaders Decry MCAS as Inequitable Testing Tool*, BOSTON GLOBE, Sept. 1, 2001, at A15.

113. Massachusetts Department of Education, *Spring 2001 MCAS Tests: State Results by Race/Ethnicity and Student Status*, *at* http://www.doe.mass.edu/MCAS/2001/results/re_ss.pdf (last visited Feb. 7, 2002). These results reveal a significant improvement since the previous year when 60% of black tenth graders and 66% of Latino tenth graders failed English/language arts, and 77% of black tenth graders and 79% of Latino tenth graders failed mathematics. *Id.* Nevertheless, the gap between the performance of non-Asian minorities and white students remains relatively constant between 2000 and 2001. *Id.*

114. *Id.* On the first administration of the MCAS retest offered in December 2001, 48% of students who previously failed the exam passed the English portion, and 31% passed the mathematics retest. Anand Vaishnav & Sandy Coleman, *Many Pass the MCAS Retest But Thousands Must Try Again*, BOSTON GLOBE, Mar. 1, 2002, at B1, 2002 WL 4114313. Added with scores from the Spring 2001 MCAS, the results of the retest indicate that 14% of the class of 2003 still must pass English and 21% still must pass math. These students have three more chances to pass the test. *Id.*

115. *See* GREGORY CAMILLI & LORRIE A. SHEPARD, METHODS FOR IDENTIFYING BIASED TEST ITEMS 6 (1994) (mentioning the Association of Black Psychologist's resolution).

Young Children."[116] And in 1991, the NAACP and the Mexican American Legal Defense and Educational Fund ("MALDEF") joined with a coalition of antitesting organizations to protest federal achievement testing proposed by President George H.W. Bush.[117]

Such charges continue to be levied against the current wave of high-stakes educational assessments, despite broad public support for the exams among the public at large, including minority parents.[118] Indeed, in many of the states that have proposed making passage of an exit exam a condition of receiving a high school diploma, activists purporting to represent the interests of ethnic and racial minorities have opposed the idea, arguing that such tests are "biased" or "discriminatory" against blacks and Latinos.[119]

In Massachusetts, for example, the high failure rates of minority students have fueled calls from some civil rights activists to suspend or eliminate the exams.[120] On the national level, Elaine Jones, the President of the NAACP Legal Defense and Educational Fund, has compared high-stakes testing to the literacy tests used in the days of Jim Crow to prevent African-Americans from voting.[121] And in November 1999, the National Urban League criticized proponents of high-stakes testing for "using low-achievers as cannon fodder in the education accountability wars."[122]

---

116. *See* THERNSTROM & THERNSTROM, *supra* note 8, at 363 (describing the resolution).

117. Cooper, *supra* note 31. Some teachers unions have also expressed strong opposition to high-stakes graduation exams. For example, the National Education Association has passed a resolution condemning the use of any standardized test whose "results are used to compare students, teachers, programs, schools, communities, and states." *See* National Education Association, *NEA 2001-2002 Resolutions*, at B-57, *at* http://www.nea.org/resolutions (last visited May 13, 2002). The resolution also condemns the use of any standardized test for "high-stakes decision making," i.e., any test with consequences. *Id.* In Massachusetts, the Massachusetts Teachers Union in November 2000 launched a $600,000 advertising campaign designed to quash public support for the test. *See* Scott S. Greenberger, *Teachers Air Anti-MCAS Ad*, BOSTON GLOBE, Nov. 9, 2000, at B9; Frank Phillips, *Cellucci Blasts Teachers on Tests*, BOSTON GLOBE, Nov. 9, 2000, at B1; *see also* Sara Mosle, *Scores Count*, N.Y. TIMES, Sept. 8, 1996, § 6 (Magazine), at 41.

118. *See supra* notes 54-59 and accompanying text.

119. *See* Gary Orfield & Johanna Wald, *Testing, Testing: The High-Stakes Testing Mania Hurts Poor and Minority Students the Most*, NATION, June 5, 2000, at 38, 39 (arguing that high-stakes tests "both discriminate against poor and minority students and are educationally unsound"); *see also* Alexakis, *supra* note 17, at 29 (describing such claims).

120. Crummy, *supra* note 112 (noting that the MCAS has been called the "new segregation"). Several years ago, the Justice Department threatened to sue the State of Massachusetts because of the disparate racial outcomes of the MCAS exam. *See* Abigail Thernstrom, *A Taboo Erodes: The Truth About Blacks and Education*, NAT'L REV., Dec. 20, 1999, at 22.

121. *See Critics Target Wrong Culprit as Minorities' Test Scores Lag*, USA TODAY, May 30, 2000, at A16; Whitmire, *supra* note 15.

122. Gail Russell Chaddock, *Adverse Impact?*, CHRISTIAN SCI. MONITOR, Nov. 30, 1999, at 14.

Some Hispanic civil rights groups have been equally critical of standardized exit exams. In a prepared statement on its web page, the National Council for La Raza argues that "[s]tandardized tests contain potential linguistic and cultural biases, and when used for 'high stakes' purposes such as being promoted to the next grade level, they can have disproportionately negative effects on disadvantaged, low-income, and limited English-proficient students."[123] And, more dramatically, MALDEF brought suit against the State of Texas in 1997, arguing that the TAAS exam unlawfully discriminated against black and Hispanic students.[124] Although MALDEF lost the case,[125] it remains committed to the use of litigation to thwart the use of high-stakes testing.[126] Other special interest groups and some legal commentators also seek to use the courts to eliminate high-stakes educational tests.[127] Indeed, in May 2001, the Harvard Civil Rights Project (an antitesting group) sponsored a closed-door conference for attorneys from various interest groups on potential litigation strategies to block the implementation of such exams or their diploma sanctions.[128]

---

123. *See* Roberto J. Rodriguez, *In Search of High Academic Achievement: The Policy Drive to End Social Promotion, at* http://nclr.policy.net/proactive (last visited Jan. 16, 2001).

124. *See, e.g.*, GI Forum v. Tex. Educ. Agency, 87 F. Supp. 2d 667 (W.D. Tex. 2000); Cooper, *supra* note 19; *see also* Erik V. v. Causby, 977 F. Supp. 384, 389 (E.D.N.C. 1997) (denying preliminary injunction in litigation alleging that minority students were more adversely affected by the school district's requirement that students satisfy specific standards as a condition of promotion from one grade to the next).

125. *See* discussion *infra* notes 244-50 and accompanying text.

126. *See* Press Release, Mexican American Legal Defense and Educational Fund, Judge Errs in Upholding Texas TAAS High School Exit Test, (Jan. 7, 2002), *at* http://www.maldef.org/news/press.cfm?ID=26.

127. *See, e.g.*, Green, *supra* note 19 (promoting the use of disparate impact lawsuits to enjoin educational institutions from relying on test scores with disproportionate racial results); Losen, *supra* note 19 (arguing in favor of disparate impact litigation to eliminate "tracking" students by ability); Silverstein, *supra* note 19 (arguing that disparate impact litigation should be mounted against the use of tests on which the average male test-taker outperforms the average female test-taker); Elul, *supra* note 16, at 495 (arguing in favor of disparate impact litigation to challenge high-stakes educational assessments); Keith Ervin, *NAACP, Schools Team Up to Reduce Racial Gap, Group Puts on Hold Suit Against District*, SEATTLE TIMES, Dec. 19, 2001, at B1 (noting that the Seattle NAACP continues to contemplate a lawsuit against the Seattle school district in response to racially disproportionate test results); *see also* ALFIE KOHN, THE CASE AGAINST STANDARDIZED TESTING, RAISING THE SCORES, RUINING THE SCHOOLS (urging parents and students to "consider filing a lawsuit" against tests which are "inherently discriminatory"); Chaddock, *supra* note 122, at 14 (quoting Gary Orfield of the Harvard Civil Rights Project as suggesting that special interest groups were prepared to launch a number of lawsuits over the use of high-stakes tests).

128. *See* Harvard Civil Rights Project, *Template for Debating/Evaluating MCAS and Other High-Stakes Policies* (on file with the author); Harvard Civil Rights Project, *Schedule of Conferences, at* http://www.law.harvard.edu/civilrights/conference/index.html (last visited Aug. 5, 2001).

III. THEORIES OF DISCRIMINATION

The minority achievement gap has caused testing critics to level sweeping public charges of "discrimination" against high school graduation exams such as the TAAS and the MCAS, and has prompted various activists and legal commentators to threaten litigation against educational institutions that employ such exams. In light of these significant, high-profile charges of discrimination, it is necessary to recall and understand the prevailing theories of discrimination and to outline the appropriate legal framework for resolving claims of discrimination in educational testing. Part III.A begins that analysis by outlining the two principal models for addressing discrimination—disparate treatment and disparate impact—and by explaining how the two approaches differ. Part III.B then sketches the major theories that have been offered to justify the disparate impact model. Lastly, Part III.C outlines the burdens of proof in a typical disparate impact case, using Title VII of the Civil Rights Act of 1964, which deals with discrimination in employment, as the prototype for disparate impact litigation. Parts IV and V will then explain how disparate impact litigation works in the educational testing context and argue that its applicability to high-stakes educational assessments is unsound as a matter of law and public policy.

A. Two Models of Addressing Discrimination: Disparate Treatment Versus Disparate Impact

There are two basic models for addressing claims of discrimination: disparate treatment and disparate impact. The disparate treatment model attempts to expose and punish intentional discrimination. Under this model, proof of discriminatory motive is critical.[129] Thus, a plaintiff in a disparate treatment case must demonstrate that the defendant treated her differently from other similarly situated individuals "because of" her race, sex, national origin, or other protected status.[130]

---

129. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 335 n.15 (1977) ("[Disparate treatment occurs where the defendant] simply treats some people less favorably than others because of their race, color, religion, sex, or national origin. Proof of discriminatory motive is critical, although it can in some situations be inferred from the mere fact of differences in treatment.").

130. In cases of employment discrimination, motive can be proved by way of direct evidence or proof of pretext. *See* Texas Dep't of Cmty. Affairs v. Burdine, 450 U.S. 248, 252-56 (1981); McDonnell Douglas Corp. v. Green, 411 U.S. 792, 805, n.18 (1973) (outlining the hurdens of proof in a disparate treatment case under Title VII).

By contrast, the disparate impact model (or the "effects test," as it is sometimes called) seeks to eliminate policies that, while neutral on their face, disproportionately harm members of a particular protected class.[131] Under the disparate impact approach, proof of the defendant's state of mind (i.e., illicit motive) is not required.[132] The disparate impact model of discrimination was not sanctioned by Congress in the original Civil Rights Act of 1964. Rather, it has been adopted piecemeal by administrative agencies[133] and federal courts[134] seeking to extend the law's reach.[135]

Although there is a broad consensus favoring the use of the disparate treatment model to eliminate purposeful discrimination in all arenas,[136] the use of the disparate impact model to curtail practices that are not intentionally discriminatory remains controversial[137] and is, therefore, limited in scope and reach.

---

131. *See Int'l Bhd. of Teamsters*, 431 U.S. at 335-36 n.15 (the disparate impact model of discrimination prohibits the use of "practices that are facially neutral in their treatment of different groups but that in fact fall more harshly on one group than another and cannot he justified by business necessity"); *see also* George Rutherglen, *Discrimination and Its Discontents*, 81 VA. L. REV. 117 (1995) (describing the disparate impact model of discrimination); Willhorn, *supra* note 18 (similar).

132. *Int'l Bhd. of Teamsters*, 431 U.S. at 335-36 n.15 (stating that in the law of disparate impact, "[p]roof of discriminatory motive . . . is not required")

133. *See, e.g.*, 34 C.F.R. § 100.3(b)(2) (2001) (applying the disparate impact model of discrimination to claims against recipients of federal funds under Title VI of the 1964 Civil Rights Act); Dept. of Housing and Urban Development, *Policy Statement on Discrimination in Lending*, 59 Fed. Reg. 18,269 (Apr. 14, 1994) (applying the disparate impact model to claims of housing and credit discrimination).

134. *See, e.g.*, Griggs v. Duke Power Co., 401 U.S. 424 (1971) (sanctioning the use of the disparate impact model in employment cases under Title VII of the 1964 Civil Rights Act).

135. In 1991, after a series of court decisions curbed the application of the disparate impact model in employment discrimination cases, Congress passed the Civil Rights Act of 1991, which amended Title VII to provide explicitly for lawsuits based on the disparate impact model. *See generally* ROGER CLEGG, DISPARATE IMPACT IN THE PRIVATE SECTOR: A THEORY GOING HAYWIRE 3 (2002) (explaining the origins of the disparate impact model).

136. *See* Owen M. Fiss, *A Theory of Fair Employment Laws*, 38 U. CHI. L. REV. 235, 244 (1971) ("The ethical basis for a law seeking equal treatment seems clear: under the view that race is an is an arbitrary criterion, *unequal* treatment . . . is a form of *unfair* treatment, a particularized wrong" (emphasis added)); David A. Strauss, *The Law and Economics of Employmcnt Discrimination: The Case for Numeric Standards*, 79 GEO. L.J. 1619, 1623 (1991) ("[A]ll of the various prohibitions against discrimination in American law—those that apply to prohibitions against discrimination by government or by recipients of federal funds, in public accommodations or in housing—forbid treating members of minority groups differently from otherwise identical nonminorities."); Willborn, *supra* note 18, at 802-03 (noting that intentional discrimination is widely regarded as immoral, as well as inefficient). *But see generally* RICHARD EPSTEIN, FORBIDDEN GROUNDS (1992) (arguing against laws which prohibit private intentional discrimination on free association grounds).

137. *See, e.g.*, Roger Clegg, *The Bad Law of "Disparate Impact"*, PUB. INT., Winter 2000, at 79 (arguing against the expansion of disparate impact law generally); Michael Evan Gold, Griggs' *Folly: An Essay on the Theory, Problems and Origin of the Adverse Impact Definition of Employment Discrimination and a Recommendation for Reform*, 7 INDUS. REL. L.J. 429 (1985)

For example, rather than apply a "pure" disparate impact model, in which all policies with adverse racial or ethnic effects would be viewed as per se illegal, courts, legislatures, and administrative agencies invoking disparate impact analysis have generally limited its scope by creating an affirmative defense to liability.[138] Under such a defense, defendants are provided an opportunity to justify their use of a policy with disproportionate demographic outcomes, and thereby avoid liability, as long as such policies are deemed "necessary."[139]

Moreover, unlike the disparate treatment model, the disparate impact model is not automatically applied to *all* areas of discrimination law.[140] In 1976, the U.S. Supreme Court held in *Washington v. Davis*[141] that the disparate impact model cannot be applied to claims of state-sponsored discrimination under the Equal Protection Clause of the Fourteenth Amendment.[142] In so holding, the *Davis* Court emphasized that the disparate impact model cannot provide a universal standard of discrimination,[143] but implied that courts and legislatures are free to adopt the model by statute on a

---

(arguing against the use of the disparate impact approach in the employment setting and in favor of an expanded intent standard under which statistical evidence would be afforded more weight); *cf.* Fiss, *supra* note 136, at 244 (conceding that the ethical basis underlying a model that requires equal achievement or equal outcomes remains uncertain).

138. *See* Michael Carvin, *Disparate Impact Claims Under the New Title VII*, 68 NOTRE DAME L. REV. 1153, 1154 (1993) (noting that a "pure" disparate impact standard would prohibit any practice that did not achieve absolute proportional representation and would require employers to engage in hiring by quota); Willborn, *supra* note 18, at 802-03 (explaining that the use of a "pure" disparate impact model would conflict with contemporary notions of fairness and would be uneconomical in practice).

139. *See, e.g.*, 42 U.S.C. § 2000(e)(2)(k) (2000) (prohibiting employment practices with a disparate impact *unless* employer can demonstrate that the practice is job-related and consistent with business necessity).

140. *Compare* Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971) (holding that disparate impact analysis is appropriate in the employment context under Title VII), § 2000(e)(2)(k) (codifying *Griggs* and defining the burdens of proof in a disparate impact employment discrimination case), *and* Policy Statement on Discrimination in Lending, 59 Fed. Reg. 18,269 (Apr. 14, 1994) (applying disparate impact standards to claims of discrimination under the Equal Credit Opportunity Act and the Fair Housing Act), *with* Washington v. Davis, 426 U.S. 229, 242 (1976) (holding that discriminatory intent, not effects, is the touchstone for proving a violation of the Equal Protection Clause), Mullin v. Raytheon, 164 F.3d 696, 701-02 (1st Cir.) (holding that *Griggs* is inapposite in the context of age discrimination claims and that proof of discriminatory intent is a prerequisite to liability under the ADEA), *cert. denied*, 528 U.S. 811 (1999), and AFSCME v. Washington, 770 F.2d 1401, 1406 (9th Cir. 1985) (rejecting the use of disparate impact theory to prove pay discrimination in claims of comparable worth).

141. 426 U.S. at 246-48 (rejecting a claim brought by black applicants to the police academy that a general literacy test had an unlawful disparate impact on African-American job seekers).

142. *Id.* (holding that the court of appeals erred in applying Title VII's disparate impact standards to a constitutional claim).

143. *See id.*; *see also* Randall L. Kennedy, *The State, the Criminal Law, and Racial Discrimination: A Comment*, 107 HARV. L. REV. 1255, 1270-77 (1994) (explaining the rationale behind the Court's holding in *Washington v. Davis*).

case-by-case basis after considering the theory's value in a particular setting.[144] Indeed, five years earlier, in *Griggs v. Duke Power Co.*, the Supreme Court approved the use of the disparate impact model in cases of employment discrimination under Title VII of the 1964 Civil Rights Act.[145]

B. Theoretical Justifications for the Disparate Impact Model

For years commentators have sought to develop a coherent justification for the disparate impact approach to claims of discrimination.[146] These theories have generally been said to fall into three broad categories: weak justifications, strong justifications, and those that are in between.[147] So-called "weak" justifications for the disparate impact model focus on its ability to ferret out covert, yet intentional, discrimination.[148] In an influential article published in the 1980s, Professor George Rutherglen rationalized the disparate impact model, not as an altogether separate cause of action, but as a different method of proof under which statistical evidence is given substantial

---

144. 426 U.S. at 247-48; *cf.* Louis A. Jacobs, *A Constitutional Route to Discriminatory Impact Statutory Liability for State and Local Government Employees: All Roads Lead to Rome*, 41 OHIO ST. L.J. 301 (1980) (advocating a piecemeal statutory approach to imposing disparate impact liability). *But see* CLEGG, *supra* note 135, at 23 (arguing that Congress cannot act to enforce the Fourteenth Amendment, which bans only intentional discrimination, by prohibiting practices with an adverse racial impact). Clegg notes, however, that this argument applies only to statutes enacted under Section 5 of the Fourteenth Amendment and not to those statutes, such as Title VI, enacted under Congress's spending authority or those statutes aimed at the private sector which are enacted pursuant to the Commerce Clause. *Id.*

145. The plaintiffs in *Griggs* argued that the employer's hiring criteria (a diploma requirement and passage of a standardized intelligence test) violated Title VII because of its discriminatory impact on African-American workers and job applicants, and the Supreme Court agreed. 401 U.S. at 431. Writing for the Court, Chief Justice Burger explained that when a hiring criterion has a disproportionately negative impact on members of a particular racial or ethnic group, the employer cannot continue to use the criterion *unless* it is "job-related" and consistent with "business necessity." *Id.* at 431, 436 ("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."). Today, the disparate impact model of employment discrimination has been codified in Title VII. But neither Title VI nor Title IX of the Civil Rights Act have been amended to incorporate this approach.

146. *See* Willborn, *supra* note 18, at 826 ("[T]here is no agreement on the [disparate impact model's] underlying theory."); *see also* Ramona L. Paezold & Steven L. Willborn, *Deconstructing Disparate Impact: A View of the Model Through New Lenses*, 74 N.C. L. REV. 325, 328 n.6 (1996) (describing various theories propounded by commentators to justify the use of the disparate impact model).

147. *See* Rutherglen, *supra* note 131, at 138-39 (describing the various justifications for disparate impact as ranging from weak to strong).

148. *See* George Rutherglen, *Disparate Impact Under Title VII: An Objective Theory of Discrimination*, 73 VA. L. REV. 1297, 1299 (1987) (stating that disparate impact theory provides objective evidence of pretext and is thus an appropriate method for smoking out intentional discrimination, which is difficult to prove).

weight, and plaintiffs' burden of proof is reduced in order to help them prove pretextual discrimination.[149]

According to this theory, the Supreme Court's approval in *Griggs* of the disparate impact approach to claims of discrimination under Title VII was to be interpreted in the context of the racial segregation in which the case arose. *Griggs* involved a challenge to an employer's requirement that all new employees or transfers to certain divisions of the company have a high school diploma or have obtained a certain score on a standardized general intelligence test. Prior to 1965, when Title VII became effective, the employer in *Griggs* openly discriminated by segregating its black employees into the division of the company with the lowest paying jobs.[150] On the day on which Title VII took effect, the employer eliminated race-based classifications from its hiring process, but instituted the hiring criteria at issue in the case.[151] The effect of the new requirements was to exclude a higher percentage of blacks than whites from certain highly coveted jobs. Because these suspicious circumstances suggested—but provided no direct evidence of—an intent to discriminate, some commentators, such as Rutherglen, argued that the *Griggs* Court approved the disparate impact model of employment discrimination as a means of smoking out pretext in the absence of any clear evidence of illicit motive.[152]

By contrast, other scholars have proffered "strong" justifications in support of disparate impact theory, focusing on the desire to achieve proportional representation or equal outcomes.[153]

---

149. *Id.* at 1309; *see also* Rutherglen, *supra* note 131, at 138 ("In its weakest form, [disparate impact] theory imposes only a light burden of justification upon the employer; it only extends the central prohibitions against discrimination and segregation to root out hidden discrimination.").

150. 401 U.S. at 426-27.

151. *Id.* at 428.

152. *See* Rutherglen, *supra* note 148, at 1309. Although this justification found some support in the language of the Court's opinion, *see Griggs*, 401 U.S. at 426 (framing the question as whether an employer may properly use job requirements that disproportionately screen out African-Americans for positions previously "filled only by white employees as part of a longstanding practice of giving preference to whites"), since Congress amended Title VII in 1991 to outline explicitly the burdens of proof in disparate impact cases, it is clear that this justification is no longer valid with respect to claims of discrimination in employment. *See infra* notes 162-75.

153. *See, e.g.*, Strauss, *supra* note 136, at 1627 (arguing that the disparate impact model is best justified as a means of achieving proportional representation); *see also* Martha Chamallas, *Evolving Conceptions of Equality Under Title VII: Disparate Impact Theory and the Demise of the Bottom Line Principle*, 31 UCLA L. REV. 305, 365 (1983) (arguing that the disparate impact theory was designed to produce equal outcomes and that it is "fundamentally results-oriented and grounded on notions of group status"); Perry, *supra* note 18 (supporting the use of disparate impact doctrine as a means of forcing the American workplace to accommodate otherwise qualified individuals with different life perspectives).

These justifications are entirely divorced from intent and are generally concerned more with notions of compensatory justice than with the principle of equal opportunity. For example, Professor David Strauss has argued that the disparate impact model bests serves the goals of Title VII by requiring proportional representation and equal group outcomes.[154]

In between these two contrasting justifications of the disparate impact model are several rationales that focus on the disparate impact doctrine's ability to eliminate certain policies or practices that unfairly harm minorities. These theories adopt the view that the disparate impact doctrine provides substantive rights independent of the right to be free from intentional discrimination, but limit the doctrine's reach to only those practices that are deemed to be "unfair." Thus, Professor Michael Perry has written that the disparate impact approach can be justified as a means of eliminating the vestiges of earlier intentional race discrimination.[155] According to Perry, in the years following the dismantlement of *de jure* segregation in American schools, an employer's use of a diploma requirement or general intelligence exam to select employees unfairly penalized African-American job applicants whose academic credentials and scores on academic tests may have been artificially depressed as a direct consequence of systemic intentional discrimination in America's primary and secondary schools.[156]

Professor Steven Willborn, on the other hand, justifies the disparate impact approach in the employment context as a means of eliminating "statistical discrimination."[157] According to Willborn, employers often lack the ability to evaluate a prospective employee's potential for productivity at a reasonably low cost. Employers therefore substitute readily available proxies such as race, sex, marital status, past experience, recommendations, or test scores for more precise, but more costly, information regarding productivity.[158] According to this rationale, fair employment laws are necessary to prevent employers from using factors such as race and sex as proxies for productivity. The disparate impact component of employment

154. *See, e.g.*, Strauss, *supra* note 136, at 1627.

155. Michael J. Perry, *The Disproportionate Impact Theory of Employment Discrimination*, 125 U. PA. L. REV. 540, 577 (1977). This view also finds support in the language of *Griggs*. *Griggs*, 401 U.S. at 432 ("Practices, procedures, or tests neutral on their face, and *even neutral in terms of intent*, cannot be maintained if they operate to 'freeze' the status quo of prior discriminatory employment practices.").

156. Perry, *supra* note 155, at 577.

157. Willborn, *supra* note 18, at 821.

158. *Id.* at 821-23 (rationalizing disparate impact as a mechanism to prohibit statistical discrimination that is only weakly linked to productivity and that adversely affects minorities).

discrimination laws operates as a mechanism to prevent employers from using proxies which, while facially neutral, are closely correlated with race and not sufficiently correlated with productivity or other economic baselines.[159]

In a similar vein, Professor Owen Fiss has argued that the disparate impact model is sustainable in the employment setting where the criterion that causes the disproportionate result is the "functional equivalent" of race.[160] In order for reliance on a particular criterion to be the functional equivalent of race, Fiss contends that: (1) the criterion must be a poor predictor of productivity, *and* (2) there must be a lack of individual control over the outcome.[161] Fiss provides the example of a company nepotism policy. Assume that a company with an all-white workforce announces that in the future it will hire only relatives of current employees. Although the policy is facially neutral, it adversely affects nonwhite workers, through no fault of their own. The practice of hiring only relatives of current workers is unrelated to productivity and cannot be justified by reference to the bottom line. At the same time, the use of such a policy is unfair to minority applicants because the ability to satisfy the hiring criterion is completely outside of their control. According to Fiss, then, the policy is the "functional equivalent" of race. In such a context, the disparate impact model would be an appropriate means of resolving a claim of discrimination based on the policy.[162] By contrast, a requirement that all candidates for secretarial positions pass a typing test would *not* be the "functional equivalent" of a race-based test because the test would measure a trait clearly relevant to productivity and not outside the control of the applicants.[163]

---

159. *Id.*

160. Fiss, *supra* note 136, at 299.

161. *Id.* at 303. According to Fiss, both elements are necessary before the disparate impact doctrine may appropriatly be applied.

162. *Id.* Consider also an example from the law of sex discrimination. Suppose an employer adopts a policy prohibiting employees who work out of doors from coming inside to use the bathrooms. This factual scenario was presented in the case of *Lynch v. Freeman*, 817 F.2d 380 (6th Cir. 1987). In *Lynch*, the Sixth Circuit found that the policy had an adverse impact on female workers in a manner which was outside of their control. *Id.* at 387-88. Moreover, the policy could not be characterized as important to the employer's productivity or its bottom line, and was therefore held unlawful. *Id.*

163. Fiss, *supra* note 136, at 299; *see also* Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 987 (1988) (O'Connor, J.) ("[T]he necessary premise of the disparate impact approach is that some employment practices, adopted without a deliberately discriminatory motive, may in operation be functionally equivalent to intentional discrimination.").

C. Title VII as the Disparate Impact Prototype

In 1991, Congress codified the disparate impact approach to employment discrimination and established with particularity the burdens of proof that must be met in order to establish unlawful disparate impact under Title VII.[164] In passing the Civil Rights Act of 1991, Congress rejected the "weak" version of the disparate impact model, which provides a means of proving intentional discrimination, in favor of a version that looks primarily at a policy's effects.[165]

Title VII now explicitly provides that, once an employment discrimination plaintiff has established a prima facie case of disparate impact (by demonstrating that the challenged employment practice causes[166] a statistically significant[167] workforce disparity), the burden

---

164. 42 U.S.C. § 2000e(2)(k) (2000). The Civil Rights Act of 1991 is generally viewed as a legislative response to the Supreme Court's 1989 decision in the case of *Wards Cove v. Atonio*, 490 U.S. 642 (1989) ("*Wards Cove*"), in which the Court held that, when proffering a business reason for a challenged employment practice, the employer's burden is one of production, and not of persuasion. Specifically, in *Wards Cove*, the Court held that a plaintiff retains the burden of proving that the business justification proffered by the defendant is not valid. *Id.* at 658-59. The 1991 Act reverses this burden of proof, making proof of business necessity an affirmative defense. § 2000(e)(k).

Some commentators have argued that, even after passage of the 1991 Act, employers need not prove that a policy be linked to actual productivity, only that it be related to the job in some sense. (Thus, for example, workplace rules about punctuality, drug use, and proper workplace attire are all job-related even when they do not influence productivity). These commentators also contend that, even after passage of the 1991 Act, the employer need not prove that the challenged practice is "essential" or "indispensable," only that it serves "legitimate employment goals." *See, e.g.*, Carvin, *supra* note 138, at 1157; C. Boyden Gray, *Disparate Impact: History and Consequences*, 54 LA. L. REV. 1487 (1994).

165. *See* Rutherglen, *supra* note 131, at 138 ("By rejecting *Wards Cove*, Congress presumably rejected the weakest form of the theory, but it did not necessarily embrace the strongest. Instead, it used equivocal language from both forms: an employer has the burden of proving that 'the challenged practice is job related for the position in question and consistent with business necessity.' ").

166. A plaintiff in a Title VII disparate impact case must identify a specific policy, practice, or standard as the *cause* of the different outcomes. A plaintiff cannot make out a prima facie disparate impact claim where the evidence shows that the same disparate results would have occurred even had the defendant not engaged in the challenged practice. *See Watson*, 487 U.S. at 994 (1988) ("[T]he plaintiff must . . . show that the practice in question has caused the exclusion of applicants for jobs or promotions because of their membership in a protected group.").

167. § 2000e-2(k); *see also Watson*, 487 U.S. at 995 (stating that in Title VII cases, plaintiffs must show that the practice in question "select[s] applicants for hire or promotion in a racial pattern *significantly* different from that of the pool of applicants" (emphasis added) (internal quotation marks omitted) (quoting Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1973))). In determining whether the complained of impact is statistically significant, courts generally apply the "eighty percent rule" (otherwise known as the "four-fifths rule") established by the EEOC. Under this rule, "[a] selection rate for any race . . . which is less than [eighty percent] of the rate for the group with the highest rate will generally be regarded . . . as evidence of adverse impact[.]" *See* Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (1990).

of proof shifts to the employer to justify its use of the challenged practice. The 1991 Act requires employers defending against disparate impact claims to prove that the challenged practice is "job related for the position in question and consistent with business necessity."[168] The 1991 Act establishes the employer's burden as an affirmative defense under which the defendant bears the burden of persuasion.[169] Should the employer meet this burden, plaintiffs may still prevail if they can demonstrate that an alternative employment practice would reduce the disproportionately adverse effects while also serving the employer's legitimate business interests.[170]

With respect to standardized employment tests, some courts interpret the statute's requirement that a defendant prove the practice at issue is "job related" and "consistent with business necessity" to require defendants to prove that the tests were properly *validated*[171] for the job or class of jobs in question.[172] Ordinarily, this means that defendants must prove that the test is "predictive of or significantly correlated with important elements of work behavior that comprise or are relevant to the job or jobs for which candidates are being evaluated."[173] Specifically, some courts have required defendants to: (1) specify the trait in question that the test seeks to measure; (2) prove that the particular trait in question is an important element of work behavior; and (3) prove by professionally accepted standards that the test is predictive or significantly correlated with the element of work behavior identified in step two.[174]

Even where a defendant succeeds in proving the validity of the test, courts will nevertheless undertake to review a defendant's choice of cutoff score (i.e., the minimum score required to qualify for consideration) using the same framework of analysis. Thus, where a defendant establishes a passing rate that tends to eliminate members

---

168. § 2000e-2(k)(1)(A)(ii).

169. *Id.* § 2000e-2(k) (reversing the burdens of proof adopted by the Supreme Court in *Wards Cove Packing*).

170. *Id.*

171. Validation is itself an extremely pliable term, easily manipulated to serve various political goals. *See infra* Part V.B.1.

172. *See, e.g.*, Ass'n of Mexican Am. Educators v. California, 231 F.3d 572, 585 (9th Cir. 2000). *But see* Carvin, *supra* note 138, at 1157 (arguing that the amended statute does not require that companies professionally validate employment tests in order to withstand disparate impact scrutiny).

173. Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(C) (1990); *see Ass'n of Mexican Am. Educators,* 231 F.3d at 585 (citing Albemarle Paper Co. v. Moody, 422 U.S. 405, 431 (1975)).

174. *Ass'n of Mexican Am. Educators*, 231 F.3d at 585 (citing Craig v. County of L.A., 626 F.2d 659, 662 (9th Cir. 1980)).

of protected groups, courts will consider whether the cutoff score is itself consistent with business necessity.[175]

For better or worse, the theory that the disparate impact doctrine should guarantee substantive group rights apart from the right to be free from purposeful discrimination is now firmly entrenched in the law of employment discrimination.[176] Indeed, the employment law version of the doctrine has become the template from which courts, legislatures, and administrative agencies craft disparate impact standards for other bodies of law. Although advocates of a blanket effects test continue to press courts and administrative agencies to import Title VII's disparate impact regime reflexively to other settings, including education,[177] the lack of consensus as to the theoretical basis of disparate impact law requires that each attempt to transfer the model to a body of law outside of the employment arena be judged separately and on its own merits.[178]

---

175. *See, e.g.*, Guardians Ass'n v. Civil Serv. Comm'n, 630 F.2d 79, 105 (2d Cir. 1980) ("[A] cutoff score unrelated to job performance may well lead to the rejection of applicants who were fully capable of performing the job. When a cutoff score unrelated to job performance produces disparate racial results, Title VII is violated."), *cert. denicd*, 452 U.S. 940 (1981); Richardson v. Lamar County Bd. of Educ., 729 F. Supp. 806, 882 (M.D. Ala. 1989) (holding that where cutoff scores on teacher certification test were found to bear no "rational relationship" to minimum teacher competence, the test could not withstand disparate impact scrutiny under Title VII), *aff'd*, 935 F.2d 1240 (11th Cir. 1991); Lanning v. SEPTA, 181 F.3d 478, 488 (3d Cir. 1999), *cert. dcnied*, 528 U.S. 1131 (2000) (ruling that facially neutral requirement that transit officers be able to run twelve-minute miles was not properly validated and that defendant must prove that cutoff score measures minimal qualifications necessary for successful job performance).

176. *Sec* Paezold & Willborn, *supra* note 146, at 326-27 (noting the centrality of disparate impact in employment law and explaining that it is in the employment arena where the doctrine has been most fully developed); *sce also* Rutherglen, *supra* note 131, at 126 (noting that, in the employment setting, "the principle [of disparate impact] develop[ed] from an evidentiary rule into a redefinition of the concept of discrimination").

177. *Sec* 34 C.F.R. § 100.3(b)(2) (1990) (prohibiting policies of fund recipients which have the "effect of subjecting individuals to discrimination"); *see also* Paezold & Willborn, *supra* note 146, at 328 n.2 and accompanying text (noting that advocates of the effects test have succeeded in convincing courts to adopt various forms of the disparate impact approach in housing discrimination, lending discrimination, and voting rights cases).

178. *Sce, e.g.*, Douglas C. Herbert & Lani Schweitzer, *A Pragmatic Argument Against Applying Disparate Impact Doctrine in Age Discrimination Cases*, 37 S. TEX. L. REV. 625, 632 (1996) (arguing against the mechanical transfer of Title VII's disparate impact regime to the Age Discrimination in Employment Act); Earl M. Maltz, *The Expansion of the Role of the Effects Test in Antidiscrimination Law: A Critical Analysis*, 59 NEB. L. REV. 345, 357-62 (1980) (arguing against the wholesale transfer of disparate impact theory from employment law to laws regarding housing and lending discrimination); Perry, *supra* note 155, at 563-86 (describing disparate impact doctrine as "contextually limited" and arguing in favor of applying the doctrine to constitutional claims relating to public employment, education, and land use planning, but not to constitutional claims regarding jury selection or legislative districting). *But see* Peter E. Mahoney, *The End(s) of Disparate Impact: Doctrinal Reconstruction, Fair Housing and Lending Law, and the Antidiscrimination Principle*, 47 EMORY L.J. 409 (1998) (endorsing the use of Title VII's disparate impact standards to prove discrimination in lending).

IV. CLAIMS OF DISCRIMINATION IN EDUCATIONAL TESTING

How has the law actually responded to claims that standardized educational tests are discriminatory? And what models have courts relied upon in determining whether a testing policy violates legal prohibitions on discrimination in education? Part IV.A explains the statutory framework of Title VI—the law that prohibits discrimination in education—and explains how claims of discrimination in educational testing are resolved by the use of the disparate treatment model. Part IV.B then considers the use of the disparate impact model under Title VI, explains how the model was applied to educational testing prior to the Supreme Court's landmark decision in *Alexander v. Sandoval*,[179] and considers the relevance of the disparate impact doctrine in the post-*Sandoval* era.

A. Title VI's Prohibition on Purposeful Discrimination in Education

Title VI of the Civil Rights Act of 1964 prohibits discrimination on the basis of race, color, or national origin by recipients of federal funds.[180] Specifically, Title VI provides:

> No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.[181]

As a result of federal financial support, most schools are covered by Title VI and subject to its prescriptions.[182]

What does Title VI's prohibition on discrimination mean in the context of high-stakes educational tests? Clearly, the plain language of Title VI prohibits the use of standardized tests *for the purpose of* screening out or segregating students on the basis of race, color, or national origin. For example, a state board of education or school board that adopted an exit exam requirement with the goal of denying diplomas to minority students would violate Title VI. Likewise, an

---

179. 525 U.S. 275 (2001).

180. Although Title VI is widely regarded as an education statute, it applies to all recipients of federal funds, not just schools. Accordingly, the legal standards for adjudicating claims of discrimination under Title VI apply broadly and can cover a wide range of entities including, inter alia, charitable institutions, job training programs, state departments of environmental protection, and state departments of motor vehicles.

181. 42 U.S.C. § 2000d (2000).

182. Title VI was enacted pursuant to Congress's power to attach conditions to grants of federal money under the Spending Clause of the U.S. Constitution. *See* Guardians Ass'n v. Civil Serv. Comm'n, 463 U.S. 582, 599 (1983). Title VI thus imposes contractual obligations on recipients of federal funds—in consideration of federal financing, recipients agree not to discriminate. *Id.* (citing 110 CONG. REC. 6546 (1964) (statement of Sen. Humphrey)).

exam adopted for purposes of "tracking" students into racially identifiable classrooms would violate Title VI.

Title VI also prohibits the inconsistent use of a test or test scores with respect to members of different racial or ethnic groups. A school that required only black students to pass an exam as a condition of graduating would clearly violate Title VI. Similarly, a school that required all students to pass an exam, but that required Latino students to obtain a higher passing score than that required of other students would violate Title VI,[183] as would a school that granted diplomas to some white students with scores just below the passing mark but not to Latinos with identical scores.[184] Such conduct would constitute unlawful disparate treatment—in other words, *intentional* discrimination.

Individuals who believe that they have been the victims of intentional discrimination in violation of Title VI can file an administrative complaint with the Office for Civil Rights at the U.S. Department of Education[185] or bring suit in federal court for injunctive

---

183. Interestingly, many commentators read Title VI's prohibition of disparate treatment on the basis of race to apply only where the educational institution requires *more* of minority students than of white students. Thus, while a school may not require minority students to obtain *higher* test scores than whites in order to obtain some benefit or privilege, these commentators endorse the practice of affirmative action, whereby colleges and universities routinely admit minority students with test scores that are *lower* than those which would be considered acceptable if presented by white applicants for admission. *See generally* Grutter v. Bollinger, 137 F. Supp. 2d 821 (E.D. Mich. 2001) (documenting the disparity between the scores of white students admitted to the University of Michigan Law School and those of minority admittees), *rev'd*, 288 F.3d 732 (2002); WILLIAM G. BOWEN & DEREK BOK, THE SHAPE OF THE RIVER: LONG-TERM CONSEQUENCES OF CONSIDERING RACE IN COLLEGE AND UNIVERSITY ADMISSIONS (1998) (defending the practice of affirmative action). *But see* Stephan Thernstrom & Abigail Thernstrom, *Reflections on the Shape of the River*, 46 UCLA L. REV. 1583 (1999) (criticizing this approach). Defenders of affirmative action find support in Title VI's regulations, which provide apparent legal authority for the proposition that that educational institutions may voluntarily implement affirmative action programs to "overcome the effects of prior discrimination" due to "race, color, or national origin." Even in the absence of past discrimination, the regulations provide that educational institutions may implement voluntary measures aimed at overcoming "the effects of conditions which resulted in limiting participation by persons of a particular race, color, or national origin." 34 C.F.R. § 100.3(6)(i)-(ii) (1990).

184. *See* NAT'L RESEARCH COUNCIL, *supra* note 15, at 52; *cf.* People Who Care v. Rockford Bd. of Educ., 851 F. Supp. 905, 958-1001 (N.D. Ill. 1994) (holding that unequal application of standards violates Title VI), *remedial order rev'd in part*, 111 F.3d 528 (7th Cir. 1997).

185. Title VI authorizes administrative agencies that dispense federal funds to conduct investigations and engage in fact finding. If the agency determines that there has been a statutory violation, it may seek to resolve the matter informally or can seek compliance by terminating the institution's federal financing after an administrative hearing. § 2000d-1.

> Compliance with any requirement adopted pursuant to this section may be effected (1) by the termination of or refusal to grant or to continue assistance under such program or activity to any recipient as to whom there has been an express finding on the record, after opportunity for hearing, of a failure to comply with such requirement, . . . or (2) by any other means authorized by law: *Provided, however*, that no such action

or monetary relief.[186] But how exactly does a Title VI plaintiff go about proving that an educational test was adopted for a discriminatory purpose? Generally, the question of whether a Title VI fund recipient has committed intentional discrimination is resolved by examining the totality of the circumstances.[187] As in the Fourteenth Amendment context,[188] evidence of a disproportionate adverse impact can support an inference of discriminatory intent where no legitimate nondiscriminatory reason exists for the practice or where other evidence indicates that the proffered reason is a pretext for discrimination. In examining claims that public policies violate the Equal Protection Clause of the Fourteenth Amendment, the Supreme Court has identified criteria to assist courts in determining whether a public policy was adopted with discriminatory intent.[189] As applied to the educational testing context, relevant factors include, but are not limited to: (1) whether the test produces a disproportionate, adverse impact on the group that alleges discrimination;[190] (2) whether the

---

> shall be taken until the department or agency concerned has advised the appropriate person or persons of the failure to comply with the requirement and has determined that compliance cannot be secured by voluntary means. In the case of any action terminating, or refusing to grant or continue, assistance because of failure to comply with a requirement imposed pursuant to this section, the head of the Federal department or agency shall file with the committees of the House and Senate having legislative jurisdiction over the program or activity involved a full written report of the circumstances and the grounds for such action. No such action shall become effective until thirty days have elapsed after the filing of such report.

Id. Fund recipients may seek judicial review of agency actions. *Id.* § 2000d-2.

186. Private individuals may sue to enforce section 601's prohibition on intentional discrimination and may obtain both injunctive relief and monetary damages. *See* Alexander v. Sandoval, 523 U.S. 275, 280 (2001) (citing Franklin v. Gwinnett County Pub. Schools, 503 U.S 60, 72, 78 (1992)).

187. Vill. of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 266 (1977) (holding that each claim of discrimination "demands a sensitive inquiry into such circumstantial and direct evidence of intent as may be available").

188. The Equal Protection Clause of the Fourteenth Amendment prohibits intentional discrimination by state actors, including public school officials. U.S. CONST. amend. XIV, § 1 ("No state shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.").

Title VI prohibits intentional discrimination by federally funded entities, including most schools, and has been interpreted to track the contours of the Equal Protection Clause. *See Sandoval,* 523 U.S. at 275; United States v. Fordice, 505 U.S. 717, 732 n.7 (1992). Accordingly, claims of disparate treatment by educational institutions are often brought under both the Equal Protection Clause and Title VI, but are analyzed only under the constitutional standard.

189. *Vill. of Arlington Heights,* 429 U.S. at 266.

190. Disparate impact may sometimes provide "an important starting point" of analysis in claims of disparate treatment, *see id.*, although courts must take care to analyze all the objective evidence carefully so that they do not "reduce the 'discriminatory purpose' requirement to a disparate impact test by another name." Columbus Bd. of Educ. v. Penick, 443 U.S. 449, 509-10 (1979) (Rehnquist, J., dissenting).

test's disproportionate impact was reasonably foreseeable or actually foreseen by policymakers,[191] and, if so, whether defendants have attempted to reduce the test's adverse impact by providing remediation and additional opportunities to retake the exam; (3) whether adoption or administration of the test can be explained on grounds other than an intent to discriminate (i.e., whether there exist legitimate pedagogical reasons for the challenged policy); (4) whether the historical background of the decision to adopt the test supports a claim of intentional discrimination (e.g., was the policy adopted in response to a desegregation order or was it adopted as part of a comprehensive reform plan intended to equalize educational opportunities, as in Texas and Massachusetts?); (5) whether the sequence of events leading up to adoption of the policy appears suspicious; and (6) whether there is direct evidence of intent to discriminate, such as statements[192] of discriminatory animus.[193]

Thus, where there is direct evidence of discriminatory animus or where the circumstances support an inference that the challenged policy is a pretext for discrimination, courts have found intentional discrimination. For example, in *People Who Care v. Rockford Board of Education*,[194] a federal district court in Illinois held that a school district's use of standardized testing to "track" students into classes for different ability levels was, in fact, a pretextual means by which to segregate children on the basis of race. In that case, the district court found that test results were used inconsistently to support essentially subjective and racially motivated decisions,[195] and that the tracking standards were developed in conjunction with a reorganization plan

191. The foreseeability of the test's consequences may constitute one of several kinds of proof of intentional discrimination, but cannot provide the sole touchstone of intent. *Columbus Bd. of Educ.*, at 509-10 (Rehnquist, J., dissenting).

192. Isolated remarks by nondecisionmakers should not suffice to support such a claim. To the contrary, the record evidence must provide an inference of system-wide bias or intent to harm a particular racial or ethnic subgroup. *Cf.* Speen v. Crown Clothing Corp., 102 F.3d 625 (1st Cir. 1996) (holding that in claims of employment discrimination, isolated remarks are insufficient, standing alone, to prove discriminatory intent).

193. *Vill. of Arlington Heights*, 429 U.S. at 268; *see also* NAT'L RESEARCH COUNCIL, *supra* note 15, at 53 n.2 (applying the *Arlington Heights* factors to educational testing).

194. 851 F. Supp. 905 (N.D. Ill. 1994), *remedial order rev'd in part*, 111 F.3d 528 (7th Cir. 1997).

195. Specifically, the court found that black students whose test scores qualified them for more than one track were more likely to be placed in lower track classes than white students whose test scores qualified them for more than one track. The court also found that "high track" classes included a number of exceptionally low-scoring white students but no black students with similar low scores. *Id.* at 914.

that the court found was an attempt to resegregate the district's elementary schools.[196]

In finding intentional discrimination in the Fourteenth Amendment context, courts have placed great weight on a history of segregation or discrimination by the district implementing the tests, and have been willing to infer discriminatory purpose in the educational context where the challenged practice is found to carry forward or preserve the effects of prior illegal discrimination.[197] *Anderson v. Banks* is illustrative.[198] In *Anderson*, a federal district court in Georgia enjoined school officials from implementing, during the 1978 school year, a diploma sanction for failure to pass a ninth grade skills test. In finding that the school district's action violated the Equal Protection Clause of the Fourteenth Amendment, the court noted that the school district had strenuously resisted integration until the 1970-1971 school year (and, even then, had maintained a segregated transportation system).[199] Moreover, the year that the school district dismantled its dual system, it instituted a tracking system that quickly created racially identifiable classrooms.[200] In light of these facts, the district court held that, although the diploma sanction was not per se illegal, the diploma sanction could not be constitutionally imposed upon students who had attended school during *de jure* segregation.[201] The court therefore enjoined the school district from making passage of the test a graduation requirement until 1983, the first year in which graduating students would not have been exposed to the dual system.[202]

Another notable case, *Debra P. v. Turlington*,[203] involved a court challenge to Florida's "functional literacy examination," the

196. *Id.* at 958-1001 (holding that unequal application of tracking standards constitutes intentional discrimination).

197. *See* NAT'L RESEARCH COUNCIL, *supra* note 15, at 52.

198. 520 F. Supp. 472 (S.D. Ga. 1981), *rev'd in part on reh'g*, 540 F. Supp. 761 (S.D. Ga. 1982), *appeal dismissed sub. nom.* Johnson v. Sikes, 730 F.2d 644 (11th Cir. 1984).

199. *Anderson*, 520 F. Supp. at 479-80.

200. *Id.* at 480-81.

201. *Id.* at 490-91.

202. *Id.* at 503; *see also* McNeal v. Tate County Sch. Dist., 508 F.2d 1017 (5th Cir. 1975) (holding that academic groupings in previously segregated school system are presumptively invalid unless the district can prove either that its assignment method was not based on the "present results" of prior *de jure* segregation or that the assignment system would remedy such results through better educational opportunities). *But compare* Ga. State Conference of Branches of NAACP v. Georgia, 570 F. Supp. 314 (S.D. Ga. 1983) (holding that a school grouping system that disproportionately placed black students in lower groups did not perpetuate prior discrimination because none of the students involved had ever attended a segregated school).

203. 474 F. Supp. 244 (M.D. Fla. 1979), *aff'd in relevant part*, 644 F.2d 397 (5th Cir. 1981), *on remand*, 564 F. Supp. 177 (M.D. Fla. 1983), *aff'd*, 730 F.2d 1405 (11th Cir. 1984).

passage of which was required in order to graduate from high school. On the test's first administration, seventy-eight percent of African-American students failed, as compared to twenty-five percent of white students.[204] The district court held that, although the exam was not "biased" against minorities, the timing of the policy's implementation was suspicious. Because requiring students educated under *de jure* segregation to pass the exam as a condition of graduation perpetuated the effects of prior intentional discrimination, the court held that the testing requirement violated the constitutional guarantee of equal protection.[205] Significantly, however, the court held that the school could constitutionally implement the exit exam requirement after students educated during legalized segregation had left the school system.[206]

In finding equal protection violations, both the *Anderson* and the *Debra P.* courts relied heavily upon lack of academic opportunities within school districts still under desegregation orders. In the absence of a history of intentional discrimination, however, claims that the use of educational tests constitutes purposeful discrimination have rarely been successful. In *Parents in Action on Special Education ("PASE") v. Hannon*, for example, a federal district court in Chicago held that the use of tests to place students in special education classes did not violate Title VI's prohibition against intentional discrimination because: (1) the tests used by the district were not "biased" against minority test-takers; (2) the erroneous placement of black children in special education classes occurred infrequently and for reasons other than racial animus or bias; and (3) the district implemented the testing and special education program with the goal of helping students.[207]

Likewise, in *GI Forum v. Texas Education Agency*, the recent challenge to Texas's high-stakes testing regime, plaintiffs argued that the state had *purposefully* discriminated against minority test-takers when it set the test's pass rate at seventy percent, knowing that a

204. *Id.* at 248. Because of these significant disparities, the court closely analyzed the accuracy of the exam and concluded that the test conformed to scientific concepts of test validity.

205. *Id.* at 255. The *Debra P.* court also held that the implementation of the test and the inadequacy of the notice provided prior to invocation of the diploma sanction violated the Due Process Clause of the Fourteenth Amendment. *Id.* On appeal, the U.S. Court of Appeals for the Fifth Circuit affirmed the district court's holding with respect to Title VI and the Equal Protection Clause, hut remanded the due process claim, ordering the district court to consider whether defendants had proved "curricular validity"—in other words, whether the state could demonstrate that the test "covered things actually taught." Dehra P. v. Turlington, 644 F.2d 397, 405 (5th Cir. 1981).

206. *Debra P.*, 474 F. Supp. at 269.

207. 506 F. Supp. 831, 875 (N.D. Ill. 1980).

substantial number of black and Hispanic students would fail an exam with this cutoff score. As circumstantial evidence of intent to discriminate, plaintiffs also cited procedural irregularities in the rush to adopt the testing regime.[208] Defendants responded that the testing regime was established to help equalize educational opportunities among students of all races and ethnicities. Far from acting with discriminatory purpose, defendants argued that they acted out of a desire to help poor and minority students obtain a meaningful education. The district court summarily dismissed plaintiffs' claim of intentional discrimination, holding that the plaintiffs had proffered insufficient evidence of intentional discrimination to warrant a trial.[209]

In one outlier case, *Larry P. v. Riles*, a federal district court in California held that California's use of an IQ test to track children into classes for the educable mentally retarded ("EMR") violated Title VI's ban on intentional discrimination.[210] In so concluding, the district court relied upon a number of factors, including: (1) the "profound" disproportionate impact on black students[211] that was not only foreseeable, but actually foreseen; (2) the timing of the decision to adopt the testing requirement in the midst of "controversy" and expressed legislative concern over the number of black students in EMR classes and so-called cultural bias[212] in IQ tests; and (3) the fact that, in the court's view, EMR classes were "dead end" classes that provided little remedial assistance for students. The district court's finding of intentional discrimination was appealed by only one of the many defendants in the case: the superintendent of schools, who was an African-American. Although the court of appeals upheld the lower

208. *See* Order Granting, In Part, and Denying, In Part, Defendants' Motion for Summary Judgment, GI Forum v. Texas Educ. Agency, CA' No. SA-97-CA-1278-EP (W.D. Tex. July 27, 1999) [hereinafter Order for Summary Judgment], *available at* http://www.nysd.uscourts/court-web/Default.html; *see also* Moran, *supra* note 31, at 123 (describing the court's unpublished opinion).

209. *See* Order for Summary Judgment, *supra* note 208. The court's resolution of plaintiffs' claims of disparate impact are discussed *infra* notes 240-45 and accompanying text.

210. 495 F. Supp. 926, 926 (N.D. Cal. 1979), *affd. in part and rev'd in part*, 793 F.2d 969, 984 (9th Cir. 1984).

211. Disproportionate adverse racial or ethnic results can contribute to the mix of evidence supporting a finding of intentional discrimination although it is insufficient standing alone to support such a finding. *See* Washington v. Davis, 426 U.S. 229, 242 (1976) ("Disproportionate impact is not irrelevant, but it is not the sole touchstone of invidious racial discrimination forbidden by the Constitution.").

212. Unlike the *PASE* court, the *Larry P.* court did not undertake to determine whether the tests were, in fact, "culturally biased." Rather, in absence of evidence to the contrary, the court assumed that the tests were "biased" and held, in essence, that where an educational institution that employs a high-stakes test fails to investigate rumors of test bias it is responsible for intentional discrimination. *Larry P.*, 495 F. Supp. at 947, 956-60. The legal relevance of so-called test bias is discussed *infra* Part V.A.2.

court's finding of a Title VI violation on grounds of disparate impact,[213] the appellate court reversed the finding of intentional discrimination with respect to the superintendent, holding that the pervasiveness and foreseeability of a policy's adverse impact is insufficient, in and of itself, to establish discriminatory motive.[214]

The cases described above reveal a pattern in which claims of intentional discrimination in educational testing are unlikely to succeed in the absence of a history of state-sponsored discrimination or segregation. Moreover, although Title VI does not differentiate between tests used for tracking and tests used to make promotion and graduation decisions, a review of the case law reveals a pattern of stricter judicial scrutiny for tests used to "track" students by ability—perhaps because they can lead to racially identifiable classrooms that evoke memories of *de jure* segregation—than for tests used to determine who will graduate or be promoted. Because, in the absence of prior *de jure* segregation or other state-sponsored discrimination, lawsuits against high-stakes tests on the basis of intentional discrimination are unlikely to succeed, testing critics have turned to the disparate impact model of litigation as a means of eliminating such tests.

B. The Use of the Disparate Impact Doctrine to Challenge High-Stakes Educational Tests

What can be said of the claims that the disparate impact model should be applied to high-stakes educational assessments? Unlike Title VII, which now specifically provides for disparate impact lawsuits against employers,[215] Title VI provides no such cause of action. To the contrary, the U.S. Supreme Court has made clear that Title VI extends only as far as the Equal Protection Clause of the Fourteenth Amendment[216] and that, accordingly, the statute bans only *intentional* discrimination, not policies or procedures that create

213. *See* discussion *infra* notes 225-27 and accompanying text.

214. Larry P. v. Riles, 793 F.2d 969, 984 (9th Cir. 1984). Because the other defendants did not appeal the intentional discrimination portion of the district court's ruling, the district court's finding of intentional discrimination with respect to the other defendants remained intact.

215. *See supra* notes 164-78 and accompanying text.

216. United States v. Fordice, 505 U.S. 717, 732 n.7 (1992) ("Our cases make clear . . . that the reach of Title VI's protection extends no further than the Fourteenth Amendment."); Regents of the Univ. of Calif. v. Bakke, 463 U.S. 265, 287 (1978) ("Title VI must be held to proscribe only those racial classifications that would violate the Equal Protection Clause or the Fifth Amendment.").

adverse *results* for racial or ethnic communities.[217] Nevertheless, the U.S. Department of Education has expanded the reach of Title VI by regulation, forbidding federally funded entities from utilizing "criteria or methods of administration which have the *effect* of subjecting individuals to discrimination."[218] In other words, Title VI's implementing regulations, adopted in 1980 by the Carter Administration, prohibit institutions receiving federal assistance—which includes most American schools and colleges—from implementing policies with a disproportionate adverse impact on minorities.

In practice, this means that the federal government may withdraw funding from any educational institution that employs tests with unequal demographic outcomes. Until recently, it also meant that private parties could bring federal lawsuits against schools that employed such tests, provided that the cases were properly pled under Title VI's implementing regulations, rather than under the statute alone.[219]

---

217. The Supreme Court has long held that the Equal Protection Clause prohibits only intentional discrimination, and not adverse racial effects. *See* Vill. of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 270-72 (1977) (holding that official action will not be held unconstitutional simply because it results in a disproportionate racial impact and that invidious discriminatory purpose must have been a motivating factor); Washington v. Davis, 426 U.S. 229, 242 (1976) (holding that disproportionate impact cannot be "the sole touchstone" in claims of state-sponsored discrimination brought under the Equal Protection Clause of the Fourteenth Amendment). Accordingly, Title VI prohibits only intentional discrimination and not policies or practices which have adverse racial or ethnic consequences. *See* Alexander v. Sandoval, 532 U.S. 275, 280-81 (2001); *Fordice*, 505 U.S. at 732 n.27; *Bakke*, 463 U.S. at 287.

218. 34 C.F.R. § 100.3(2) (1980) (emphasis added).

219. In *Sandoval*, the Supreme Court held that private litigants may *not* sue to enforce Title VI's adverse impact regulations. 532 U.S. at 293. Previously, it had been widely assumed that private parties could bring disparate impact lawsuits for injunctive relief pursuant to Title VI's regulations, although not pursuant to the statute itself. *See* Guardians Ass'n v. Civil Serv. Comm'n, 463 U.S. 582, 645 (1983) ("*Guardians*") (White, J.) (Marshall, J., dissenting) (Brennan, Blackmun & Stevens, JJ., dissenting) (arguing that irrespective of whether Title VI itself provides a private right of action arising from a policy's disparate racial or ethnic impact, a plaintiff nevertheless can pursue a private right of action on grounds of disparate impact under Title VI's implementing regulations); *see also* Powell v. Ridge, 189 F.3d 387, 396 (3d Cir.) (adopting this view), *cert. denied*, 528 U.S. 1046 (1999); Elston v. Talladega Co. Bd. of Educ., 997 F.2d 1394, 1406 (11th Cir. 1993) (concerning private lawsuit regarding school reorganization plan brought, inter alia, pursuant to Title VI's regulations which proscribe actions having a racially disparate impact); Larry P., 793 F.2d at 982 (upholding injunctive relief for private plaintiffs who challenged California's use of an IQ test to track children into classes for the educable mentally retarded under, inter alia, Title VI's disparate impact regulations); GI Forum v. Tex. Educ. Agency, 87 F. Supp. 2d 667 (W.D. Tex. 2000) (concerning private lawsuit against Texas's graduation exam brought pursuant to Title VI's implementing regulations); Cureton v. NCAA, 37 F. Supp. 2d 687, 689, 697 (E.D. Pa.) (holding that, under *Guardians*, private plaintiffs may bring suit to enforce Title VI's disparate impact regulations), *rev'd on other grounds*, 198 F.3d 107 (3d Cir. 1999).

1. The Resource Guide: Importing Title VII's Disparate Impact Regime to Title VI

In 2000, the Clinton Administration issued a manual entitled *The Use of Tests as Part of High-Stakes Decision-making for Students: A Resource Guide for Educators and Policymakers.*[220] The *Resource Guide* grafted Title VII's disparate impact regime directly onto Title VI and effectively put schools on notice that relying on standardized test scores to make important educational decisions could place them in legal jeopardy.[221]

According to the *Resource Guide*, once the complaining party has made out a prima facie case of disparate impact (by identifying a specific policy or practice that caused a statistically significant disparity in the award of benefits or services to students based on race, national origin, or sex)[222], the burden of proof shifts to the educational institution to demonstrate that the testing practice is "educationally necessary."[223] As in the employment context, the *Resource Guide* states that even where the defendant establishes that its policy is "necessary," the party challenging the test has the opportunity to demonstrate that there exists "a feasible alternative practice that is effective in meeting the institution's goals and that would eliminate or reduce the adverse impact."[224] If the plaintiff can demonstrate that such an alternative exists, the *Resource Guide* contends that the institution is in violation of the statute.

---

220. *See* U.S. DEP'T OF EDUC., OFFICE FOR CIVIL RIGHTS, THE USE OF TESTS AS PART OF HIGH-STAKES DECISION-MAKING FOR STUDENTS: A RESOURCE GUIDE FOR EDUCATORS AND POLICYMAKERS (2000) [hereinafter RESOURCE GUIDE], *available at* http://www.ed.gov/offices/OCR/testing/TestingResource.pdf. Interestingly, the Clinton Administration did not speak with one voice on the question of accountability and high-stakes tests. In his 1997 State of the Union Address, President Clinton announced a "national crusade for education standards." *See* President Clinton, *Message to Congress on the State of the Union*, N.Y. TIMES, Feb. 5, 1997, at A20. Calling for an end to social promotion, Clinton stated, "no child should graduate from high school with a diploma he or she can't read" and proposed establishing voluntary national tests to assess the proficiency of fourth and eighth graders in reading and math. *Id.* Yet, despite Clinton's resounding endorsement of standards-based reform, critics of educational standards and tests found an ally in Clinton's own Department of Education, which published the *Resource Guide* encouraging lawsuits against school districts that employ high-stakes testing with disparate racial results. *See* RESOURCE GUIDE, *supra*; *see also* RAVITCH, *supra* note 2, at xx (noting that, even as President Clinton praised Arkansas's requirement that every student pass an eighth-grade exam before entering high school, his own Department of Education was threatening to launch a federal investigation into Ohio's use of a high school graduation exam).

221. RESOURCE GUIDE, *supra* note 220, *passim*; *see* Clegg, *supra* note 137, at 1.

222. RESOURCE GUIDE, *supra* note 220, at 37 (incorporating as well claims of disparate impact on the basis of sex under Title IX of the Education Amendments of 1972, 20 U.S.C. § 1681(a)).

223. *Id.* at 38.

224. *Id.* at 40.

2. Case Law

Although the Clinton Administration's *Resource Guide* claimed merely to summarize existing law, in truth very few cases have actually considered the question of whether the disparate impact model is appropriately applied to tests used to make promotion and graduation decisions.[225] Indeed, most of the cases cited by the *Resource Guide* (and most of the cases that have considered the application of Title VI's disparate impact regulation to educational testing) have involved the use of tests to make decisions regarding "tracking" or the eligibility of students for certain benefits or privileges.

In *Larry P.*, for example, a divided panel of the U.S. Court of Appeals for the Ninth Circuit upheld on grounds of unlawful disparate impact a district court's injunction barring the State of California from assigning students to classes for the educable mentally retarded on the basis of their performance on a standardized IQ test.[226] Affirming findings entered by the district court, the majority held that the tests were not "educationally necessary" because they were not separately validated for use with black children and because the EMR classes were educational "dead-ends" that provided no academic remediation.[227]

Dissenting from the majority's determinations on disparate impact, U.S. District Judge William B. Enright (sitting by designation) wrote separately to make clear his view that the district court had erred in assuming, based on the raw disparities in the racial composition of EMR classes, that black children were *erroneously* placed in EMR classes; the trial court, in Judge Enright's view, had failed to determine whether, and to what extent, school officials *actually erred* in making placement decisions. Judge Enright emphasized that, before educational tests can be found to discriminate, there must be, at a minimum, a finding that the tests

---

225. Most of the cases involving challenges to high-stakes exit exams have focused largely on the question of whether such exams, or the implementation of such exams, violate the Equal Protection Clause of the Fourteenth Amendment—in other words, whether the exams were enacted for a discriminatory *purpose*, whether the exams are themselves "biased" and thus discriminatory, or whether the exams preserve the effects of prior intentional discrimination.

226. 793 F.2d 969 (9th Cir. 1984). The district court had previously enjoined the policy on the grounds that it was intentionally discriminatory under the Fourteenth Amendment as well as on grounds that it had an unlawful adverse racial impact under Title VI's disparate impact regulations. *See* Larry P. v. Riles, 495 F. Supp. 987, 993 (N.D. Cal. 1979); *see also supra* note 212 (discussing case). The court of appeals, however, affirmed on regulatory grounds of disparate impact only. *Larry P.*, 793 F.2d at 970.

227. *Larry P.*, 793 F.2d at 973.

were invalid in that they produced improper placements in EMR classes.[228]

A second line of cases involves claims that tests used to determine an individual's eligibility for certain benefits or privileges result in an unlawful disparate impact. In *Groves v. Alabama,*[229] for example, a federal district court in Alabama held that the State's reliance on a minimum cutoff score on the ACT to determine eligibility to enter an undergraduate teacher training program had an unlawful disparate impact on African-American students in violation of Title VI's disparate impact regulation. After finding that plaintiffs had made out a prima facie case of disparate impact, the court rejected defendants' claim that the minimum required score was "educationally necessary" to ensure that students entering the teaching program possess a certain level of knowledge and academic attainment. The court held that defendants failed to demonstrate that the level of achievement designated by the cutoff score was required in order to ensure that program participants would turn out to be minimally competent teachers at the completion of the program.[230]

In *Sharif v. New York State Education Department,* a federal district court in Manhattan held that the State of New York's reliance on SAT scores to determine eligibility for state merit scholarships had an unlawful disparate impact on female students.[231] In particular, the court held that the SAT: (1) "underpredicts" female performance in college; and (2) has never been validated as a measure of *prior* academic achievement.[232] Accordingly, the district court held that the practice of using SAT scores to award merit scholarships did not bear *even* a "reasonable relationship," much less a "manifest relationship" (which the court held to be the requisite threshold), to the stated goal of rewarding high academic achievement.[233]

Similarly, in *Cureton v. NCAA,*[234] a case later reversed on other grounds, a federal district court in Pennsylvania held that the

228. *Id.* at 989.

229. 776 F. Supp. 1518 (M.D. Ala. 1991).

230. *Id.* at 1531-32. *But see* United States v. LULAC, 793 F.2d 636 (5th Cir. 1986) (rejecting a claim that a similar entrance exam was instituted with a discriminatory *purpose* and reversing a district court's order preliminarily enjoining, on Fourteenth Amendment grounds, the State of Texas's requirement that any student wishing to enroll in more than six hours of education courses first pass a preprofessional skills test).

231. 709 F. Supp. 345, 346 (S.D.N.Y. 1989) (Walker, J.). Because the disparate impact claim presented in *Sharif* dealt with claims of sex discrimination, it was decided under Title IX of the Education Amendments of 1972, 20 U.S.C. §§ 1631-1688 (2000). 709 F. Supp. at 360-61.

232. 709 F. Supp. at 353-54.

233. *Id.* at 362.

234. 37 F. Supp. 2d 687 (E.D. Pa.), *rev'd on other grounds,* 198 F.3d 107 (3d Cir. 1999).

National Collegiate Athletic Association's ("NCAA") "initial eligibility" rule had an unlawful disparate impact on minority students in violation of Title VI. The plaintiffs in *Cureton* sought to enjoin the NCAA from enforcing Proposition 16, which prevents students who have not obtained a minimum score on either of two standardized tests (the SAT or the ACT) from participating in intercollegiate athletics or receiving athletically related financial aid during their freshman year.[235] The NCAA defended Proposition 16 on the grounds that the rule is necessary in order to: (1) raise student-athlete graduation rates; and (2) close the gap between white-athlete and black-athlete graduation rates.[236]

In ruling for the plaintiffs, the district court first considered whether the NCAA's proffered justifications constituted substantial and legitimate educational goals and next considered whether the test score policy bore a "manifest relationship"[237] to achieving the stated goals. While the court approved of the first of the NCAA's stated goals—raising the graduation rate for all student-athletes—it rejected as illegitimate and insufficiently substantial the organization's goal of closing the black-athlete/white-athlete graduation gap.[238] Moreover, the *Cureton* court held that the NCAA failed to demonstrate a manifest relationship or nexus between Proposition 16 and the legitimate goal of raising graduation rates of all athletes because the SAT—while a valid predictor of freshman grades—had not been scientifically validated as an accurate predictor of who will graduate.[239]

3. *GI Forum v. Texas Education Agency*

The first case in many years to confront directly the question of whether and how the disparate impact doctrine might apply to high-stakes educational assessments is *GI Forum v. Texas Education Agency*.[240] In *GI Forum*, the Mexican American Legal Defense and Educational Fund sought an injunction barring the Texas Education Agency from using failure on the TAAS as a basis for denying high school diplomas. The complaint, filed by MALDEF on behalf of the GI Forum, Image de Tejas, and seven minority students, alleged that by making passage of TAAS a graduation requirement, the State of Texas

235. *Id.* at 701.
236. *Id.*
237. *Id.*
238. *Id.*
239. *Id.* at 701-03.
240. *See* 87 F. Supp. 2d 667 (W.D. Tex. 2000).

"denies diplomas to Mexican-American and African-American students at a rate significantly higher than that of Anglo students."[241]

After a bench trial on the merits, the district court rejected the plaintiffs' claim that the State's use of TAAS scores to determine which students should earn a high school diploma violated Title VI's disparate impact regulation and the Due Process Clause of the Fourteenth Amendment.[242] With respect to plaintiffs' disparate impact claim, the court held that: (1) the test was validly constructed (i.e., that the test is an accurate measurement of students' mastery of skills and knowledge that the state has deemed important); (2) the State's chosen passing score was not arbitrary or unjustified; and (3) the use of the TAAS exam as a graduation criterion bore a "manifest relationship" to the "legitimate" educational goal of accountability[243] by "guarantee[ing] that students will be motivated" to learn.[244] Moreover, the court rejected as factually unsupportable plaintiffs' contention that the State's use of the exam was causing minority dropout rates to rise as well as plaintiffs' argument that other educational measures would motivate students and teachers as effectively as the TAAS.[245]

Although *GI Forum* was widely regarded as good news for high-stakes educational assessments across the country, the case's actual relevance to standardized tests other than TAAS remains unclear. The uncertainty stems in part from the court's acceptance of the State's assertion that the test "guarantees" motivation.[246] In a

241. Complaint, GI Forum v. Tex. Educ. Agency, C.A. No. SA-97-CA-1278EP (W.D. Tex. Oct. 14, 1997), *reprinted in Testing in Texas: Accountability for Bilingual Students*, READ PERSPECTIVES, Fall 2000, at 15.

242. *GI Forum*, 87 F. Supp. 2d at 668. Earlier in the litigation, the court dismissed on summary judgment plaintiffs' claim of intentional discrimination. *See* Order for Summary Judgment, *supra* note 208; *see also supra* notes 208-09 and accompanying text.

243. In particular, the court reasoned that the State had the right to expect schools and students to be held accountable for their performance. *GI Forum*, 87 F. Supp. 2d at 679.

244. *Id.* at 681.

245. *Id.* at 682. In rejecting the plaintiffs' procedural due process claim, the district court held that the TAAS meets currently accepted standards for curricular validity and that students have had a reasonable opportunity to learn the subjects covered on the exam. The district court also rejected the plaintiffs' claim that the test violated their rights to substantive due process, noting that the test did not constitute a "substantial departure from accepted academic norms." *Id.*

246. In *GI Forum*, the court did not require defendants to offer any statistical proof that the diploma sanction motivates learning, and instead simply accepted at face value the State's contention that the policy motivates learning. *Id.* at 681. Another court might be less deferential and, indeed, might even question whether "accountability" is itself a legitimate educational goal. *See, e.g.*, Cureton v. NCAA, 37 F. Supp. 2d 687, 701-06 (E.D. Pa. 1999) (questioning the legitimacy of the educational goals before considering whether the policy is "manifestly related" to said goals).

different case, however, a court could reach a contrary result if evidence presented at trial indicated that the diploma sanction had less than a "guaranteed" influence on student motivation. And, although the court rejected plaintiffs' assertions that TAAS caused minority students to drop out of school at higher rates than they ordinarily would have, the language of the opinion suggests that if plaintiffs could prove a connection between dropout rates and high-stakes testing, they might prevail on their Title VI claim.[247]

4. *Alexander v. Sandoval*

In April 2001, the U.S. Supreme Court dealt a significant blow to proponents of disparate impact in general, and foes of standardized tests in particular, when it held that Title VI provides no private right of action to enforce the statute's disparate impact regulations. In *Alexander v. Sandoval,* the Court considered a challenge to Alabama's policy of administering state driver's license examinations only in English.[248] In December 1996, Martha Sandoval (a native Spanish speaker) filed a class action lawsuit seeking relief from the state's English-only rule. Among other things, Sandoval alleged that the Alabama policy discriminated on the basis of national origin in violation of Title VI's implementing regulations.[249] In particular, the complaint alleged that defendants' refusal to administer the drivers' examination in languages other than English or to allow for the use of translators or interpretive aids had a disproportionate, adverse impact on foreign-born residents and that this disparate impact rendered the policy unlawful.[250]

The question presented to the Supreme Court in *Sandoval* was whether private plaintiffs can pursue a cause of action to enforce Title VI's disparate impact regulations. By a vote of five to four, the Court held that they cannot.[251]

---

247. *GI Forum*, 87 F. Supp. 2d at 676. No doubt seizing on this language from *GI Forum*, antitesting groups such as tbe Harvard Civil Rights Project have sought to commission research into these very issues with the goal of establishing a record that the tests fail to motivate learning and in fact lead to increased dropout rates. *See* Harvard Civil Rights Project, *supra* note 128.

248. 532 U.S. 275 (2001). Because the Alabama Department of Public Safety accepts grants of financial assistance from the U.S. Department of Transportation, it is a covered entity under Title VI. *See id.* at 278.

249. *See* Sandoval v. Hagan, 7 F. Supp. 2d 1234 (M.D. Ala. 1998), *aff'd*, 197 F.2d 484 (11th Cir. 1999), *rev. sub. nom.* Alexander v. Sandoval, 532 U.S. 275 (2001); *see also* 34 C.F.R. § 100.3 (1990) (prohibiting federally funded institutions from utilizing "criteria or methods of administration which have the effect of subjecting individuals to discrimination").

250. *Sandoval*, 7 F. Supp. 2d at 1244 (citing the plaintiff's complaint).

251. 532 U.S. at 293.

In rejecting Sandoval's claim, the Court, in an opinion by Justice Antonin Scalia, acknowledged that individuals may sue to enforce their statutory right to be free from *intentional* discrimination under Section 601 of Title VI.[252] The Court assumed, for the sake of argument, that the administrative regulations promulgated pursuant to Section 602 of Title VI may lawfully prohibit policies with discriminatory effects.[253] The *Sandoval* majority noted, however, that because Section 602 is phrased as a directive to administrative agencies engaged in the distribution of public funds and makes no mention of the individuals protected by the Act or of the recipients of federal funds, there was no reason to infer that Congress intended to allow private plaintiffs to enforce the administrative regulations.[254] The Court thus held that private parties cannot bring disparate impact lawsuits against recipients of federal funds under Title VI or its implementing regulations.[255]

Sandoval's reach remains unclear. In his dissenting opinion, Justice John Paul Stevens argued that private parties may still sue to enforce Title VI's disparate impact regulations pursuant to 42 U.S.C. § 1983, which authorizes suits for deprivation, under color of state law, of rights secured by the Constitution and the laws of the United States.[256] At least one federal court of appeals has rejected this approach, holding that *Sandoval* precludes all private actions to enforce Title VI's disparate impact regulations.[257] But other courts may decide otherwise. Thus, while *Sandoval* may present some

---

252. *Id.*

253. For purposes of deciding the *Sandoval* case, the Court assumed the validity of the disparate impact regulation itself. *Id.* at 281-82.

254. *Id.* at 289 (citing Univs. Research Ass'n, Inc. v. Coutu, 450 U.S. 754 (1981)).

255. *Id.* at 293. The Court left open, however, the questions of whether the disparate impact regulations (which proscribe activities permissible under the Act alone) are themselves valid and whether the federal government may properly deny federal funding to educational institutions that utilize polices with an adverse racial or ethnic impact.

256. *Id.* at 299 (Stevens, J., dissenting).

257. S. Camden Citizens in Action v. N.J. Dep't of Envtl. Prot., 274 F.3d 771, 774 (3d Cir. 2001). *South Camden Citizens in Action* involved a challenge to the New Jersey Environmental Protection Agency's decision to grant certain air permits to a new cement plant in a predominantly minority neighborhood. Pointing to an already high rate of asthma and other respiratory problems among the largely black and Hispanic residents of the community, plaintiffs argued that the EPA's decision to grant the permits would create an adverse environmental impact on blacks and Hispanics in violation of Title VI's implementing regulations. In holding that the case could proceed even in the advent of *Sandoval*, the district court accepted the suggestion of Justice Stevens in his *Sandoval* dissent that lawsuits to enforce Title VI's disparate impact regulation be allowed to proceed under 42 U.S.C. § 1983. S. Camden Citizens in Action v. N.J. Dep't of Envtl. Prot., 145 F. Supp. 2d 505 (D.N.J. 2001). On appeal, the U.S. Court of Appeals for the Third Circuit reversed, holding that *Sandoval* precludes all private rights of action to enforce Title VI's disparate impact regulations. *S. Camden Citizens in Action*, 274 F.3d at 774.

obstacles to disparate impact challenges to standardized tests, it is by no means the final word on the matter.[258]

Furthermore, in the long run, it is quite possible that Congress will seek to amend Title VI to authorize disparate impact lawsuits explicitly.[259] And, even if Congress refrains from amending the statute, schools that utilize high-stakes tests nevertheless remain subject to investigation by the U.S. Department of Education and risk losing federal financial assistance if they administer exams that exhibit a disparate racial or ethnic impact. Thus, a significant normative question remains—namely, whether it makes sense to apply the disparate impact model of employment discrimination to the education context in general and to the testing arena in particular.

V. ARGUMENTS REGARDING THE APPLICATION OF DISPARATE IMPACT THEORY TO EDUCATIONAL TESTING

Should federal discrimination law prohibit those tests, such as the Texas Assessment of Academic Skills and the Massachusetts Comprehensive Assessment System, that (at least in the short term) may have the effect of preventing disproportionately large numbers of blacks and Latinos from earning a high school diploma? This part considers whether the disparate impact model of discrimination should be applied to educational policies in general and educational assessments in particular. After analyzing arguments for and against the importation of the disparate impact model to the education setting, I conclude that the use of the disparate impact model in the context of high-stakes educational tests is inappropriate. Instead, legal claims of discrimination in the educational testing arena should be limited to claims of *intentional* discrimination. This does not mean, of course, that evidence of disproportionate adverse impact is irrelevant. To the contrary, evidence of disparate impact may be considered as part of the totality of the circumstances in claims of purposeful discrimination[260] and may be politically relevant to determinations regarding test use and consequences. But absent additional evidence of purposeful discrimination, a disparate racial or ethnic impact should not provide a legal barrier to legitimate efforts to

258. *See* Bradford C. Mank, *Using Section 1983 to Enforce Title VI's Section 602 Regulations*, 49 KAN. L. REV. 321 (2001) (advocating use of § 1983 to enforce disparate impact regulations).

259. This is precisely what happened after the Supreme Court acted in *Wards Cove Packing Co. v. Atonio*. 490 U.S. 642 (1989), to cabin the disparate impact approach to proving employment discrimination. 42 U.S.C. § 2000e2-k (2000); 137 CONG. REC. S15,273, S15,276 (daily ed. Oct. 25, 1991) (introducing the 1991 Civil Rights Act and interpretive memorandum).

260. *See* Washington v. Davis, 426 U.S. 229, 242 (1976).

improve the educational system for all students, including those who are disadvantaged.

A. Rationales for Applying the Disparate Impact Model to High-Stakes Tests

Scholars and special interest groups have suggested a variety of rationales for using the disparate impact model to prevent the use of high-stakes educational tests.[261] These advocates, and accounts in the popular press, have sometimes conflated and confused these various rationales. Nevertheless, it is possible to discern three distinct arguments for applying the doctrine of disparate impact to high-stakes educational tests: (1) claims that the disparate impact approach will help smoke out covert, intentional discrimination by educators or educational policymakers; (2) claims that the disparate impact model is necessary to eliminate so-called "biased" tests; and (3) claims that the disparate impact approach is necessary to screen out tests that, although technically unbiased, lead to troubling social outcomes or otherwise entrench existing inequalities. In the following pages, each of these arguments is examined in turn and rejected. In addition, I consider more generally why theories supporting the use of disparate impact in the employment setting are inapplicable in the educational testing arena.

1. Disparate Impact as Objective Evidence of Purposeful Discrimination

Claims that the disparate impact model should be applied to high-stakes educational assessments in order to smoke out covert intentional discrimination have their roots in Professor George Rutherglen's "objective theory of discrimination."[262] According to this theory, the disparate impact model serves as a mechanism for identifying intentional discrimination in the absence of direct evidence of racial or ethnic animus. Such arguments are indeed compelling. After all, who could possibly oppose an improved method of uncovering and punishing intentional discrimination?

This justification is fatally flawed, however, in both the employment and educational settings. To begin with, federal discrimination laws do not, in fact, require plaintiffs to provide direct evidence—that is, "smoking gun" evidence—of discrimination. Indeed,

261. *See, e.g.*, KOHN, *supra* note 127, at 60; NAT'L RESEARCH COUNCIL, *supra* note 15; Moran, *supra* note 31, at 111; Elul, *supra* note 16, at 518.

262. *See supra* notes 148-52 and accompanying text.

the Supreme Court has stressed repeatedly that circumstantial proof can be sufficient to support a finding of intentional discrimination.[263] Thus, the argument that the disparate impact cause of action is the only way for a plaintiff who lacks direct evidence of discrimination to recover for injuries suffered due to racially motivated decisions holds little weight.[264]

In addition, the disparate impact model cannot be justified as an objective method of proving intentional, covert discrimination because the model is too often inaccurate. As Professor Steven Willborn has noted in the employment law context, the disparate impact model is an imprecise mechanism for identifying purposeful discrimination because it elevates one form of evidence of discriminatory intent (statistics) over all others (e.g., a history of discrimination by the defendants). At the same time, the model elevates one form of evidence of innocent motive (the relationship between the test and legitimate business or educational objectives) over others (such as contemporaneous statements or legislative history).[265] As such, the disparate impact model underdeters ill-motivated, purposeful discrimination and overdeters well-intended practices with disproportionate racial or ethnic consequences.

In other words, the disparate impact model is unable to eliminate an exam policy enacted with *covert* racial animus where the exam can be defended on legitimate pedagogic grounds. Suppose, for example, that a local school board is comprised entirely of bigots who seek to prevent black and Latino students from receiving high school diplomas in order to keep them from qualifying for certain jobs in the local economy.[266] Suppose further that the racist board members do not publicly admit their motives or otherwise outwardly demonstrate their bigotry. The board members approve a new policy requiring students to pass a standardized exit exam in order to receive a high school diploma. The test they select is educationally sound and linked to the curriculum. During the first few years of the test's administration, the policy results in a disproportionate number of

263. *See, e.g.*, Reeves v. Sanderson Plumbing Prod., Inc., 530 U.S. 133 (2000) (holding tbat jury may infer intentional discrimination from the plaintiff's prima facie case combined with disbelief of defendant's proffered nondiscriminatory justification for its action).

264. Roger Clegg has also pointed out the absurdity of justifying disparate impact as a means of proving intentional discrimination. *See* CLEGG, *supra* note 135, at 9 ("[T]he fact that an offense is difficult to detect would not in any other context justify redefining the offense.").

265. Willborn, *supra* note 18, at 808.

266. It is unclear how often a situation such as this will actually occur. Nevertheless, the objective evidence theory is premised on its occurrence with some regularity. The scenario thus provides a useful illustration for testing the value of the objective evidence theory as a justification for the disparate impact model.

minorities failing to graduate. Nevertheless, the use of the test also results in improved academic achievement by students of all backgrounds. In this situation, the use of a neutral exam was approved with a discriminatory purpose. Because the motives of board members are well disguised, direct evidence of intentional discrimination is lacking. In this hypothetical, the disparate impact model would fail to "smoke out" an intent to discriminate because the policy itself is pedagogically sound and sufficiently correlated with legitimate educational goals.[267]

Conversely, the model overdeters in that it can result in the elimination of policies which were adopted without racial animus—indeed, it can eliminate policies adopted in order to *help* students of color—but that cannot be proved to result in the attainment of a particular educational goal. For example, suppose that a state legislature enacts a high-stakes testing regime with the intention of equalizing educational opportunities across racial and ethnic lines and across various regions of the state. The legislative history indicates that the policy's sponsors hoped that the testing regime would achieve this goal by: (1) providing incentives for teachers in disadvantaged communities to focus on a uniform, core curriculum; (2) motivating teaching and student learning; and (3) targeting scarce resources toward the most disadvantaged schools. Suppose further that, despite the state's best efforts to provide remedial assistance to failing students and increase financial support for disadvantaged school districts, after five years the policy has resulted in only minor improvements in student achievement overall and across racial and ethnic groups. Moreover, the policy has had no measurable effect upon student or teacher motivation. Finally, suppose that during the first five years of the policy's implementation, the policy has led to falling graduation rates among African-American and Hispanic students. There is evidence that the falling minority graduation rates are due not only to high failure rates on the test, but also to increased dropout rates among black and Hispanic students in response to the threat of the diploma sanction.

---

267. Some commentators have argued that if an educational policy proves beneficial, then the fact that those who sponsored the measure were motivated by discriminatory animus should be irrelevant. *See, e.g.*, THERNSTROM & THERNSTROM, *supra* note 8, at 20-21 (criticizing the reasoning of *United States v. Fordice*, 505 U.S. 717, 733-34 (1992), in which the U.S. Supreme Court held that a current and valid educational justification cannot cleanse a policy originally enacted for a discriminatory purpose). Irrespective of the merits of this position, the objective theory of discrimination is clearly premised on the notion that the law should prohibit those actions motivated by racial or ethnic animus, and as a tool for achieving this objective, the disparate impact model fails.

Given these facts, courts applying the disparate impact model of discrimination would probably strike down the testing regime as violative of Title VI. A policy that has not measurably improved overall student achievement or narrowed the racial achievement gap, and which can be causally linked to rising minority dropout rates, can hardly be said to be "educationally necessary," as courts have thus far defined the term. Yet, whatever one thinks of this result, it is inconsistent with the notion that the disparate impact model is a vehicle for eliminating policies enacted with actual racial or ethnic animus.

The hypothetical scenario offered above may indeed be politically and socially unacceptable. After five years, policymakers might reasonably conclude that, despite their best intentions, the testing regime has failed to begin the process of equalizing educational opportunities between various demographic groups. They may therefore conclude that the policy should be abandoned. But while the hypothetical raises legitimate concerns about the policy's social impact,[268] it certainly does not implicate concerns regarding *purposeful* discrimination, which, after all, is the basis for the objective evidence rationale. Thus, the objective evidence theory fails to provide a coherent theoretical justification for applying the disparate impact model to educational testing.

2. The Bias Rationale

Some commentators argue that the disparate impact model should be applied to the education setting not so much to root out purposeful discrimination by those who enacted or implemented the testing regimes, but rather to eliminate so-called biased tests.[269] This view is rooted in the notion that disproportionate demographic outcomes on standardized tests are a product of faulty tests, rather than the result of an actual academic gap.[270] According to this view, standardized exams fail to measure accurately the actual skill level or

268. *See infra* notes 300-14 and accompanying text.

269. *See, e.g.*, Elul, *supra* note 16, at 517.

270. *Id.* ("Disparate impact analysis in the education context . . . rejects the belief that group test score differences reflect actual differences in test-takers' abilities. Rather, the disparate test scores would reflect discrimination in the tests themselves."); *see also* Regents of Univ. of Calif. v. Bakke, 438 U.S. 265, 306 n.43 (1978) (Powell, J.) (noting the possibility, but lack of record evidence indicating, that tests are "culturally biased" and thus inaccurate with respect to minority test-takers); Lani Guinier & Susan Sturm, *The Future of Affirmative Action: Reclaiming the Innovative Ideal*, 84 CAL. L. REV. 953, 987 (1996) (arguing that affirmative action is just a band-aid solution to the problem of biased tests that fail to predict or measure accurately the abilities of people of color); White, *supra* note 16, at 90 (arguing that test scores mask the true ability of minority test-takers).

knowledge base of minority test-takers vis-à-vis their white counterparts.[271] The argument is thus that the exams themselves "discriminate."

a. Bias Defined

Before discussing whether claims of test bias provide a coherent rationale for the use of disparate impact litigation against high-stakes educational tests, it is necessary to define the term briefly. Although in recent years "test bias" has become a popular catch-phrase often used by lay persons to explain racial or ethnic differences on test scores, the term "test bias" actually has a precise, scientific meaning. Psychometricians define "bias" as technical flaws in an examination that lead to incorrect measurements for one or another subgroup of test-takers.[272] In other words, test bias refers to the failure of a test to assess accurately the abilities of examinees from a particular demographic group. This problem often arises when language barriers prevent one or more groups of test-takers from understanding the questions asked by a test. Consider, for example, a test administered in English to Latino students who speak only Spanish. If the test seeks to measure a trait, such as IQ, which is *unrelated* to the ability to speak English, the test will inevitably underestimate the abilities of the Spanish-speaking children.[273] In other words, the test would be "biased" against non-English-speaking Latinos. If, however, the test actually aims to measure the students' English language proficiency or the ability to solve problems using the dominant language of our culture (i.e., English), then the source of the

271. White, *supra* note 16, at 95-106 (arguing that low minority test scores are "the results of defects in the measurement processes themselves").

272. *See* GREGORY CAMILLI & LORRIE A. SHEPARD, METHODS FOR IDENTIFYING BIASED TEST ITEMS 8 (1994). Camilli and Shepard have analogized test bias to "clocking individuals to measure their running speed, but using a stopwatch that runs too slowly for black runners." *Id.* Accordingly, relative comparisons between black runners would remain accurate, while comparisons between whites and blacks or between group averages would be distorted by bias. The test would be biased with respect to blacks because the scores obtained by the black runners do not mean the same thing as the identical score obtained by a white runner. *Id.*; *see also* ANNE ANASTASI, PSYCHOLOGICAL TESTING (4th ed. 1976) ("In the psychometric sense, test bias refers to over prediction or under prediction of criterion measures. If a test consistently under predicts criterion performance for a given group it shows unfair discrimination or 'bias' against that group.").

273. This was the scenario in *Diana v. Board of Education*, No. C-70-37RFP (N.D. Cal., settled 1973), in which an IQ test administered in English to Spanish-speaking children was used to place Latino students in classes for the mentally retarded. *See id.* (cited in CAMILLI & SHEPARD, *supra* note 272, at 7-8); *see also* Christopher Jencks, *Racial Bias in Testing*, *in* THE BLACK-WHITE TEST SCORE GAP 55, 56 (Christopher Jencks & Meredith Phillips eds., 1998) (drawing a similar analogy).

Latino students' difficulty with the exam is *relevant* to the trait being measured, and the test is not "biased."

Some commentators and lay persons reflexively assume that any group test score difference must have been caused by a flawed and discriminatory exam.[274] But this view confuses the analytically distinct concepts of "disparate impact" and "test bias."[275] Although a biased test (or test question) may result in disparate demographic outcomes, it also is possible for a biased test (or question) to yield demographically proportional results.[276] Moreover, while a biased test (or question) can yield statistically significant disparities in outcome, such disparities plainly can be caused by factors other than bias. Thus, while differences in the mean scores of demographic groups could theoretically occur because of test bias, different group outcomes might just as easily be understood as reflecting *actual* differences between the groups in the area measured by the test.[277] Absent additional evidence, a finding of disparate impact points nowhere. It simply begs the question: Are racial or ethnic differences in average test scores caused by defects in the exam (in which case the exam is properly said to be "biased") or are such differences indicative of actual educational gaps (in which case the test is not "biased")?

---

274. *See, e.g.*, Elul, *supra* note 16, at 517.

275. *See, e.g.*, CAMILLI & SHEPPARD, *supra* note 272, at 7; Irving Lorge, *Difference or Bias in Tests of Intelligence*, *in* TESTING PROBLEMS IN PERSPECTIVE (Anne Anastasi ed., 1966) (explaining as faulty the notion that where there are group differences, there is bias, and where there are no group differences, there is no hias).

276. Jencks, *supra* note 273, at 67 (describing research by McGurk in which the black-white test score gap was found to be twice as large on items deemed "least cultural" than on those rated "most cultural"). Despite widespread claims that so-called "culturally biased" test questions are to blame for uneven demographic test results, evidence suggests that "cultural bias" is not the source of disproportionate outcomes on most tests. Indeed, empirical evidence indicates that African-Americans perform equally or better on questions deemed to be "culturally loaded" than those deemed to be culturally neutral. *See, e.g.*, LEE J. CHRONBACH, ESSENTIALS OF PSYCHOLOGICAL TESTING 336 (1990) (noting that the black-white gap is virtually nonexistent on many items that critics regard as unfair); Flaugher, *supra* note 1, at 671 (describing research by Bianchini in which eliminating thirteen items judged to be "culturally biased" from a test of eighty-two items did not improve the performance of schools with high minority populations relative to their performance on the original "biased" version); *see also* Janice Dowd Scheuneman & Thomas Oakland, *High Stakes Tests*, *in* TEST INTERPRETATION AND DIVERSITY: ACHIEVING EQUITY IN ASSESSMENT 85 (Sandoval et al. eds., 1998) (conceding that the evidence is "substantial" that tests are unbiased and valid across cultures). This does not mean, of course, that exams containing so-called "culturally biased" test questions can never produce adverse racial or ethnic outcomes. To date, however, evidence of such a phenomenon is lacking.

277. Indeed, there are a variety of factors (including income, level of educational resources, family structure, and cultural differences in family emphasis on education), that might cause, or contribute to, the creation of real educational deficits among different demographic groups that are then reflected in test scores.

The concept of so-called test bias is also frequently confused with that of test "invalidity." Test "invalidity" refers to the inaccuracy of an exam across subgroups.[278] A test that is an inaccurate measurement of the abilities of *all* test-takers (irrespective of race, ethnicity, sex, etc.) is invalid, but not biased.[279] Although a biased test is invalid with respect to those subgroups it fails to measure accurately, a test that is an invalid measurement device for all subgroups does not suffer from test bias.

A third set of issues that are often incorrectly referred to as involving "bias," but which do not technically fall into the realm of bias, are those sociopolitical difficulties that may result from improper test naming and from public misperceptions that a test reflects evidence of something other than that which it was designed to measure. When an achievement test (like a high school exit exam) is mislabeled or misinterpreted (by test administrators, policymakers, or the public at large) as a test of *innate* ability, certain groups may suffer adverse public policy consequences.[280] For example, as Ronald Flaugher has pointed out, when a test is viewed as a measure of achievement, low performance can appropriately create public pressure for an increase in resources to improve performance. If, on the other hand, the test result is improperly understood to be a measure of inherent ability, then a low score might be wrongly viewed as a basis for withholding additional resources on the ground that spending more money on those who lack the ability to improve is wasteful.[281] The fear of these adverse policy consequences, along with a fear of the stigma[282] that may result when an achievement test is improperly perceived or portrayed as an ability test,[283] has led many civil rights activists to call for elimination of the tests altogether.[284]

---

278. CIZEK, *supra* note 33.

279. CAMILLI & SHEPARD, *supra* note 272, at 8. Modifying Camilli and Shepard's analogy regarding runners, invalidity—but not bias—would be present if the same broken stop watch were used to clock the times of all runners, irrespective of race. Thus, in the educational testing context, when a high school exit exam administered to students statewide is found to contain flaws which create measurement errors among all demographic groups, the test is invalid, but not biased.

280. *See* Flaugher, *supra* note 1, at 672; Jencks, *supra* note 273, at 55-56; Christopher Jencks & Meredith Phillips, THE BLACK-WHITE TEST SCORE GAP: AN INTRODUCTION, *in* THE BLACK-WHITE TEST SCORE GAP, *supra* note 273, at 1, 13.

281. Flaugher, *supra* note 1, at 672.

282. *See* Jencks, *supra* note 273, at 56 (explaining this concern).

283. Flaugher, *supra* note 1, at 672-73 (noting that some people view low test scores "not as a diagnosis that will lead toward attempts to cure the illness, but rather . . . as an official certification that no help is possible").

284. *See id. But see* Jencks, *supra* note 273, at 56 (arguing that the proper way to eliminate this problem is to change the names of the tests).

But while the failure of policymakers or the public to understand the purpose of a test certainly poses a sociopolitical problem, it is not a problem of "bias" (defined as a measurement error with respect to certain subgroups of test-takers).[285] Unfortunately, the tendency to confuse problems of impact, validity, or labeling with problems of "bias" muddies the public discourse and confuses the legal landscape.

In the context of high-stakes achievement tests, it is unclear exactly what testing critics mean when they attack tests as "biased." The educational assessments that are the topic of this paper necessarily measure, among other things, a student's ability to read, write, and analyze problems in standard English and, therefore, are not biased simply because those who have not mastered the English language underperform in comparison with others. Nor has credible evidence been presented that any particular test fails to measure accurately the achievement of minority students. To the contrary, there is a consistent body of evidence that indicates that disparate demographic test scores on high school exit exams reflect real and persistent educational gaps between ethnic groups and are not simply measurement errors. For example, the achievement gap between black and Latino students, on the one hand, and their white peers, on the other hand, has been found to be present across tests and across assessment devices. Thus, data from the National Assessment of Educational Progress ("NAEP"),[286] the National Educational

285. Harvard social scientist Christopher Jencks refers to this set of problems as "labeling bias." Jencks, *supra* note 273, at 58. In my view, the term "bias" is not appropriate to describe this set of concerns; it appears, rather, that Jencks has identified an important, yet distinct, sociopolitical problem of mislabeling. *See id.* at 55-56, 88.

286. The NAEP is a federally mandated series of tests, often referred to as the "nation's report card," which reports on academic achievement of fourth, eighth, and twelfth grade students across the country. Administered every two years through the U.S. Department of Education's National Center for Education Statistics, the NAEP is currently the only assessment that provides information on the level of achievement of a representative sample of American students. The NAEP does not provide individual student or school-specific results. (Students who are tested are selected randomly and their names are not collected.) Rather, it provides a composite picture of American students' strengths and weaknesses in basic and higher-order skills; comparisons of achievement by race/ethnicity, gender, type of community, and region; and trends in performance across the years. *See* THERNSTROM & THERNSTROM, *supra* note 8, at 352; Krista Kafer, *A Guide to the NAEP Academic Achievement Test*, HERITAGE FOUND. BACKGROUNDER (Heritage Found., Washington, D.C., Mar. 15, 2002); Georgia Department of Education, *Research, Evaluation, & Testing, National Assessment of Educational Progress*, *at* http://www.doe.k12.ga.us/sla/ret/naep.html (last visited May 14, 2002).

On the 1999 NAEP, for example, the average black seventeen-year-old was able to read only at the level of the average white thirteen-year-old. In the same year, the average black seventeen-year-old demonstrated the math skills of the average white thirteen-year-old. In other words, NAEP data indicates that by the age of seventeen, the typical black student lags four years behind his or her white counterparts. THERNSTROM & THERNSTROM, *supra* note 7; *see also* NAT'L CTR. FOR EDUC. STATISTICS, U.S. DEP'T OF EDUC., NAEP 1998 READING REPORT CARD:

Longitudinal Survey ("NELS"),[287] and the SAT[288] all confirm the result of state educational assessments: African-American and Latino students lag behind their peers from other ethnic groups at every educational level.[289] And it is not just standardized test scores that

---

NATIONAL & STATE HIGHLIGHTS 8-9 (1999), *available at* http://nces.ed.gov/pubsearch/pubsinfo.asp?-pubid=1999479; NAT'L CTR. FOR EDUC. STATISTICS, U.S. DEP'T OF EDUC., NAEP 1999 TRENDS IN ACADEMIC PROGRESS: THREE DECADES OF STUDENT PERFORMANCE 33 (2000), *available at* http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000469.

287. The NELS is a large-scale longitudinal study of high school students conducted by the U.S. Department of Education's National Center for Education Statistics ("NCES"). Begun in 1988, it provides trend data about critical transitions experienced by eighth grade students as they progressed through high school, secondary school, and/or the workforce. *See* Center for Health & Wellbeing, Woodrow Wilson School, Princeton University, *National Education Longitudinal Survey—1988, at* http://www.wws.princeton.edu/~kling/surveys/NELS88.html. Data from the NELS confirms that black and Hispanic high school seniors lag behind their white counterparts in both reading and math. *See* NAT'L CTR. FOR EDUC. STATISTICS, U.S. DEP'T OF EDUC., STATISTICAL ANALYSIS REPORT—NATIONAL EDUCATION LONGITUDINAL STUDY OF 1988: TRENDS AMONG HIGH SCHOOL SENIORS, 1972-1992, at ii, 25-27 (1995) [hereinafter NCES, TRENDS AMONG HIGH SCHOOL SENIORS], *available at* http://nces.ed.gov/pubs95/95380.pdf. (noting that in 1992, the effect size of the difference between black and white reading scores was 0.74; the effect size of the difference between Mexican-American and white reading scores was 0.61, the effect size of the difference between Puerto Rican and white reading scores was 0.45, and the effect size of the difference between other Hispanic and white reading scores was 0.43); *id.* (noting that in 1992, with respect to math scores, the effect size of the difference between black and white seniors was 0.97; the effect size of the difference between Mexican-American and white seniors was 0.70; the effect size of the difference between Puerto Rican and white students was 0.58; and the effect size of the difference between other Hispanic and white students was 0.38).

288. The SAT I is a test that measures verbal and mathematical reasoning abilities. Educational Testing Service, *Complete Test Directory: Test Directory-S, at* http://www.ets.org/tests/stest.html (last visited Apr. 14, 2002). Among students who planned to enter college in the fall of 1999, the average score of African-Americans students on the SAT I Verbal was ninety-three points below that of the average white student. Blacks scored, on average, 106 points below whites on the SAT I Math. Frontline: Secrets of the SAT—The Test Score Gap (PBS television broadcast), *available at* http://www.pbs.org/wgbh/pages/frontline/shows/sats/etc/gap.html (last visited Apr. 14, 2002).

289. *See generally* THERNSTROM & THERNSTROM, *supra* note 7; *see also* Jencks & Phillips, *supra* note 280, at 1 (noting that the average African-American student scores below seventy-five percent of whites on most standardized tests). There is some data that indicates that the racial achievement gap has begun to close since the 1970s. Thus, comparing data from the NELS with data from two other longitudinal studies (the National Longitudinal Study of the Class of 1972 and High School and Beyond) seems to indicate that the achievement gap between white high school seniors and black and Latino seniors has narrowed, although disparities still remain. *See* NCES, TRENDS AMONG HIGH SCHOOL SENIORS, *supra* note 287, at ii, 25-27 (noting that with respect to reading scores, the effect size of the difference between black and white seniors decreased from 0.97 in 1972 to 0.74 in 1992 and that during the same time period the effect size of the difference between Mexican-American and white students decreased from 0.89 to 0.61, the effect size of the difference between Puerto Rican and white students decreased from 0.93 to 0.45, and the effect size of the difference between other Hispanic and white students decreased from 0.91 to 0.43); *id.* (noting that with respect to math scores, the effect size of the difference between black and white seniors decreased from 1.09 in 1972 to 0.97 in 1992, and that during the same time period the effect size of the difference between Mexican-American and white seniors decreased from 0.86 to 0.70, the effect size of the difference between Puerto Rican and

reveal this learning deficit. Grade point averages, graduation rates, and class rankings of students across the country are, regrettably, also consistent with this pattern,[290] indicating that claims of bias are, at best, exaggerated.[291]

---

white students decreased from 1.1 to 0.58, and the effect size of the difference between other Hispanic and white students decreased from 0.86 to 0.38). However, as Abigail Thernstrom and Stephan Thernstrom explain in their forthcoming book, since 1988 the trend has been toward a widening achievement gap in reading. THERNSTROM & THERNSTROM, *supra* note 7.

290. *See, e.g.*, Fredrick E. Vars & William G. Bowen, *Scholastic Aptitude Test Scores, Race, and Academic Performance in Selective Colleges and Universities, in* THE BLACK-WHITE TEST SCORE GAP, *supra* note 273, at 457, 461, 465. In a survey of eleven selective colleges and universities, African-American students were found to have lower college graduation rates than whites at every level of SAT score and lower average GPAs than their white counterparts (2.80 versus 3.30). *Id.* An even larger difference in GPAs was found when other differences including gender, athletic participation, institution, and field of study were controlled. *Id.* Pedro A. Noguera and Antwi Akom have written that "[f]or years, African-American, Latino and Native American students have lagged far behind their white and Asian peers on most standardized tests. The gap is also present in [high school] graduation and dropout rates, [high school] grades and most other measures of student performance." Pedro A. Noguera & Antwi Akom, *Disparities Demystified: Causes of the Racial Achievement Gap All Derive from Unequal Treatment*, NATION, June 5, 2000, at 29, 29. Unlike critics who refuse to recognize that the test gap represents an actual achievement gap, Noguera and Akom openly and refreshingly recognize that the problem of minority underachievement is real. They blame this problem, however, on discrimination in the K-12 education system itself. *Id.*

291. Although the reasons why black and Latino students lag behind their white and Asian-American peers is beyond the scope of this Article, some discussion is required. For years education experts argned that differences in test scores could be traced to differences in resources between minority schools and white schools. Recent studies, however, indicate that funding differences, more than likely, are not the reason for the disparities, as most school resources are, in fact, equally distributed between blacks and whites. As Christopher Jencks and Meredith Phillips have explained, "the average black child and the average white child now live in school districts that spend almost exactly the same amount per pupil." Jencks & Phillips, *supra* note 280, at 9. Instead, Jencks and Phillips argue, "the most important resource difference between black and white schools seems to be that teachers in black schools have lower test scores than teachers in white schools." *Id.* at 10. It is therefore reasonable to hypothesize that a significant cause of poor minority performance on standardized tests is less academically prepared teachers. *Id.* With respect to factors outside of the school itself, research by Jencks and Phillips suggests that the attitude of a student's parents toward learning and education may be a significant factor in the achievement gap. *Id.* at 24. Indeed, Jencks and Phillips have found that parenting practices and the value placed on education in the home is more closely correlated with a student's test scores than household income or even the parents' own level of educational attainment. *Id.* For further discussion of the academic achievement gap, see generally THERNSTROM & THERNSTROM, *supra* note 7 (explaining some of the cultural bases for the gap), and John McWhorter, *Why the Blaek-White Test Gap Exists*, 5 AM. EXPERIMENT Q. 45 (2002) (rejecting the notion that the average black student underperforms vis-à-vis the average white student because of underfunding, poverty, or discrimination in the educational system, and arguing that the performance gap is a result largely of the "acting-white syndrome"—that is, the phenomenon whereby many black students pejoratively label blacks who succeed academically as "acing white").

b. *Test Bias as Rationale for Applying the Disparate Impact Model to High-Stakes Educational Tests*

The fact that the average black or Hispanic student lags behind the average white student on *all* accepted measures of educational achievement strongly suggests that no one measurement device is aberrational or "biased."[292] Yet the disparate impact model enshrines the opposite presumption into law: tests with disproportionate demographic results are presumed unlawful unless the educational

---

292. Scheuneman & Oakland, *supra* note 276, at 85 ("Overall, the evidence is substantial that tests are unbiased and valid for use with both sexes and with minority populations.").

When faced with this consistent body of evidence, antitesting commentators often respond that *all* measurement devices are biased or that the domains that tests seek to measure are themselves biased. These scholars and advocates have sought to vastly expand the definition of bias to include consideration of so-called "unfairness" in the underlying criterion measured by the tests as well as global claims of societal discrimination. For example, William Kidder has argued that conventional models of test bias mask the bias of the Law School Admissions Test ("LSAT") because they consider the ability of the test to predict law school grades—a criterion that Kidder believes is also infected with gender and racial bias. Kidder, *supra* note 19, at 169-70 n.6 (arguing that validity cannot be established merely by correlating test scores with outcomes and that the outcomes must also be evaluated for bias). Other commentators have even argued that standardized tests are biased because they seek to measure areas of knowledge that are familiar to students from white-European backgrounds and fail adequately to assess the alleged separately acquired knowledge base of minority students. *See* THERNSTROM & THERNSTROM, *supra* note 8, at 364-65 (describing this critique). This view is expressed in claims that a test of American history is biased because it seeks to assess students' knowledge of historical figures who are predominantly white and male as well as in claims that science exams are biased because they seek to measure a student's ability to think linearly, a trait which some radical scholars claim is itself "wbite." *Id.*

These critiques are rooted in broader attacks on traditional definitions of merit, wbich contend that merit and standards are themselves subjective concepts—mere social constructs created by those in power as a means of protecting their place in society. *See, e.g.*, PATRICIA A. WILLIAMS, THE ALCHEMY OF RACE AND RIGHTS 99 (1991) (adopting the radical view of merit and bias); Richard Delgado, *Rodrigo's Tenth Chronicle: Merit and Affirmative Action*, 83 GEO. L.J. 1711, 1719, 1740-45 (1995) (similar); Robin West, *Constitutional Fictions and Meritocratic Success Stories*, 53 WASH. & LEE L. REV. 995, 1018 (1996) (similar); Daria Roithmayr, *Deconstructing the Distinction Between Bias and Merit*, 85 CAL. L. REV. 1449 (1997) (similar); *see also* DANIEL FARBER & SUZANNA SHERRY, BEYOND ALL REASON: THE RADICAL ASSAULT ON TRUTH IN AMERICAN LAW (1997) (describing the radical critique of merit and bias). Radical critiques posit that merit standards, in general, and standardized tests, in particular, can never be fair or objective and that the entire enterprise of testing is pernicious because it fails to consider principles of distributive justice. *See, e.g.*, Delgado, *supra*, at 1719, 1740-45.

Because the radical critiques reject the concepts of merit and testing altogether, they are not helpful in determining the validity of the disparate impact approach in the educational testing context. Indeed, the very premise of the bias rationale for the disparate impact model is that this model can belp to separate out those tests that are biased against certain subgroups from those that are unbiased in a way that the disparate treatment model cannot. If, however, one subscribes to the view that *all* tests are inherently biased and unfair, or that tbe criteria or standards a test seeks to measure are culturally unfair, then disparate impact litigation becomes unnecessary. We need only eliminate all standardized tests.

system can affirmatively prove their value. This approach is not only counterintuitive, it is ineffective.

Indeed, the application of the disparate impact model fails to shed light on the question of whether disproportionate test results are due to test bias, test invalidity, actual gaps in knowledge, or some other set of factors. That is so because the disparate impact doctrine does not expressly require consideration of whether the objectionable disparate impact was actually caused by bias in the exam itself. It requires only that the educational institution prove that the test bears a *substantial relationship* to a *legitimate* educational objective. Depending on how this legal standard is applied, disparate impact litigation could prove to be both underinclusive and overinclusive in achieving the professed goal of the bias rationale—i.e., eliminating tests that inaccurately measure minority achievement.

Suppose, for example, that a court accepts as offered in good faith a school district's goal of accountability. Assume further that defendants proffer testimonial and statistical evidence demonstrating that the test and the threat of the diploma sanction help to motivate students to study and teachers to focus on particular aspects of the curriculum. The court could then plausibly find that the school district had demonstrated the "educational necessity"[293] for the testing regime, even if the test is *actually biased* against minority test-takers.

Even where a test is deemed "educationally necessary," plaintiffs may argue that there are less discriminatory assessment devices that can achieve the same goal. But suppose that the alternatives cited by the plaintiffs fail to motivate learning in the same manner as the test. Or suppose that other assessment devices have a similar adverse impact as the challenged test. Under the disparate impact model, a court could very well uphold this exam, even though it is technically biased. In this scenario, the disparate impact model fails to eliminate a biased exam; it is therefore underinclusive.

More likely, however, is the opposite scenario, in which use of the disparate impact model results in the elimination of a fair and unbiased test. Suppose that the plaintiffs proffer statistical evidence demonstrating that over a five-year period the requirement that students pass an exit exam in order to receive a diploma has failed to motivate learning and, indeed, has caused the dropout rate among minority students to increase significantly. In these circumstances, the school district could be found liable for disparate impact discrimination even if the testing policy was well intentioned and the

293. *See supra* Part V.B.1.

exam proved bias free (and even, it should be noted, if the test might have improved outcomes in the long run).[294] Thus, the disparate impact model fails to achieve the stated goal of the bias rationale—that is, eliminating tests that are *inaccurate* measurements of minority performance due to "cultural bias." "Bias," therefore, cannot be a viable basis for applying disparate impact doctrine to standardized tests.

c. *Reformulating the Disparate Impact Model to Conform to the Bias Rationale*

Several recent proposals for reform appear, on their face, to address the underinclusiveness and overinclusiveness of the bias rationale. In 2000, the *Resource Guide* developed by the Clinton Administration's Department of Education suggested that "educational necessity" be redefined to require proof of the test's scientific validity (in other words, a test's accuracy and reliability for measuring the constructs that it purports to measure across demographic groups).[295] Defining "educational necessity" to mean "scientific validity" would indeed reduce the disparate impact model's potential for underinclusiveness under the bias rationale. Requiring proof of scientific validity would eliminate those tests whose adverse impact and invalidity result from bias. At the same time, however, the problem of overinclusiveness would be exacerbated. This is so because, as explained previously,[296] while all biased tests are invalid tests, not all invalid tests underestimate the skills or knowledge of *particular subgroups* of test-takers vis-à-vis the entire group. To the contrary, a test can be invalid for a host of reasons that affect all examinees. Thus, even a disparate impact model that requires proof of scientific validity in order to prove educational necessity would be overinclusive from the perspective of the bias rationale in that it would screen out exams that, while perhaps flawed with respect to all test-takers, are not racially or ethnically biased.[297] And while eliminating flawed

294. As previously stated, this scenario is indeed troubling and may give educators and policymakers cause to reconsider or reformulate the testing policy. But a test's undesirable social impact does not call into question the legality of the test itself.

295. *See* RESOURCE GUIDE, *supra* note 220, ch. 2. This approach is also favored by Professor Jay Heubert who advocates giving binding authority to the American Educational Research Associations' Standards for Educational and Psychological Testing. *See* NAT'L RESEARCH COUNCIL, *supra* note 15, at 226.

296. *See supra* Part V.A.2.a and text accompanying notes 280-81.

297. It is perhaps for this reason that, with respect to employment testing, the Supreme Court has expressed a reluctance to require proponents of standardized tests to introduce formal validation studies demonstrating that the exams predict actual performance. *See, e.g.*, Watson v. Fort Worth Bank & Trust, 487 U.S. 997, 998 (1988).

exams is, indeed, a worthy educational goal, it is not the role of antidiscrimination law to do so.

The critical normative question, then, is this: Why must we go through the whole disparate impact charade of trying to prove "educational necessity," or even scientific validity, if what really concerns us as a matter of antidiscrimination law is test bias? Federal disparate impact litigation seems a costly and imprecise means of attacking this problem (if, indeed, the problem exists at all). To be intellectually honest, to avoid confusion in the law, and to achieve a fairer testing regime, those who are concerned with test bias should promote reforms that would require explicit consideration of test bias—either through regulation to eliminate unfair questions (but not the testing regime itself)[298] or as part of the "totality of the circumstances" inquiry under a claim of intentional discrimination.[299]

3. The Social Impact/Inequality Rationale

Distinct from the intentional discrimination and bias rationales is the argument that disparate impact litigation is a necessary mechanism for the elimination of barriers to social progress. According to this rationale, the crux of the issue is not whether the tests are biased or otherwise flawed (most of them are not), but rather the social impact and harm that such policies cause to vulnerable demographic groups.[300] Many of the commentators who support the

298. In fact, concerns of so-called test bias are already being addressed by test makers and local regulatory agencies, which today go to great lengths to make sure that test questions are appropriate to a wide range of cultural backgrounds. Indeed, testing companies and state regulatory agencies already employ high level "sensitivity reviews" in order to remove test items that may prove offensive or insulting to certain demographic subgroups. In Massachusetts, for example, state law requires that "assessment instruments . . . be designed to avoid gender, cultural, ethnic or racial stereotypes." MASS. GEN. LAWS ANN. ch. 69, § 1I (West Supp. 2002). And in Texas, each question on the TAAS exam is screened nine times for any racial or ethnic bias by a special committee. Reviewers are responsible for considering whether a test question with an adverse demographic outcome is culturally contingent *and* unrelated to the skills or body of knowledge the test aims to measure. Where a biased question produces an inaccurate result, the offending question is eliminated. Significantly, however, the testing regime itself remains in place. *See* GI Forum v. Tex. Educ. Agency, 87 F. Supp. 2d 667, 672 (W.D. Tex. 2000); Texas Education Agency, *Test Development Process*, *at* http://www.tea.state.tx.us/student.assessment-/develop/devproc.html (last visited Jan. 30, 2001); Jeff Webb, *Litmus Tests for Racial Bias*, TEX. A & M BATTALION, 1999.

299. *See, e.g.*, Parents in Action on Special Educ. v. Hannon, 506 F. Supp. 831, 875 (N.D. Ill. 1980); Larry P. v. Riles, 495 F. Supp. 926, 959-66 (N.D. Cal. 1979). In both of these cases, the courts considered claims of test bias in conjunction with plaintiffs' claims of intentional discrimination.

300. *See, e.g.*, Chamallas, *supra* note 153, at 377 n.334 (citing former EEOC Chairman Eleanor Holmes Norton as arguing that the concept of validity, or technical accuracy, provides an

use of disparate impact theory on this basis acknowledge that group differences in test scores reflect an *actual* achievement gap, but they worry that penalizing such students by withholding a diploma will have a negative social outcome, thus further entrenching current inequalities.[301] According to this argument, high short-term failure rates in minority communities will unfairly disadvantage students who have not been afforded an equal educational opportunity to master the material on the test and will result in a caste-like system in which the majority of students from the dominant class receive a diploma credential while large numbers of students from disadvantaged communities are forced to navigate life without it.

This argument mirrors Professor Michael Perry's argument in favor of applying the disparate impact model to public employment policies so as not to "reinforce social evils" created by a racist state in the first instance.[302] Yet the two contexts—employment and educational testing—are different. In the employment context, the concept of harm to a disadvantaged job seeker is obvious: not being hired. Such outcomes may directly serve to entrench existing inequalities. While the employer may benefit, there is no corresponding direct benefit to the individuals affected, either immediately or in the long run. On the other hand, in the educational arena it is quite possible that a testing regime will operate not to entrench inequalities, but to alleviate them, both in the short term and in the long run.

"out" for those who create and/or administer tests with a racially harmful effect and, thus, avoids the more significant question of social harm).

301. *See* Jay P. Heubert, *Nondiscriminatory Use of High-Stakes Tests: Combining Professional Test-Use Standards with Federal Civil-Rights Enforcement*, 133 EDUC. L. REP. 17, 27 (1999). In the Executive Summary to the National Research Council's study on the use of high-stakes tests, editors Jay Heubert and Robert Hauser implicitly recognize that the lower test scores of racial and ethnic minorities reflect a real educational deficit. NAT'L RESEARCH COUNCIL, *supra* note 15, at 4 (noting that disparate racial outcomes on high-stakes tests "reflect persistent inequalities in American society and its schools"). Heubert and Hauser, nevertheless, object to the use of high-stakes tests to make determinations regarding graduation on tbe ground that such policies simply "reinforce these inequalities." *Id.*; *see also* White, *supra* note 16, at 95 (describing as separate from the bias rationale the view that test scores, while accurate, are nevertheless unfair because the poor performance of disadvantaged groups is caused by current inequality and/or past discrimination).

302. Perry, *supra* note 155, at 558-59. Significantly, in early testing cases concerning testing of students who began their education under *de jure* segregation, courts have had little trouble enjoining on the grounds of intentional discrimination the use of a test where the test's use would entrench inequality. *See* Anderson v. Banks, 520 F. Supp. 472 (S.D. Ga. 1981); Debra P. v. Turlington, 474 F. Supp. 244 (M.D. Fla. 1979); *see also supra* note 35 and accompanying text (discussing *Anderson* and *Debra P.*). These results indicate that the disparate impact model is not necessary to root out policies that carry forward the legacy of discrimination.

With respect to educational testing, and, indeed, with respect to educational policies in general, notions of harm and benefit are less clear-cut than in the employment context. Although the failure to receive a diploma may cause harm to disadvantaged students, an equal or greater harm may be caused by receiving an inadequate education that does not meet some minimal standards.[303] Accountability proponents argue that the *absence* of strict accountability measures has contributed significantly to the minority achievement gap and current inequalities. From this perspective, high-stakes exams, far from hurting minority students, actually help to break the cycle of failure by allowing policymakers to target scarce resources where they are most needed and by forcing schools to ensure that a greater percentage of students achieve at least minimal levels of competency.[304]

Significantly, while critics of standardized testing in the employment arena rarely recognize any benefit to employees or potential employees from standardized employment tests, critics of educational testing are likely to concede that high-stakes educational exams create both costs *and* benefits for minority students.[305]

Instead of simply creating a "disparate impact"—that is, a purely *adverse* racial or ethnic outcome—such high-stakes educational assessments create a "dual impact" on both minority communities and individual students.[306] Thus, those African-American and Hispanic

---

303. *See GI Forum*, 87 F. Supp. 2d at 674 ("The receipt of an education that does not meet some minimal standards is an adverse impact just as surely as failure to receive a diploma."); Taylor, *supra* note 12 ("[A]rguments [that denying diplomas to students who fail graduation exams will entrench social inequalities] confuse the symhol of the diploma with the level of education it represents. It misses the central point that the real high stakes are imposed not by a test, but by educational neglect. . . . Students who receive a bad education are penalized whether or not they receive a diploma.").

304. Because the Texas accountability system requires the collection of data on the passage rates of each ethnic or racial group within a school district, the Texas plan forces school districts to focus on assisting disadvantaged students. *See* Order for Summary Judgment, *supra* note 208; *see also GI Forum*, 87 F. Supp. 2d at 667.

305. *See, e.g.*, Heubert, *supra* note 34 (noting that both arguments are "plausible" and that, in reality, "high-stakes testing present[s] both opportunities and risks for students of color"). Professor Heubert has also noted that, in the tracking context, courts often split on whether low-track educational placements are "remedial," and thus educationally beneficial, or educationally harmful "dead ends." *See* Heubert, *supra* note 301, at 27.

306. My position regarding education policies that result in disparate racial outcomes complements the view of those scholars who have analyzed the concomitant positive and negative effects of the criminal laws on communities of color. Professor Randall L. Kennedy, for example, has written that facially neutral criminal statutes often harm some members of the African-American community while simultaneously assisting other segments of that same community. Kennedy, *supra* note 143. Professor Kennedy writes that

> [a]lthough blacks subject to relatively heavy punishment for crack possession are burdened by it, their black law-abiding neighbors are presumably helped by it (insofar

students who fail the MCAS or the TAAS will be burdened by the testing policy if they are ultimately unable to earn a high school diploma. At the same time, however, many of the minority students who fail the exam in the first instance will be benefited by the academic assistance and remedial services they receive as a result of the their initial failure—and many of these students may ultimately pass the exam and receive diplomas. In some cases, a student who repeatedly fails the exit exam, but who participates in all of the available remedial classes, may not receive a diploma, but may, nevertheless, benefit from improving his reading and writing skills. Even those African-American and Hispanic students who pass the exam on the first attempt may benefit from the policy if, as a result of the testing regime, they graduate from high school with skills or knowledge they might otherwise not have obtained.

On a broader community level, many primarily black and Hispanic communities benefit from the increased resources that may flow to their schools as a result of testing regimes.[307] And standardized tests establish an incentive for school systems to recruit and promote better teachers.

In addition to these contemporaneous dual impacts, high school exit exams with disparate racial or ethnic results in the short term may actually eliminate or reduce racial or ethnic disparities in the long run, as the beneficial disparate impact has the consequence of reducing the harmful impact over time. Minority communities, and society as a whole, may therefore conclude that the short-term racial or ethnic disparities are acceptable if the testing policy eventually

---

> as the statute deters and punishes drug trafficking that typically takes place in their midst). Although black youngsters who wish to stay out late are burdened by a curfew, blacks who feel more secure because of the curfew are benefited. Although black members of violent gangs are burdened by police crackdowns on such gangs, blacks terrorized by gangs are aided. Although some black women who use illicit drugs harmful to unborn babies are burdened by prosecutions that punish them for this conduct, it is at least plausible to suppose that the deterrent effect of such prosecutions will help other black unborn children.

Id. at 1273-74. Because criminal laws can simultaneously "help" and "harm" members of the same community, Professor Kennedy concludes that—in the absence of evidence of *purposeful* discrimination—courts "should refrain from condemning a state criminal policy as violative of the Equal Protection Clause." (Professor Kennedy leaves open, however, the possibility that policies with a racially disparate impact may be prohibited through regulation or legislation.) *Id.* at 1274; *see also* Gary S. Becker, *Tough Justice Is Saving Our Inner Cities*, BUS. WK., July 17, 2000, at 26 (arguing that minorities are the major beneficiaries of "tough on crime" policies and that the focus on the disproportionate number of blacks and Hispanics in the U.S. prison population ignores the disproportionate number of black and Hispanic victims of crime).

307. *See* Alexakis, *supra* note 17 (explaining that MCAS has helped Massachusetts target resources to disadvantaged communities and noting the related support for high-stakes testing policies in disadvantaged communities).

helps to narrow the achievement gap.[308] Thus, at worst, high-stakes exams can be said to create not a disparate impact but a "dual impact"[309]—both positive and negative—on minority communities and individual minority students.[310]

Because education policies will always affect different segments of minority communities differently, it is inappropriate to generalize and claim that a policy with disparate racial or ethnic outcomes favors the interests of white students over the interests of minority students.[311] To the contrary, where a policy creates both costs *and* benefits for minorities, both courses of action—the implementation of the policy and the failure to institute the policy—can plausibly be described as entrenching inequality.[312] Thus, if, as testing proponents argue, testing is an important means of ensuring that black and Latino students, poor students, and students

---

308. *See GI Forum*, 87 F. Supp. 2d at 671 (holding that the disproportionate short-term effect on minority graduation rates is outweighed by the concomitant long-term societal results).

309. Examples of "dual impact" educational policies abound. For example, the practice of teaching classroom lessons in English necessarily makes it more difficult for immigrant students from non-English speaking countries to compete with native English speakers. Yet, using English also benefits such students by teaching them to navigate the language of the dominant culture and benefits all students and, indeed, society at large, by fostering a society united by a common language.

Consider, also, the "zero tolerance" disciplinary policies promulgated by a number of schools in recent years. Some civil rights activists argue that such policies are discriminatory because they have resulted in a disproportionate number of African-American students being expelled from school. *See* Chaddock, *supra* note 122, at 14 (discussing high-stakes tests and zero-tolerance policies). But a plausible argument can also be made that such policies actually help those who are expelled by giving them a "wake-up call" and forcing them to take responsibility for their lives. Even if one rejects the argument that strict disciplinary measures are a benefit to those who are disciplined, it is difficult to dismiss the argument that such policies may benefit other members of the same demographic community by removing troublemakers from the learning environment. *Cf.* Kennedy, *supra* note 143, at 1255 (arguing that criminal laws may burden black lawbreakers while simultaneously benefiting the majority of the black citzens who are law abiding).

310. A similar argument can be made with respect to public employment. Thus, while minority applicants to the police academy might be disproportionately burdened by a requirement that applicants pass a written exam, the public, and the minority community in particular, might be better served by a better-educated police force. The same is true of teacher certification exams. While black and Latino prospective teachers might be heavily burdened by the certification requirement, students—particularly poor or otherwise disadvantaged students—would be better served by teachers who (at a minimum) understood the material they are required to teach than by those who do not. *See generally* Michael A. Rebell, *Disparate Impact of Teacher Competency Testing on Minorities: Don't Blame the Test-Taker—or the Tests*, 4 YALE L. & POL'Y REV. 375 (1986). As strong as this argument is in the public employment context, it is even stronger in the education setting where dual impact can be felt not only across the community but within individuals.

311. *Cf.* Kennedy, *supra* note 143, at 1253 (advancing similar argument with respect to drug sentencing policy).

312. *Cf. id* (explaining how arguments about black genocide are used both in favor of and against drug legalization).

from inner-city schools receive the same educational opportunities as their white suburban counterparts, then it is the *failure* to hold schools accountable, not the attempt to do so, that causes the more pernicious harm.[313]

Any determination as to whether the costs of high-stakes tests outweigh the benefits to a particular subgroup necessarily involves both empirical predictions and value judgments. The empirical questions involve the extent to which the negative disparate impact (i.e., greater failure rates among black and Latino students) will be reduced over time due to the positive impact (greater accountability, remediation, and resources) or mitigated by other circumstances (i.e., the extent to which even those who fail to graduate have improved their knowledge and skills as a result of the accountability system). The value judgment involves deciding which negative externality—a short-term drop in graduation rates for minority students or the continued failure of schools to educate minority students properly—constitutes the greater evil. This decision—as to who (both at a point in time and over time) should bear the burden of the educational policy—belongs in the realm of local policymaking and not in the realm of the courts or federal administrative agencies that generally consider the viewpoints of only two parties to an adversary process.[314]

The notion that high-stakes tests will inevitably entrench social inequality is by no means obvious or universally accepted. Therefore, the use of federal antidiscrimination laws to halt or interfere with well-intentioned education reforms is inappropriate.

4. Other Rationales

The previous subsection sought to demonstrate that the three most commonly offered rationales for using disparate impact litigation

---

313. *High Standards: Giving All Students a Fair Shot*, *supra* note 10, at 2 (arguing that without tests, inequities between black and white and rich and poor will likely persist beneath the radar).

314. Those who file administrative claims or federal lawsuits over the use of high-stakes tests are usually concerned only with the negative impact the test may have on a narrow class of individuals. Indeed, the nature of the adversarial system necessarily focuses advocates single mindedly on the short-term cost of testing policies borne by minority students who fail the exam and ignore the beneficial impact on minority students who are better educated as a result of the policy. Because the adversarial system (whether played out at an administrative agency or in federal court) inevitably obscures the forest for the trees, it is not an appropriate forum for resolving educational policy disputes. *Cf.* Lambert, *supra* note 27, at 1186 (warning that a regime that permits private plaintiffs to bring suit every time a public policy affects groups of students differently inappropriately gives individuals from the adversely affected group the power to override state policy and legitimate efforts at governmental reform).

against standardized tests (the objective evidence theory, the bias rationale, and the social impact theory) fail to support the model's use in the educational testing context. Other theoretical bases are also unpersuasive.

a. "Statistical Discrimination"

As explained above, statistical discrimination in the employment context occurs when employers, lacking perfect information regarding a job applicant's potential for success on the job, rely upon proxies that are closely correlated with race and only loosely correlated with productivity.[315] Because employers face economic incentives to engage in practices that may result in statistical discrimination,[316] this theory posits that the disparate impact approach is a necessary mechanism to prevent the use of illegitimate proxies and the compounding effects of discrimination in the labor market.[317]

The statistical discrimination rationale is not transferable to the education arena. In the employment context, proponents of the statistical discrimination rationale for disparate impact reject the use of tests or other policies that are only marginally helpful in predicting future "success" on the job. In the educational assessment context, however, this model fails—both because the concept of future "success" is more difficult to define with respect to education than with respect to employment, and because high-stakes educational assessments are not, in fact, used as predictive devices.

Although it may be reasonable to link workplace "success" with productivity, in the education arena productivity has no appropriate analog. The concept of educational success cannot be cabined by economic baselines and, indeed, can be defined any number of ways by any number of entities. Legal scholars, primary and secondary school teachers, college professors, parents, and students themselves may all have different conceptions of what counts as educational "success" and hence what measure of future performance is appropriate in determining the predictive validity of a test.[318]

---

315. *See supra* notes 157-59 and accompanying text.

316. *See supra* Part III.B.

317. As articulated by Professor Steven Willborn, the "statistical discrimination" theory tracks the first prong of Professor Owen Fiss's "functional equivalence" theory, discussed *supra* notes 160-63 and *infra* notes 320-25, in that it views with suspicion employment criteria which are poor predictors of productivity and which are strongly correlated with race.

318. *See, e.g.*, Kidder, *supra* note 19, at 190-95 (arguing that when measuring the validity of the LSAT, the appropriate analog to productivity is not success in the first year, but completion of the degree program or passage of the bar exam).

How, then, should courts determine whether a test accurately predicts educational success? Should courts look to the correlation between performance on the test and performance at the next educational level as measured by grades? (If so, should courts compare test scores with cumulative grades or only with grades in courses substantively related to the test itself?) Or should courts look to see whether test scores correlate with "success" in the "real world" as measured by future earnings? The lack of consensus as to how precisely we should measure future success in this context renders the disparate impact model highly problematic in the educational setting. Moreover, while employers generally control the workplace and prescribe the tasks to be performed by various employees, secondary schools and colleges exert limited control over a student's academic choices, making future "success" impossible to predict accurately. The statistical discrimination rationale may therefore provide a useful rationale for the disparate impact model in the employment context, but, where, as in the case of high-stakes educational assessments, the definition of future success is both disputed and variable, the approach loses its utility.

More importantly, unlike standardized exams utilized by employers to select employees, high-stakes educational tests that are used to make promotion and graduation decisions do not in fact aim to *predict* anything. Rather, they are used to measure levels of achievement and to certify that students at each educational level have gained the minimal skills and knowledge that educators and policymakers have determined to be appropriate. In view of this assessment purpose, it is inappropriate to use *future* success (however defined) as a benchmark for validity in educational testing. Indeed, in an educational setting, it is entirely appropriate to measure general knowledge and skills *for their own sake*, even if an examinee's score on the challenged test correlates with nothing else. In other words, in the education setting, there is only one relevant question: Does the test measure that which it purports to measure? If so, the fact that the test fails to correlate with success at some other point in time on some other benchmark is simply irrelevant. Because the failure to predict future performance should not render an educational assessment invalid, the statistical discrimination rationale cannot properly justify the use of the disparate impact model in cases involving educational tests used solely as a means of assessment.

b. Functional Equivalence

Under the functional equivalence theory advanced by Professor Owen Fiss, the disparate impact model is appropriately used where the test in question proves to be a poor or weak predictor of productivity *and* the criterion that the test measures is one over which examinees lack individual control.[319] As explained above with respect to the statistical discrimination rationale, because educational success is easily manipulated and difficult to quantify, and because standardized promotion or graduation exams are generally used to measure prior achievement, not future performance, the functional equivalence rationale is not an appropriate basis on which to justify the use of disparate impact theory with respect to high-stakes educational assessments.[320]

Even putting aside these problems with the first prong of the functional equivalence rationale, additional problems with the second prong of the functional equivalence analysis—individual control—would render the disparate impact model inappropriate with respect to educational testing. This is because knowledge and academic skills—unlike inheritance (as in the nepotism example used by Professor Owen Fiss)[321]—are in large part within the control of the students themselves.

Writing in 1971, Professor Fiss argued that, in some places, intentional racial discrimination in the public education system might provide an explanation for poor test performance by minorities.[322] In such circumstances, a person's performance on a general intelligence battery or achievement test might be said to be outside of his or her own control. For instance, where, as in *Griggs*,[323] a black job

319. *See supra* Part III.B and notes 162-65 and accompanying text.

320. Some states have proposed making students who do not pass the state-mandated high school exit exam ineligible for admission to that state's public colleges and universities. *See, e.g.*, Scott S. Greenberger, *MCAS Required for State Colleges: Local Certificates Not Acceptable*, BOSTON GLOBE, Sept. 28, 2000, at B1. This use of test results does not change the analysis because the tests are still not intended to predict anything; rather, they are intended to certify mastery of a previous course of study.

Even where high-stakes tests such as the SAT or LSAT are used to predict future performance, the "statistical discrimination" and "functional equivalence" rationales fail to justify the use of disparate impact theory to challenge such tests. In the employment setting, these rationales rely on the fact that some exams (and other selection devices) only weakly correlate with success as measured by productivity. But, as explained previously, *see supra* Part V.A.4.a, the concept of "educational success" cannot be cabined by measurable productivity or any other economic baselines.

321. Fiss, *supra* note 136, at 303.

322. *Id.*

323. Griggs v. Duke Power Co., 401 U.S. at 424, 427 (1971).

applicant's lack of a high school diploma or poor performance on an employment examination was due to systemic intentional discrimination in the educational system, the use of such a hiring criterion would be appropriately subject to rigorous disparate impact analysis.

According to Fiss, it is appropriate to infer a lack of individual control over one's ability to learn if the racial discrimination in the educational system is of "a unique, systematic character"—as in the case of *de jure* segregation.[324] Yet this inference seems implausible in the year 2002. Today our educational institutions do not practice the sort of "unique, systematic" discrimination of which Fisk wrote.[325] And it would be outrageous, if not outright racist, to conclude that black and Latino students lack all control over their ability to learn. Indeed, the implementation of high-stakes tests rests on the premise that academic achievement can be improved through the efforts of teachers and students. To insist, instead, that black and Hispanic students are uniquely unable to exert some control over their ability to learn would be an absurd claim of inability and inferiority and would do nothing to ameliorate the achievement gap that now exists.[326]

To be sure, we can expect that some students who hail from impoverished family backgrounds and who attend schools in high-poverty areas will have a more difficult time passing the MCAS or the TAAS than will students from wealthier suburban districts. But, in this regard, poor white students will be as disadvantaged as poor black students; as a general matter, black students attending suburban schools will be as privileged as white students attending suburban schools. To the extent that economically disadvantaged students "lack individual control" over their ability to pass standardized tests, this is not due to their race, but rather to the fact that they have received an inadequate education from a dysfunctional school system.

c. Strong Theories of Disparate Impact

As explained previously, "strong" justifications for the disparate impact model of employment discrimination presume that

---

324. Fiss, *supra* note 136, at 299.

325. *But see* Noguera & Akom, *supra* note 290, at 29 (arguing that racial and ethnic discrimination infects all aspects of the educational system).

326. Professor Fiss does not argue that all students who attend inferior schools lack individual control over their ability to learn, only that those who have suffered intentional discrimination in the school system lack control over their ability to learn. Fiss, *supra* note 136, at 303.

the model can compensate minorities for past injustices by forcing employers to distribute jobs on the basis of race, or at least in proportion to the racial makeup of the labor market.[327] But the notion of compensatory justice—that is, the idea that societal goods, such as jobs, should be distributed in a way that compensates for past injustice—makes little sense with respect to educational testing. The notion that a black or Hispanic student deserves a high school diploma, not because he may have earned it, but because society owes it to him, makes the concept of a diploma (which, after all, is supposed to be a certificate of mastery) meaningless. Proponents of the compensatory justice rationale might argue that individuals who are denied a high school diploma will be disadvantaged in the job market and, thus, the awarding of a diploma is equally as important as the awarding an actual job. There is little empirical evidence, however, that those students who are unable to obtain a high school diploma are worse off after leaving school than those students who obtain a diploma but who, nevertheless, have failed to master basic skills. It is a lack of skills—not the absence of a piece of paper—that prevents individuals from succeeding in the workplace. Giving someone a worthless diploma cannot, therefore, provide compensation for previous societal wrongs.

B. Normative Arguments Against Applying the Disparate Impact Model to Education

Part V.A argued that none of the proffered theoretical bases for the disparate impact model justifies its use with respect to educational testing. This section offers other reasons why the disparate impact model should not be used to thwart high-stakes educational tests. On a doctrinal level, the concept of "business necessity" is not easily transmuted into a coherent concept of "educational necessity." At a more basic level, the underlying enterprises of businesses (on the one hand) and schools (on the other) are very different. Given the different organizing principles of these two types of activity, the use of testing in the educational context cannot be directly analogized to testing in the employment context. Although the disparate impact doctrine might be coherently applied to private sector employment, the doctrine cannot be coherently applied to educational testing.

327. *See supra* text accompanying note 154.

1. The Failure of the "Educational Necessity" Standard

By what objective standard are courts supposed to evaluate educational testing policies? We are told by proponents of the disparate impact model that courts should apply the standard of "educational necessity,"[328] although this standard is ambiguous, if not opaque. Under Title VII, the necessity of an employer's policy is generally judged by its relationship to productivity, profit margins, or other economic baselines—all of which are arguably quantifiable. These guideposts, however, cannot help ascertain the "necessity" of an educational policy or practice. That is because, as difficult as it might be to identify accurately the connection between a specific employment policy and productivity, it is even more difficult to identify the connection between a particular educational practice (e.g., testing) and an educational institution's purported goal (e.g., student motivation). And it is nearly impossible to identify accurately the connection between a particular educational goal (e.g., mastery of ninth-grade level mathematics) and "productive citizenship."

In the employment setting, proponents of the disparate impact model have argued that even where a test is generally consistent with "business necessity," a high cutoff score that results in disproportionate demographic outcomes should be considered unlawful if the threshold score does not marginally increase productivity over other potential cutoff scores.[329] Once again, however, this argument is meaningless when educational achievement, not productivity, is the issue. Unlike the skills measured by employment tests, which are supposed to correlate with success on the job, the domains measured by educational tests—levels of achievement—are not necessarily intended to correlate with anything other than mastery of the areas tested and current academic skills. Hence, the appropriate cutoff score on a test in the school setting cannot be determined by reference to factors other than the educational objectives of the curriculum itself. Educational policymakers should have the right to set the bar for academic achievement as high (or as low) as they deem reasonable. The selection of a cutoff score in this context is entirely a policy choice, and thus—by definition—cannot be an "educational *necessity*." Any attempt to apply the "educational necessity" standard to cutoff scores in the education setting

328. NAT'L RESEARCH COUNCIL, *supra* note 15, at 59.

329. *See, e.g.*, Willborn, *supra* note 18, at 821.

transforms an antidiscrimination law into a device for establishing the educational standards themselves.[330]

In the absence of clear and objective criteria by which to judge the "necessity" of a particular educational policy, courts have, not surprisingly, developed a variety of approaches for measuring a test's legality.[331] Some courts have required proof that the objective underlying the challenged policy is "substantially legitimate,"[332] while others require proof that the educational objective is "legitimate, important, and integral" to the institution's mission.[333]

Even when courts accept the legitimacy of the educational objectives underlying a particular testing policy or practice, there is no clear or obvious legal standard that courts can apply to decide whether the policy or practice is sufficiently tailored to meet these objectives. Some courts have adopted a stringent standard, while others have been more permissive. Some courts require proof that the test (or other policy or practice) bears a "manifest relationship" to the educational goal.[334] At the other extreme, at least one court has held that a defendant fails to meet its burden of proving educational necessity "only if the evidence reflects that the test falls so far below acceptable and reasonable minimum standards that the test could not be reasonably understood to do what it purports to do."[335]

---

330. *Cf.* Groves v. Ala. State Bd. of Educ., 776 F. Supp. 1518 (M.D. Ala. 1991). In *Groves*, a federal trial court held unlawful Alabama's requirement that students seeking admission to undergraduate teaching programs at state universities have received a certain minimum score on the ACT. *Id.* at 1519. The *Groves* court reasoned that the passing score set by the state was insufficiently correlated with minimal competence as a teacher. In so holding, the court inappropriately applied the productivity theory used in the employment setting to an educational test, the goal of which was to measure general knowledge and skills, not productivity or competence. *See* Maltz, *supra* note 178, at 360-61 (noting that in the credit context, application of the disparate impact model would turn antidiscrimination laws into vehicles for controlling profit margins).

331. *See* Elul, *supra* note 16, at 518 (noting that courts have reached mixed results in disparate impact challenges to educational policies); *see also* Heubert, *supra* note 301, at 27 (describing the varying approaches to "educational necessity" adopted by different courts).

332. *See* Cureton v. NCAA, 37 F. Supp. 2d 687 (E.D. Pa.) (rejecting as neither legitimate nor substantial defendant's proffered educational objective), *rev'd on other grounds*, 198 F.3d 107 (3d Cir. 1999).

333. *See* Elston v. Talladega County Bd. of Educ., 997 F.2d 1394, 1413 (11th Cir. 1993).

334. *See, e.g.*, Ga. State Conference Branches of NAACP v. Georgia, 775 F.2d 1403, 1418 (11th Cir. 1985) (requiring a "manifest demonstrable relationship"); Larry P. v. Riles, 793 F.2d 969, 982 n.9 (9th Cir. 1984) (holding that test must bear a "manifest relationship" to the educational goal); Sharif v. N.Y. State Educ. Dep't, 709 F. Supp. 345, 362 (S.D.N.Y. 1989) (requiring a "manifest relationship").

335. Richardson v. Lamar County Bd. of Educ., 729 F. Supp. 806, 822-23, 825 (M.D. Ala. 1989) (holding that Alabama's teacher certification test was not educationally necessary under Title VII because it failed to measure that which it claimed to measure).

In sum, the "educational necessity" standard provides a fact-finding court with broad latitude in the exercise of its discretion—from a heightened deference to educational policy choices to virtually no deference to such policies. Yet the goals and policies adopted by educational institutions are based on a variety values that are difficult to quantify, and it is manifestly within the province of our political and democratically elected branches of government—not the judiciary—to determine what those values are or what those values should be.[336] Recognizing that courts possess neither the expertise nor the political mandate to support unwarranted intrusions into education policy, the U.S. Supreme Court has cautioned that federal courts should not overrule state or local education policy absent a clear-cut violation of the law.[337] The application of the "educational necessity" standard to public education in disparate impact litigation flies in the face of the Court's instruction and invites courts to usurp the role of state and local governments and second-guess their efforts to establish educational goals and methods for achieving them.

Moreover, even if we could be certain that courts could resist the temptation to second-guess educational policymakers, use of the disparate impact model with respect to educational testing would be inappropriate. Disparate impact cases, by their nature, do not involve clear-cut violations of the law—those cases are brought under the disparate treatment model of discrimination. Accordingly, courts that adhere to the Supreme Court's admonition to defer to local educational decisions absent clear legal violations will tend almost invariably to uphold educational policies from disparate impact challenges, rendering the model a useless exercise and a waste of scarce judicial, administrative, and educational resources.

2. Recent Efforts to Refine and Codify the Definition of "Educational Necessity"

Some commentators have claimed that the concept of "educational necessity" is unworkable precisely because Congress and the U.S. Department of Education have failed to define the term by

336. *See* Maltz, *supra* note 178, at 359 (making a similar argument regarding the application of the disparate impact model to zoning laws).

337. *See* San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 42 (1973) (stating that the judiciary's lack of specialized knowledge in the area of education counsels against its interference with informed educational judgments made at the local level); *see also* GI Forum v. Tex. Educ. Agency, 87 F. Supp. 2667, 683 (W.D. Tex. 2000); *cf.* Epperson v. Arkansas, 393 U.S. 97, 104 (1968) (noting in the context of the First Amendment that "[j]udicial interposition in the operation of the puhlic school system of the Nation raises problems requiring care and restraint").

statute or regulation.[338] They argue that the creation of legally binding statutory or regulatory standards would ensure that lower courts respond consistently in cases involving high-stakes educational tests.[339] Professor Jay Heubert, for example, recommends that the U.S. Department of Education adopt standards issued by the American Educational Research Association in its *Standards for Educational and Psychological Testing*.[340] The adoption of such standards would require courts to consider, among other things: (1) the "[c]onsequences resulting from the uses of the test, both intended and unintended" (including the dropout rate among students who do not pass the test, the narrowing of the curriculum, and the use of other practices designed to raise test scores rather than teach);[341] (2) the consistency between test scores and other performance indicators (such as grades);[342] and (3) the "relationship among particular test scores, the instructional programs, and desired student outcomes [including] . . . evidence about whether students, in fact, benefit from the differential [or remedial] instruction."[343] The *Resource Guide* issued by the Office for Civil Rights in the Clinton Administration's Department of Education takes a similar approach, arguing that the "educational necessity" of a high-stakes test should be assessed by examining test validity, reliability, and fairness, as well as the institutional goals and objectives of the school, the educational consequences to students, and the relationship of the educational institution to the student.[344]

Yet such approaches hardly cabin the concept of educational necessity. To the contrary, by demanding consideration of a host of inherently subjective factors (including goals, relationships, and fairness) and factors that are ordinarily left to the discretion of local policymakers (such as a policy's social consequences), the proposed standards greatly complicate the task of courts, while holding no promise of greater clarity or consistency in the law.

Even the scientific concept of test "validity" has become confused and the subject of much disagreement and, thus, an

338. Heubert, *supra* note 301, at 27 (complaining that "neither Title VI nor its current regulations provide standards that would help define the circumstances under which it is 'educationally necessary' to use a high-stakes test that has a disproportionate, adverse impact by race or national origin").

339. *See, e.g., id.*

340. NAT'L RESEARCH COUNCIL, *supra* note 15, at 266; Heubert, *supra* note 301, at 27.

341. AM. EDUC. RESEARCH ASS'N ET AL., STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING, Standard 13.1, at 145 (1999) [hereinafter JOINT STANDARDS].

342. JOINT STANDARDS, *supra* note 341, Standard 13.3, at 146.

343. *Id.* Standard 13.9, at 147.

344. RESOURCE GUIDE, *supra* note 220, at 34, 39.

improper basis for determining the necessity of an educational test. Recall that "validity" generally refers to a test's accuracy—that is, its ability to predict accurately future performance (as in the case of the SAT) or its ability to measure accurately the knowledge and skill level that the test purports to measure (as in the case of educational assessments).[345] In recent years, however, other less rigorous theories of test validity have infiltrated the field of educational measurement. For example, the "instructional validity" theory asserts that a test can be valid only if it measures students' knowledge of curriculum actually taught.[346] As Michael Rebell has noted elsewhere, it is impossible to use instructional validity as a legal benchmark in the context of district-wide or statewide assessments because any attempt to measure precisely what students across schools and classrooms have *actually* been taught is impossible. Rebell explains:

> In order to establish that every student has had a fair opportunity to learn each of the many subjects covered by an examination, the practices of every school district, perhaps even of every school and classroom in the state over the twelve year span of a public school education would have to be analyzed [by the court].[347]

Moreover, because the experts disagree as to how "instructional validity" can be measured with respect to a statewide testing program, this species of validity is particularly subject to manipulation.[348]

Equally malleable[349] and subject to manipulation is the "consequentialist" approach to test validity, which maintains that a

---

345. *See supra* Part V.A.2.a. As explained previously, in the case of educational assessments, predictive validity is not particularly important because such tests are not designed to predict anything, but rather seek to measure achievement. *Id.*

346. Rebell, *supra* note 310, at 386. The concept of "instructional validity" was first given legal weight by the Fifth Circuit in *Debra P. v. Turlington*, the lawsuit challenging Florida's minimum competency exam. 644 F.2d 397, 405 (5th Cir. 1981). On appeal, the court held that, in order to satisfy constitutional requirements of due process, the state must prove that the graduation exam "covered things actually covered in the classroom." *Id.*

347. Rebell, *supra* note 310, at 387; *see also* Anderson v. Banks, 540 F. Supp. 472, 765 (S.D. Ga. 1981) ("[T]o require school officials to produce testimony that every teacher finished every lesson and assigned every problem in the curriculum would impose a paralyzing burden on school authorities."); *cf.* Debra P. v. Turlington, 564 F. Supp. 177, 186 (M.D. Fla. 1983) (ruling that, as long as the curriculum includes basic objectives of which teachers are aware, the court will not intervene in educational policy under the guise of the Due Process Clause).

348. Rebell, *supra* note 310, at 387.

349. As evidence that the concept of validity is too pliable to provide a legal rule for determining the educational necessity of a test, one need look no further than an essay written by Arthur L. Coleman, a former Clinton Administration Deputy Assistant Secretary for Civil Rights at the U.S. Department of Education. Coleman approvingly notes that the concept of validity is fluid and ever-changing. He writes that "a conclusion that a test was valid yesterday may yield to information suggesting that it may be invalid tomorrow." Coleman, *supra* note 19, at 103 (citing Samuel Messick, *Validity*, *in* EDUCATIONAL MEASUREMENT 13, 13 (Robert Linn ed., 3d ed. 1989) for the proposition that validity is an evolving property and validation is a continuing process).

determination as to whether a test is valid cannot be made without considering whether the use of a test is socially just.[350] This politically charged, scientifically dubious, approach has recently ascended in importance in the psychometric literature.[351] Consequentialist theories of test validity, however, conflate the concept of validity with that of raw adverse impact, and are thus circular. Upon close examination, consequentialist arguments can be distilled into the following set of propositions:

1) A test should not survive disparate impact scrutiny if it is not valid.
2) In order for a test to be valid, the test's use must be socially just.
3) In order for the use of a test to be socially just, it must not adversely burden particular minority groups.
4) Where the result of test failure is denial of some educational benefit (such as a diploma or advancement to the next grade), a lower mean score for minority students than for white students constitutes proof that the lower scoring minority group will be adversely burdened by the test.
5) Because minority groups are disproportionately burdened by the consequences of the test, use of the test is unjust, invalid, unnecessary, and, therefore, illegal.

In other words, under the consequentialists' circular approach, where a test has high-stakes consequences, disparate racial outcomes are per se illegal. This approach thus conflates the analytically distinct concepts of technical accuracy and social impact. Accordingly,

---

350. Linda F. Wightman, *An Examination of Sex Differences in LSAT Scores from the Perspective of Social Consequences*, 11 APPLIED MEASUREMENT IN EDUC. 255, 273 (1998) (arguing that the social consequences of standardized testing must be incorporated into any assessment of a test's validity). *See generally* Messick, *supra* note 349, at 13 (arguing that validity cannot be divorced from social outcomes).

351. *See* CAMILLI & SHEPARD, *supra* note 272, at 154 (warning that "test developers should never claim on the basis of [technical measures of test bias] that a test has been guaranteed free of bias and valid for all possible uses" without considering a test's sociopolitical impact). According to this view, validity can be determined only by considering the social and political consequences of test use. *See id.* at 153 (contending that questions regarding whether a test is accurate, reliable, and technically bias free cannot answer the larger normative question about whether test use is reasonable and fair).

The Clinton Administration's *Resource Guide* implicitly adopts the consequentialist approach, stating that, in addition to considering whether a test is accurate, reliable, and bias-free, courts evaluating a test's scientific validity must also consider: (1) the purpose for which the test is used; (2) whether an individual's score on the test is the sole criterion for the educational decision; (3) the nature and quality of the opportunity provided to students to master required content; and (4) the educational bases for establishing passing or cutoff scores. RESOURCE GUIDE, *supra* note 220.

statutorily defining "educational necessity" to mean "validity"—without limiting the concept of validity to scientific accuracy (as opposed to fairness)—would serve only to confuse the law and to eviscerate a defendant school system's affirmative defense and adopt, through the back door, a per se ban on high-stakes tests with disparate racial effects.

3. Adversarial Versus Cooperationist Models

An additional fundamental difficulty with applying to educational tests the disparate impact model developed in the employment context is that employers and educational institutions, serving fundamentally different interests, are organized around different operational models. While employers act out of self-interest, public educational institutions operate, by definition, in the public interest.[352]

The adversarial model of employment relations posits a tension between the objectives of management—which seeks to maximize profits—and the needs of the employees—for whom profit maximization is generally secondary to wage maximization and other aspects of individual self-interest.[353] For example, the National Labor Relations Act ("NLRA") is premised on the idea that there is "a fundamental conflict of interest between labor and management—a chasm—that is thought to require structural guarantees to keep separate their respective spheres of influence."[354] In order to preserve

352. This is true even when we consider the role of private schools. The decision by a private school to require students to pass a test before graduating, while not a public action, is nevertheless closer to the realm of policy than to the realm of production, and thus the interests of public and private schools are not analytically distinct in this regard.

353. *See, e.g.*, Charles B. Craver, *The Vitality of the American Labor Movement in the Twenty-First Century*, 1983 U. ILL. L. REV. 633, 673-74 (describing the view that conflict of interest is inherent in the employment relationship).

354. Samuel Estreicher, *Labor Law Reform in a World of Competitive Product Markets*, 69 CHI.-KENT L. REV. 3, 11 (1993); *see also* Packard Motor Co. v. NLRB, 330 U.S. 485, 494-95 (1947) (Douglas, J., dissenting) (stating that the National Labor Relations Act is based on a strict labor-management dichotomy). Recently, some scholars have begun to question the assumption that the NLRA is a strictly adversarial statute that discourages cooperation between management and labor. *See, e.g.*, Mark Barenberg, *The Political Economy of the Wagner Act: Power, Symbol, and Workplace Cooperation*, 106 HARV. L. REV. 1381, 1492 (1993) (arguing that the adversarial interpretation of the NLRA evolved over time and is not inherent in the statute); Thomas C. Kohler, *Models of Worker Participation: The Uncertain Significance of Section 8(a)(2)*, 27 B.C. L. REV. 499, 549-50 (1986) (same). Nevertheless, the prevailing view remains that the drafters of the Act intended to create an adversarial framework. *See* Marion Crain & Ken Matheny, *"Labor's Divided Ranks": Privilege and the United Front Ideology*, 84 CORNELL L. REV. 1542, 1555 (1999); Shannon Browne, Note, *Labor-Management Teams: A Panacea for American Businesses or the Rebirth of a Laborer's Nightmare?*, 58 OHIO ST. L.J. 241 (1997) (finding continued historical, statutory, and policy support for the adversarial model).

these separate spheres, the NLRA creates a legal wall between managers and employees. It thus prohibits employers from interfering with or in any way assisting a labor organization, prohibits supervisors and managers from unionizing, and defines the scope of mandatory bargaining strictly so as to keep the domains of labor and management separate and distinct.[355]

Judicial interpretations of the NLRA likewise reflect adversarial assumptions about the workplace. Thus, in *NLRB v. Insurance Agents' International Union*,[356] the Supreme Court noted that "[employers and employees] still proceed from the contrary and to an extent antagonistic viewpoints and concepts of self-interest."[357] The NLRA and judicial interpretations of the NLRA, therefore, not only reflect adversarial assumptions about employer-employee relations, they enshrine those assumptions firmly in American law.[358]

Title VII is also premised on the assumption that an employer's unfettered self-interest will often cause it to act in ways that disadvantage certain employees or potential employees. According to economic models of employment discrimination, a rational employer may discriminate, even in the absence of racial animus, if its customers or clients prefer that it do so.[359] Similarly, a rational employer might discriminate if it concludes that a potential employee's status is a useful proxy for job qualifications.[360] Because

355. 29 U.S.C. § 152(a)(5) (2000); *id.* § 158(a)(2); Estreicher, *supra* note 354, at 11; *see also* Note, *Collective Bargaining as an Industrial System: An Argument Against Judicial Revision of Section 8(a)(2) of the National Labor Relations Act*, 96 HARV. L. REV. 1662, 1679-80 (1983) (arguing that the National Labor Relations Act is premised on the dicbotomy between labor and management).

356. 361 U.S. 477, 488 (1960).

357. *Id.*

358. *See* Crain & Matheny, *supra* note 354, at 1555 ("The NLRA cements the adversarial model by mandating a united front on labor's side of the table: it obligates the employer to bargain with the duly certified labor organization representing a majority of its employees.").

359. Economists refer to this as "taste-based discrimination" since it is caused by a customer's preference (or taste) for dealing with people from certain backgrounds. *See* Strauss, *supra* note 136, at 1621 (citing GARY BECKER, THE ECONOMICS OF DISCRIMINATION (2d ed. 1971)).

360. Economists refer to this as "statistical discrimination." *See* Strauss, *supra* note 136, at 1621; Willborn, *supra* note 18, at 821. For example, assume that women, on average, take more time out of the paid labor market than men for the purpose of bearing and raising children. If this is so, it might make economic sense for a company concerned with the continuity of its workforce and reducing employee absences to hire only men. In other words, in the absence of Title VII, many employers might "rationally" decide to discriminate against women of child-bearing age. Yet Title VII rightly forbids this. Similarly, it might make good business sense for a company to dismiss employees who have brought lawsuits against them. Employees embroiled in ongoing litigation against their employers may be insubordinate, may encourage other workers to sue the company, and may become less productive employees. Yet Title VII prohibits

employers lack perfect information to evaluate the productivity of prospective employees, it is argued that they may use race or ethnicity as a proxy for other traits even if they harbor no dislike for members of a particular race.[361] Fair employment laws are therefore necessary to prevent employers from taking those actions that, while economically rational, are socially undesirable.

In sum, both Title VII and the NLRA embody the notion that employers and employees operate from separate vantage points and that, absent government intervention, employers will often act in ways that fundamentally disadvantage employees (or prospective employees), particularly racial and ethnic minorities and women.

Viewed through this lens, the use of disparate impact litigation to challenge employment tests with a disparate racial or ethnic outcome may appear to make sense. In the business context, tests are used as a screening device. An employer may be able to hire or promote only a small subset of the workers who apply for a particular job, and so may rely on tests to streamline the pool of candidates. In doing so, the employer is not attempting to help the applicants; rather, it is serving its own institutional interests in efficiently sorting the applicants so as to be able to differentiate between them. Particularly where the number of candidates far exceeds the number of available positions, an employer may set a particularly high cutoff score on an employment test as a threshold to eliminate all but a small group of top qualifiers. In this situation, the argument in favor of the disparate impact approach is that the use of a cutoff score unfairly eliminates from consideration a disproportionate number of candidates from particular demographic groups who did not meet the threshold score but who, nevertheless, could perform the functions of the job adequately. At the same time, use of the cutoff score may only marginally advance the employer's legitimate interest in productivity.[362]

The assumption underlying the use of the disparate impact model in the employment setting is that, at some point, an employee's (or potential employee's) interest in obtaining a job that he is capable

---

employers from retaliating—however rationally—against employees who exercise their statutory rights.

361. *See, e.g.*, Strauss, *supra* note 136, at 1621.

362. Some commentators have argued that the willingness of courts to scrutinize the cutoff score for an otherwise valid test unfairly limits an employer's ability to select "the best and the brightest" and leads to forced mediocrity. *See, e.g.*, Brian W. Jones, *The U.S. Department of Education and Two Court Decisions Probe the Limits of "Disparate Impact" Theory*, 3 CIVIL RIGHTS PRACTICE GROUP NEWSL. (The Federalist Society for Law and Public Policy, Washington, D.C.), *available at* http://www.fed-soc.org/Publications/practicegroupnewsletters/civil-rights/decisionscivv3i2.htm (last visited June 28, 2002).

of performing outweighs an employer's interest in any incremental increase in productivity, particularly when identifiable minority groups are disproportionately affected. Thus, where an employment policy or practice results in harm to a particular demographic group of workers (or prospective workers) an employer may not continue that practice absent proof that the practice is job related and consistent with business necessity.[363]

The adversarial model on which American labor and employment law are based clearly do not apply to education. State boards of education, local school boards, principals, teachers, and students all share a common goal: the education of the children entrusted to the school system. Unlike businesses that operate out of self-interest, public schools are intended to serve the public interest and are judged by their ability to educate all students safely and effectively. Indeed, schools and students are not so much adversaries as they are partners in a common enterprise. And, unlike employment tests, which are adopted in order to serve the interests of the employer, legitimate educational assessments are adopted in order to improve educational opportunities for the students. Educational assessments are, therefore, a form of pedagogic practice. And while educators and policymakers may disagree about which practices best achieve the objective of educating students, the overarching purpose—education—is never in doubt. Thus, if educational policymakers adopt testing policies to ensure that all students graduate from school with a mastery of certain subjects and skills, the law should not distrust such tests on the basis of disproportionate demographic outcomes, unless, of course, circumstances suggest an *intent* to discriminate.

4. Rewarding Willful Blindness

Ironically, application of the disparate impact model to educational testing would mean that states or school districts that implement high-stakes testing as part of an effort to equalize educational opportunities would open themselves to the risk of federal lawsuits or investigations by the federal government, while those educational institutions that do nothing to improve the educational opportunities of disadvantaged youth would remain immune from costly lawsuits and government inquiries.[364] In other words, the use of

363. 42 U.S.C. § 2000e-2(k) (2000); Griggs v. Duke Power Co., 401 U.S. 424 (1971).

364. *See* James S. Liebman & Charles F. Sabel, *A Public Labaratory Dewey Barely Imagined: The Emerging Model of School Governance and Legal Reform*, N.Y.U. REV. L. & SOC. CHANGE (forthcoming 2002) (unpublished manuscript at 104), *available at* http://www.law.colum-

the disparate impact model with respect to educational testing rewards willful blindness and punishes those who dare to announce that the emperor is wearing no clothes.

Some scholars have argued that the solution to this problem is not to eliminate the use of Title VI disparate impact claims altogether, but rather to alter the presumptions governing such cases. Thus, rather than view testing regimes with disparate racial consequences as inherently suspect, Professors James Liebman and Charles Sabel argue that courts and administrative agencies should view institutions that *fail* to implement accountability regimes as presumptively in violation of Title VI.[365]

While this approach is certainly better than the current model, it fails to address—and indeed highlights—the central problem with the disparate impact model as applied to education policy: namely, that there is no obvious answer to the question of which negative impact—the failure of schools to test students and provide remedial assistance where necessary or the failure of schools to award diplomas to large numbers of minority students—produces the greater evil. By simply taking an opposing normative viewpoint, the Liebman and Sabel approach fails to resolve the inherent tension in Title VI disparate impact policy, while still requiring that the federal government usurp local decisionmaking.

VI. CONCLUSION

In this Article, I have sought to demonstrate that the disparate impact model of discrimination should not be applied to challenges to educational tests under Title VI of the Civil Rights Act of 1964. Application of the disparate impact model to educational testing fails to achieve any of the stated goals of the disparate impact doctrine and is unsustainable on grounds of public policy.

After decades of poor academic achievement in our public schools—particularly those that are urban and those that are poor—policymakers have finally begun serious efforts at educational reform through accountability. A critical component of such reform efforts is the testing of primary and secondary school students and the attaching of consequences to those test results. High-stakes testing

---

bia.edu/sabel/papers.htm (last visited May 14, 2002) ("Districts that scrutinize themselves carefully in the interest of the new reforms are disproportionately likely to become the target of Title VI suits. Conversely, obstinate ignorance of the details of local conditions, manifest, for example, in the unwillingness to collect data or set standards, can immunize schools and districts from Title VI liability." (footnote omitted)).

365. *See id.*

regimes seek not only to obtain critical information in order to help policymakers target corrective and remedial services, but also to create incentives for success and consequences for failure that will motivate teaching and learning.

The primary objectives of the reform movement are the improvement of student achievement across the board and the equalization of educational opportunities across class and race. Ironically, however, reform efforts that include regular student assessments with actual consequences for repeated failure have been greeted by threats of legal action from activists and commentators who claim that standardized academic tests "discriminate" against racial and ethnic minorities and who seek, in the name of minority students, to use federal antidiscrimination laws to block creative and well-intended efforts at reform. While this approach may in the short term help more African-American and Latino students to obtain diplomas or advance to the next educational level, the use of federal antidiscrimination laws to block education reform will do nothing to help improve educational opportunities for these students or the next generation of minority youth.

The decision to adopt, reject, discontinue, or modify an educational testing policy should, of course, be based on careful consideration of the educational and social science literature, empirical evidence regarding the success or failure of the policy or other like policies, evidence regarding the policy's social impact, and political realities.[366] These considerations are best weighed by legislative bodies, state boards of education, local school boards, and state or local regulatory agencies that are institutionally positioned to balance competing social and political interests and to assess the relevance of social science research.

The adversary system is ill-suited to these tasks. Asking courts to apply complicated psychometric concepts and social science principles will inevitably create inconsistency and confusion in the law. Perhaps most importantly, however, the use of disparate impact litigation to resolve concerns regarding high-stakes testing requires courts to choose between two compelling notions of justice, and thus, in the end, to set educational policy, which is not their role in our

366. It is here that evidence of disparate impact is most relevant. That is, policymakers may very well determine that establishing a graduation requirement that results in large numbers of Latino and African-American students failing to graduate or be promoted is unwise public policy. But this is not a problem that our antidiscrimination laws can or should resolve. Absent a showing that the test is being purposefully used to subjugate minorities, local education officials should be able to weigh the costs and benefits of educational policy and determine, without judicial interference, whether a testing regime should be implemented or maintained.

constitutional regime. Accordingly, courts should reject attempts to thwart education reform under the guise of enforcing our nation's antidiscrimination laws.

For the same reasons, Congress should reject calls to codify the disparate impact model of discrimination in the text of Title VI. Should Congress nevertheless decide to amend the statute in the wake of the Supreme Court's *Sandoval* decision, it should exempt school accountability efforts from any provision authorizing disparate impact lawsuits. Such an amendment would create a safe harbor for states and school districts that attempt to equalize educational opportunities and improve educational outcomes through accountability regimes.

To be sure, the current wave of reforms may not succeed in improving student achievement or closing the minority achievement gap. Perhaps even where such policies succeed, policymakers may, nevertheless, decide that the short-term social costs are too high to continue the testing policy. But until we can agree not to ask our nation's courts to kill the messengers that bring us the news of poor student achievement, we will be unable to engage in a productive dialogue on education reform and thereby begin to eliminate the chronic instructional failures that afflict so many of the schools attended by poor and minority students.

* * *