

SORTING AND REFORMING: HIGH-STAKES TESTING IN THE PUBLIC SCHOOLS

by

Rachel F. Moran*

I. INTRODUCTION

During the recent television season, the surprise hit was a game show called “Who Wants to be a Millionaire.”¹ Defying the common wisdom that programs today must appeal to “niche” markets, “Millionaire” attracts people of all ages and from all walks of life.² There has been considerable speculation about the reasons for the show’s success,³ but one thing is certain. At the heart of this entertainment phenomenon is multiple-choice testing. In fact, “Millionaire” is the first quiz show to give contestants multiple-choice questions, prompting one expert on television and film to comment: “The difference between ‘Millionaire’ and the 50’s game shows tells you where education is going in this country. In the 50’s game shows, you had to know the exact answer. Today we get multiple choice on a show whose title isn’t even punctuated correctly.”⁴

The show’s format demonstrates that multiple-choice tests have become an accepted part of American life, so commonplace that millions of people tune in to see who will reap the rich rewards of selecting the correct answer. In fact, the format on “Millionaire” has triggered many of the same concerns that generally arise about multiple-

* Robert D. and Leslie-Kay Raven Professor of Law, University of California School of Law (Boalt Hall). I would like to thank the sponsors of the University of Akron’s conference on “Education and the Constitution: Shaping Each Other and the Next Century” for inviting me to participate.

¹ Bill Carter, *‘Millionaire’ Cashes In*, N.Y. TIMES, May 21, 2000, at § 4, at 5 (describing how the show’s unprecedented success has enabled ABC to demand the highest advertising prices in history); Bill Carter, *‘Millionaire’ Adds a Night*, N.Y. TIMES, May 17, 2000, at E10 (noting that the show is so popular that ABC has expanded it to four nights a week on prime time television).

² Bill Carter, *A Million Reasons the TV Quiz Show Has Struck It Rich*, N.Y. TIMES, Nov. 14, 1999, § 1, at 1; Frank Rich, *Who Doesn’t Want to Be a Millionaire*, N.Y. TIMES, Nov. 20, 1999, at A13.

³ See, e.g., Carter, *supra* note 2 (noting the long hiatus since game shows were popular in the 1950s); Rich, *supra* note 2 (greed and materialism coupled with economic insecurity due to a growing divide in wealth); Julia Chaplin, *Quiz Shows Make Every Contestant a Video Gladiator*, N.Y. TIMES, Jan. 16, 2000, at § 9, at 1 (showing high-tech effects and computer anxiety); Joseph Siano, *The Extraordinary Thrill of Watching Ordinary Winners*, N.Y. TIMES, Feb. 6, 2000, §13, at 4 (connecting to ordinary contestants in a depersonalized world); Paul Brownfield, *He’s America’s Dealer of Choice; With Regis Philbin Holding the Cards, ‘Who Wants to Be a Millionaire’ Feeds Off the Same High-Stakes Tension That Lights Up Las Vegas*, L.A. TIMES, Calendar, at 3 (love of gambling for large stakes).

⁴ Carter, *supra* note 2 (citing Robert Thompson, a professor at Syracuse University).

choice tests. In the first place, there have been doubts about the accuracy of the scoring. Early on, two contestants were informed that they had given wrong answers, when in fact their responses were correct.⁵ In addition, some viewers have called in to complain that questions have been drafted in misleading and inappropriate ways.⁶ Moreover, some have expressed reservations about the validity of the test. That is, the questions do not seem well-calculated to identify who is entitled to one million dollars. The most widespread criticism has been that the questions are too easy. The insurance company that guarantees the prize money insisted that the show pose more difficult problems to avoid an excessive number of winners.⁷ Alex Trebek, host of the high-brow game show Jeopardy, joked that contestants on “Millionaire” were asked simple questions like “What’s the usual color of Post-Its?”⁸ Finally, and perhaps most perniciously, the show has been widely criticized because its contestants are overwhelmingly white and male. Indeed, the show’s host, Regis Philbin, issued an on-air appeal to women and minority viewers to enter the selection process.⁹ Critics have alleged that a time-pressured format and biased questions have disproportionately advantaged white men.¹⁰

America’s fascination with “Millionaire” provides a glimpse into the power of high-stakes testing and its

⁵ Sarah Shannon, *U.S. Imports Millionaire Quiz, Plus the Mistakes*, EVENING STANDARD, Aug. 24, 1999, at 17; Bill Carter, *Hit Quiz Show Misses as Correct Contestant Loses*, N.Y. TIMES, Aug. 20, 1999, at B1.

⁶ Amy Pagnozzi, *Betting On a Million*, HARTFORD COURANT, May 19, 2000 at A3 (describing how one viewer complained that a question that attributed the term computer “bug” to a 1947 incident involving a moth that flew inside a computer ignored far earlier uses of the term “bug,” for example, by Thomas Edison in 1889).

⁷ Brownfield, *supra* note 3 (describing a lawsuit that Goshawk Syndicate, brokers at Lloyd’s of London, filed in February 2000 based on allegations that the questions had been “dumbed down,” making the game too easy).

⁸ Bill Hoffmann, *Trek: Millionaire Wins No Prizes*, N.Y. POST, Mar. 20, 2000, at 83; Michael Precker, *Brain Games: The Eggheads Compare Quiz-Show Questions*, DALLAS MORNING NEWS, April 22, 2000, at 1C. Another game show host, Ben Stein, did a parody of “Millionaire” in which a contestant was asked to choose his first name from four possibilities. Allison Fass, *Who Wants to Parody ‘Millionaire’? There Is No Final Answer*, N.Y. TIMES, Apr. 6, 2000, at C10.

⁹ Robert Schaeffer, *Who Wants to Be a Contestant?*, N.Y. TIMES, Feb. 19, 2000, at A15 (noting Philbin’s on-air plea after barely 10% of those who competed on the show were women and none were African-American).

¹⁰ *Id.*; Bill Carter, *Quiz Show Draws More Women Contestants*, N.Y. TIMES, Feb. 28, 2000, at C13 (describing efforts to include more “female material” and “black culture”); Amy Alexander, *Who Wants to Be a Millionaire? We All Do!*, BOSTON GLOBE, May 14, 2000, at E1 (asserting that the show’s questions are biased because they are written by white men); *see also* Alex Kuczynski, *Quiz Screening Seems to Reflect Gender Skew*, N.Y. TIMES, Apr. 24, 2000, at C16 (claiming that the show had resorted to skewing the questions used in the selection process to enable more women to qualify).

hold on the popular imagination. But while this television show affects the fortunes of only a few contestants each night, other forms of standardized testing have daily consequences for millions of students in America. These tests are being used to determine who will be promoted from one grade to the next, who will receive a high school diploma, and who will gain admission to elite colleges and universities. As with “Millionaire,” these practices are widely accepted, roundly applauded, but nevertheless quite controversial. In this article, I will first examine the historical origins of high-stakes testing. Next, I will describe the growing interest in these tests in elementary and secondary schools as well as the tensions that have resulted. Then, I will explore the most significant challenges to the use of high-stakes testing as a requirement for graduation or promotion to another grade. This article will close by contemplating the likely future of the movement for testing and accountability.

II. THE ROOTS OF HIGH-STAKES TESTING

The use of standards in American education is far from a contemporary innovation. During the late nineteenth century, schools took on newfound importance as America confronted the combined forces of urbanization, industrialization, and immigration. Attendance became compulsory as educators assumed the task of preparing students to become productive workers and good citizens. The “common” school was designed to take students from all walks of life, assimilate them to a shared set of values, and inculcate in them a necessary set of skills.¹¹ Concerned about accountability and standardization, particularly in urban schools, business and professional leaders pressed for centralized administrative control of school districts. The hope was that schools could adopt the same model of governance as corporations used and apply a “scientific” approach to teaching.¹²

In response to the growing importance of education, national reformers took an interest in developing curricular guidelines that would help schools perform their critical public function. In 1892, the National Education Association created the Committee of Ten to draft recommendations for strengthening the curriculum in America’s high schools. This group was the first “blue ribbon panel” to address such an issue.¹³ The Committee’s work was influential in developing not only secondary school course offerings, but also standards for college admissions. The

¹¹ DAVID B. TYACK, *THE ONE BEST SYSTEM: A HISTORY OF AMERICAN URBAN EDUCATION* 28-59 (1974).

¹² *Id.* at 126-76; Elliot W. Eisner, *Standards for American Schools: Help or Hindrance?*, 76 *PHI DELTA KAPPAN* 758, 759-60 (1995).

¹³ DIANE RAVITCH, *NATIONAL STANDARDS IN AMERICAN EDUCATION: A CITIZEN’S GUIDE* 36-37 (1995).

resulting recommendations focused on giving modern academic subjects the same weight as classical ones and on ensuring that all students received the same preparation for “the duties of life” through a liberal education, regardless of whether they were college-bound.¹⁴

The Committee of Ten’s report prompted the formation of yet another committee, this time to review college admissions criteria.¹⁵ As a result of efforts to standardize high school graduation and college entry requirements, the College Entrance Examination Board was established. The Board developed a unified examination system for admission to college. The topics chosen for the test shaped high school offerings at schools with students headed for higher education; the scope and difficulty of the tests established something akin to voluntary national standards in each subject.¹⁶ The College Entrance Examination Board’s work laid a foundation for the development of standardized testing in higher education. After World War II, these tests became a commonplace feature of college admissions. The impetus for multiple-choice entrance examinations came from elite institutions. Top educators wanted to shake up the traditional system of merely admitting the children of alumni without much concern for their scholastic aptitude. To compete with European universities, college presidents at institutions like Harvard and Yale believed that they needed a meritocratic system of admissions. Because a wide pool of graduating high school seniors would have to be evaluated to identify the most academically talented, the system had to be relatively inexpensive to administer. Thus was born the Scholastic Aptitude Test, now a fixture of the college application process.¹⁷

The success of standardized testing in colleges and universities paved the way for their use in elementary and secondary education. At first, the tests were primarily a way to sort students by ability and to place them in appropriate programs.¹⁸ In the 1970s, however, a movement emerged to implement testing for promotion and graduation. This “minimum competency” movement enjoyed some success, but it also met with strong opposition

¹⁴ *Id.* at 37-39.

¹⁵ *Id.* at 39.

¹⁶ *Id.* at 40-41.

¹⁷ For an engaging account of the development of the standardized testing movement in college admissions, see NICHOLAS LEMANN, *THE BIG TEST: THE SECRET HISTORY OF THE AMERICAN MERITOCRACY* 1-52 (1999).

¹⁸ RAVITCH, *supra* note 13, at 46-47.

because of concerns about fairness. The United States had only recently embarked on efforts to desegregate the schools and to undo past inequities in education based on race. As 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. Moreover, placing students who failed the test in remedial classes could resegregate schools that had only recently been ordered to achieve racial balance in the student body.¹⁹ These objections slowed down the implementation of high-stakes testing, but interest in standards and accountability was rekindled in the early 1980s, leading to the contemporary push for testing requirements.

III. THE RECENT RISE OF HIGH-STAKES TESTING

In 1983, the National Commission on Excellence in Education published a report entitled *A Nation at Risk*, which warned that America's children were falling behind their peers in other countries and that an alarming number were functionally illiterate and innumerate.²⁰ The report went on to urge that Congress undertake national educational reform to reverse "a rising tide of mediocrity that threatens our very future as a Nation and a people."²¹ Despite this ominous warning, federal officials could not agree on comprehensive educational reform, nor was Congress willing to appropriate substantial additional monies for the public schools.²²

Instead, policymakers grew increasingly interested in the use of national standards and accountability.²³ For the accountability movement to remain a low-cost reform initiative, it had to rely heavily on standardized tests. Moreover, to address the outcry over social promotions and meaningless high school diplomas, the tests had to have real consequences for a student's future. The result was high-stakes testing, which the National Research Council defines as a test on which "an individual student's score determines not just who needs help but whether a student is allowed to take a certain program or class, or will be promoted to the next grade, or will graduate from high

¹⁹ *Id.* at 47-50.

²⁰ NATIONAL COMM'N ON EXCELLENCE IN EDUC., UNITED STATES DEP'T OF EDUC., *A NATION AT RISK: THE IMPERATIVE FOR EDUCATIONAL REFORM* 8-9 (1983).

²¹ *Id.* at 5.

²² Hagit Elul, *Making the Grade, Public Education Reform: The Use of Standardized Testing to Retain Students and Deny Diplomas*, 30 COLUM. HUM. RTS. L. REV. 495, 497-500 (1999); Michael W. Kirst & Allan Odden, *National Initiatives and State Education Policy*, 4 STAN. L. & POL'Y REV. 99, 99 (1992-93).

²³ Elul, *supra* note 22, at 499; Kirst & Odden, *supra* note 22, at 99-100.

school.”²⁴ Testing was touted as both a scientifically legitimate and a relatively inexpensive way to “get tough” on the public schools.

High-stakes tests have become standard practice throughout much of American public education, but this school reform strategy has hardly been free of controversy. Because of the popularity of the movement, policymakers have battled over who should set the standards and get the political credit. In his 1997 State of the Union address, President Clinton called for “a national crusade for education standards—not federal government standards, but national standards, representing what all our students must know to succeed in the knowledge economy of the twenty-first century.”²⁵ He followed through with efforts to develop voluntary national tests that would assess fourth-graders’ and eighth-graders’ proficiency in reading and mathematics.²⁶ In 1998, Congress brought the drive for federally imposed accountability to a halt when it barred the Department of Education from using appropriations to develop a national standards test.²⁷ At present, the only federally sponsored testing is the National Assessment of Educational Progress (NAEP). Developed in the mid-1980s, the NAEP is administered on a purely voluntary basis to make state-by-state comparisons of student achievement. The test is used exclusively to generate research data; it has no stakes attached for students, teachers, local school districts, or state agencies.²⁸

So far, then, efforts to impose uniform, national standards have been stymied. If Congress were to hold school districts receiving federal monies accountable by administering a mandatory national test, state and local officials fear that their traditional authority over designing and implementing the public school curriculum would be undermined. Even voluntary national tests are controversial because they make state and local educational tests seem at best parochial and at worst pointless.²⁹ Congressional intermeddling seems particularly inappropriate in light

²⁴ COMMISSION ON BEHAV. & SOC. SCI. & EDUC., NAT’L RES. COUNCIL, HIGH STAKES: TESTING FOR TRACKING, PROMOTION, AND GRADUATION 14 (Jay P. Heubert & Robert M. Hauser eds. 1999) [hereinafter NAT’L RES. COUNCIL].

²⁵ *President Clinton’s Message to Congress on the State of the Union* N.Y. TIMES, Feb. 5, 1997, at A20; see also NAT’L RES. COUNCIL, *supra* note 24, at 13.

²⁶ NAT’L RES. COUNCIL, *supra* note 24, at 14.

²⁷ Department of Education Appropriations Act of 1998, Pub. L. No. 105-277, § 447, 112 Stat. 2681 (1998).

²⁸ RAVITCH, *supra* note 13, at 70-71.

²⁹ NAT’L RES. COUNCIL, *supra* note 24, at 15; Elul, *supra* note 22, at 500; Kirst & Odden, *supra* note 22, at 104; Arthur L. Coleman, *Excellence and Equity in Education: High Standards for High-Stakes Tests*, 6 VA. J. SOC.

of the steady decline in the federal government's financial support for public schools throughout the 1980s and much of the 1990s.³⁰

Because of this impasse in developing a national examination, states have taken the primary responsibility for adopting standards testing. At present, 19 states require students to pass an exit examination to receive a high school diploma, and seven plan to adopt such tests in the next few years.³¹ The tests have proven particularly popular in the South. In 1998, eleven of 15 states below the Mason-Dixon line required students to pass a high school exit examination. By contrast, of the remaining 35 states, only six had imposed such requirements.³² State reliance on high-stakes testing continues to grow. Although the Wisconsin legislature's joint finance committee last year voted 13 to 3 to leave an exit testing program unfunded and to delete the law creating the examination, this decision bucked the trend toward expanded use of standards testing.³³

There has been ongoing debate about the propriety of high-stakes testing. One aspect of the controversy is technocratic. Experts accuse politicians of misunderstanding or willfully ignoring the proper applications of standardized tests. In particular, when state and local officials are under pressure to do something about failing schools, they tend to look for a quick fix. Rather than develop new tests, they use existing ones, even though they were constructed for a wholly different set of purposes.³⁴ In Chicago, for example, the school district decided to administer the Iowa Test of Basic Skills (ITBS) to identify low-performing schools and students. However, the ITBS is a norm-referenced test, one that by definition classifies some children as above average and others as below

POL'Y & L. 81, 89 (1998).

³⁰ Lois Salisbury, et al., *One Simple Standard: Every Child Can Learn*, 4 STAN. L. & POL'Y REV. 87, 92-94 (1992-93); David Cohen, *What Standards for National Standards?*, 76 PHI DELTA KAPPAN 751, 753 (1995). For a general account of the tensions among federal, state, and local governments over standards testing, see RAVITCH, *supra* note 13, at 150-53, 158-60.

³¹ John Gehring, *'High Stakes' Exams Seen as Test for Voc. Ed.*, 19 EDUC. WEEK ON THE WEB 1 (February 2, 2000) <<http://www.edweek.org/ew/ewstory.cfm>>.

³² Linda Jacobson, *State Graduation Tests Raise Questions, Stakes*, 17 EDUC. WEEK ON THE WEB 22 (June 24, 1998) <<http://www.edweek.org/ew/1998/41gradof17.cfm>>.

³³ Julie Blair, *Wisconsin Committee Votes to Kill State Graduation Test*, 18 EDUC. WEEK ON THE WEB 22 (June 23, 1999) <<http://www.edweek.org/ew/1999/41wisc.h18>>; Jacobson, *supra* note 32.

³⁴ See Coleman, *supra* note 29, at 90.

average based on their performance relative to each other. Moreover, the ITBS's scoring procedure is based on national, aggregate performance without any special consideration of Chicago's curricular priorities and academic standards.³⁵ Indeed, the test's own developers have indicated that it should not be used as the sole basis for retaining students at a grade level or for selecting students for special instructional programs.³⁶ Although the Chicago public school system conducted a pilot test of the Chicago Academic Standards Examination in early 1999, the chief accountability officer expressed fears that dropping the ITBS would cause the district to lose credibility with the media and the public.³⁷

Another line of confrontation has emerged between teachers and school administrators who must undertake the day-to-day business of educating children on the one hand and politicians who pronounce testing and accountability a panacea for the school system's ills on the other. Teachers worry that standards are being developed without much input from them and that the tests cast a shadow over daily instruction.³⁸ When districts face negative labels, loss of funds, and constraints on their autonomy if scores are low, teachers are under intense pressure to teach to the test, even at the cost of displacing other important curricular material. The incentive to focus on the test to the exclusion of other academic subjects is particularly strong in low-income, minority school districts, where at-risk students jeopardize a school's chance of gaining a favorable rating.³⁹ In addition, standards testing

³⁵ NAT'L RES. COUNCIL, *supra* note 24, at 30-31.

³⁶ *Id.* at 261. *But see* Robert A. Davis, *Public School Test Policy Hit; Too Much Rides on Scores, Critics Say*, CHICAGO SUN-TIMES, Jan. 28, 1999, at 12 (noting that the publisher of the ITBS defended Chicago's decision to use it as a factor in grade promotions).

³⁷ NAT'L RES. COUNCIL, *supra* note 24, at 31; Rosalind Rossi, 'Too Many Tests,' *Students Complain*, CHICAGO SUN-TIMES, Jan. 12, 2000, at 20; Michael Martinez, *Exam Protest Taken to School Board Office*, CHICAGO TRIBUNE, Jan. 12, 2000, at 2.

³⁸ *See* Anne C. Lewis, *An Overview of the Standards Movement*, 76 PHI DELTA KAPPAN 744, 749-50 (1995). A notable exception is an experimental program in California that was designed to allow local school districts to play a central role in developing their own standards. The program operates on the premise that "meaningful change must begin with introspection by the educators charged with daily responsibility for the school's organizational culture and extend outward to all those affected by change—including students, teachers, parents, administrators, and the community." Salisbury, et al., *supra* note 30, at 91.

³⁹ Linda McNeil & Angela Valenzuela, *The Harmful Impact of the TAAS System of Testing in Texas: Beneath the Accountability Rhetoric* 9-28 (n.d.) (unpublished manuscript on file with author); William P. Moore, *The Devaluation of Standardized Testing: One District's Response to a Mandated Assessment*, 7 APPLIED MEASUREMENT IN EDUCATION 343, 352-59 (1994); M. Gail Jones, et al., *The Impact of High-Stakes Testing on Teachers and Students in North Carolina*, 81 PHI DELTA KAPPAN 199, 200-02 (1999); Ann Cook, et al., *Unmasking the Low Standards of High-Stakes Tests*, 19 EDUC. WEEK ON THE WEB (Mar. 8, 2000)

may force teachers to try to fit all students into the same mold, regardless of differences in their learning styles. As

one scholar in the field of education has noted:

Uniform standards may be appropriate for business—a manufacturer wants all its microwave ovens to meet specified standards of quality. That’s good. But to what extent do we want all students to be alike? Of course, there are certain basic skills we want all students to master, but should all students be required or expected to attain them at exactly the same age or grade level?⁴⁰

Students who cannot fit the mold are in danger of being pushed out of the regular testing system to keep a school’s scores up. Students may be placed in special education, held back, encouraged to be absent the day of the examination, or even steered toward transfer to another school or dropping out.⁴¹ When these strategies fail, some desperate districts have resorted to cheating to improve their test performance.⁴²

Finally, the tests create conflict between disadvantaged communities and state and local proponents of standards testing. Interestingly, some accountability initiatives have been adopted as a corollary of school finance reform. In both Kentucky and Texas, for example, the state supreme courts held that reliance on local property taxes to fund the public schools resulted in such gross disparities among districts that the finance system was unconstitutional.⁴³ When the state legislatures developed plans that equalized school funding, they also adopted accountability measures to ensure that the state’s enlarged role in allocating funds actually led to improved academic

<<http://www.edweek.org/ew/ewstory.cfm>>; Robert C. Johnston, *In Texas, the Arrival of Spring Means the Focus Is on Testing*, EDUC. WEEK ON THE WEB (Apr. 29, 1998) <<http://www.edweek.org/1998/33tex.h17>>. [But see Ulrich Boser, *Teaching to the Test?* EDUC. WEEK, June 7, 2000, at 1 \(reporting on a forthcoming University of Wisconsin study that finds that “there may be little overlap between what state assessments test and what teachers teach.”\)](#)

⁴⁰ Charles M. Reigeluth, *Educational Standards: To Standardize or to Customize Learning?*, 78 PHI DELTA KAPPAN 202, 204 (1997).

⁴¹ Linda Darling-Hammond, *Creating Standards of Practice and Delivery for Learner-Centered Schools*, 4 STAN. L. & POL’Y REV. 37, 39-40 (1992-93); Debra Viadero, *Testing System in Texas Yet to Get Final Grade*, EDUC. WEEK, May 31, 2000, at 1; Gary Orfield & Johanna Wald, *The High-Stakes Testing Mania Hurts Poor and Minority Students the Most: Testing, Testing*, THE NATION, June 5, 2000, at 38; Lorrie A. Shepard, *Will National Tests Improve Student Learning?*, 73 PHI DELTA KAPPAN 232, 234 (1991).

⁴² David J. Hoff, *As Stakes Rise, Definition of Cheating Blurs*, EDUC. WEEK, June 21, 2000, at 1.

⁴³ *Rose v. Council for Better Educ.*, 790 S.W.2d 186 (Ky. 1989); *Edgewood Indep. Sch. Dist. v. Kirby*, 777 S.W.2d 391 (Tex. 1989) (Edgewood I); *Edgewood Indep. Sch. Dist. v. Kirby*, 804 S.W.2d 491 (Tex. 1991) (Edgewood II); *Carrollton-Farmers Branch Indep. Sch. Dist. v. Edgewood Indep. Sch. Dist.*, 826 S.W.2d 489 (Tex. 1992) (Edgewood III); *Edgewood Indep. Sch. Dist. v. Meno*, 917 S.W.2d 717 (Tex. 1995) (Edgewood IV). For a description of the lengthy battle over school finance reform in Texas, see J. Steven Farr & Mark Trachtenberg, *The Edgewood Drama: An Epic Quest for Education Equity*, 17 YALE L. & POL’Y REV. 607 (1999).

outcomes.⁴⁴ In short, school finance reforms coupled with accountability measures were seen as a way to ensure that students in less affluent districts received not just money, but also a meaningful education.

Even so, parents, students, and community leaders in low-income and minority communities have questioned the fairness of standardized tests.⁴⁵ High-stakes testing has been concentrated in states and districts with substantial low-income and minority populations.⁴⁶ Not only are students of color more likely than whites to have to take an examination, but available evidence suggests that they are disproportionately likely to fail it.⁴⁷ Although defenders of accountability testing argue that standards are needed to protect at-risk students from educational neglect, disparities in passage rates have led to charges that the tests do not measure merit but instead mark class privilege. As one critic has charged, “America’s elites now perpetuate themselves with gatekeeping rules of their own making, rules legitimated by scientific objectivity.”⁴⁸ The tests have been attacked because they require the kind of cultural capital that low-income children of color are unlikely to have. According to critics, questions drafted with white, middle-class students in mind seldom resonate with the lives of non-white or poor children. Perhaps the most striking example of this cultural insensitivity is a mandate that students with limited English proficiency take high-stakes tests only in English, a practice that effectively prevents them from fully demonstrating their academic skills.⁴⁹

Alternatively, testing opponents argue that students in historically underfinanced and typically segregated

⁴⁴ Ellen Williams, *Annual Survey of Texas Law–Education*, 49 SMU L. REV. 901, 902-03 (1996); Brian Gong & Edward F. Reidy, *Assessment and Accountability in Kentucky’s School Reform*, in I PERFORMANCE-BASED STUDENT ASSESSMENT: CHALLENGES AND POSSIBILITIES 215, 216-18 (Joan Boykoff Baron and Dennie Palmer Wolf eds., 1996); C. Scott Trimble & Andrew C. Forsaith, *Achieving Equity and Excellence in Kentucky Education*, 28 U. MICH. J.L. REFORM 599, 599-600 (1995); Charles J. Russo, *School-Based Decisionmaking in Kentucky: Dawn of a New Era or Nothing New Under the Sun?*, 83 KY. L.J. 123, 123-25 (1994-95).

⁴⁵ Richard Whitmire, ‘Pushback’ Movement, Which Resists Education Reforms, *Gaining Ground*, GANNETT NEWS SERVICE, Dec. 6, 1999; Marjorie Coeyman, *Urban Parents Protest ‘Unrealistic’ Standards*, CHRISTIAN SCIENCE MONITOR, Oct. 19, 1999, at 17.

⁴⁶ Sean F. Reardon, *Eighth-Grade Minority Competency Testing and Early High School Dropout Patterns* 4-5 (Apr. 1996) (paper presented at the annual meeting of the American Educational Research Association).

⁴⁷ Coleman, *supra* note 29, at 91.

⁴⁸ Peter Sacks, *Standardized Testing: Meritocracy’s Crooked Yardstick*, 29 CHANGE 24, 31 (Mar./Apr. 1997).

⁴⁹ See ANGELA VALENZUELA, *SUBTRACTIVE SCHOOLING: U.S.-MEXICAN YOUTH AND THE POLITICS OF CARING* 172-73 (Christine E. Sleeter ed., 1999).

schools have not had an equal opportunity to meet the standards imposed by accountability tests. The tests stigmatize those who are held back in a lower grade or denied a high school diploma, even if the fault lies with a deficient educational system. Far from motivating children, this message of failure could prompt them to quit school or withdraw from their studies. As a result, those students most in need will be lost to the educational system.⁵⁰ Moreover, critics contend that labeling schools as low-performing in traditionally disadvantaged communities can backfire as a reform strategy. When these schools are placed into receivership or denied funding because of low test scores, teachers and administrators become discouraged. Instead of continuing to work in schools that serve at-risk students, these educators seek jobs in affluent districts with students who arrive well-fed, healthy, rested, and ready to learn. Rather than enhance the educational environment for the disadvantaged, the accountability system may make it that much more tempting for teachers and administrators to leave schools already considered undesirable, low-status institutions.⁵¹

Although the federal government has not undertaken to mandate tests itself, concerns about fairness have prompted some legislative and administrative activity. Late in 1999, the President's Advisory Commission on Educational Excellence for Hispanic Americans endorsed standards-based reform but criticized high-stakes testing programs that fail to account for the special needs of Hispanic students. The Commission called on the Department of Education's Office for Civil Rights (the OCR) to launch a state-by-state inquiry into testing practices that discriminate against Hispanic students. Of particular concern was the treatment of students with limited English proficiency.⁵²

⁵⁰ See *supra* note 41.

⁵¹ See Richard J. Stiggins, *Assessment, Student Confidence, and School Success*, 81 PHI DELTA KAPPAN 191, 192-93 (1999); Linda Darling-Hammond, *The Implications of Testing Policy for Quality and Equality*, 73 PHI DELTA KAPPAN 220, 223 (1991); Jones, et al., *supra* note 39, at 202; Martin L. Maehr & Jane M. Maehr, "Schools Aren't As Good As They Used to Be"; *They Never Were*, 25 EDUC. RESEARCHER 21, 22-23 (1996); Milbrey W. McLaughlin, *Test-Based Accountability As a Reform Strategy*, 73 PHI DELTA KAPPAN 248, 250 (1991); Kenneth A. Sirotnik & Kathy Kimball, *Standards for Standards-Based Accountability Systems*, 81 PHI DELTA KAPPAN 202, 211 (1999).

⁵² PRESIDENT'S ADVISORY COMMISSION ON EDUCATIONAL EXCELLENCE FOR HISPANIC AMERICANS, A REPORT TO THE NATION: POLICIES AND ISSUES ON TESTING HISPANIC STUDENTS IN THE UNITED STATES (1999) <<http://www.ed.gov/offices/OIIA/hispanic/PAC/September99.html>>; Mary Ann Zehr, *Hispanic Students 'Left Out' By High-Stakes Tests, Panel Concludes*, 19 EDUC. WEEK ON THE WEB 5 (Sept. 22, 1999) <<http://www.edweek.org/ew/ewstory.cfm>>. The Office for Civil Rights responded cautiously to the request for an investigation. Richard T. Cooper, *Inquiry Into English-Only Tests Ordered*, L.A. TIMES, Sept. 16, 1999, at A3.

Although the OCR has not launched a special inquiry into the treatment of Hispanic students, it did circulate a revised draft of guidelines on high-stakes testing a few months after the Commission's report was released. The proposed guidelines are designed to clarify the practices permissible under Title VI and Title IX of the Civil Rights Act of 1964. The advisory standards specify the principles of measurement that should be observed to minimize dangers of discrimination, and they indicate that the tests should be "educationally justified" if they are to survive legal scrutiny.⁵³ So far, the OCR has investigated complaints about high school exit examination requirements in several states, including Ohio, Nevada, North Carolina, and Texas. Chicago's testing program also has triggered a complaint. In both Ohio and Texas, federal and state officials agreed that the tests could be used, but that all students must have access to remedial assistance, multiple chances to take the examination, and appropriate classroom instruction to prepare for it. The investigations in Nevada and North Carolina are likely to yield similar results.⁵⁴

In addition, Senator Paul Wellstone of Minnesota has sponsored the Fairness and Accuracy in Student Testing Act, which would ban elementary and secondary schools receiving federal financial assistance from using standardized tests as the sole factor in making "any decision about the retention, graduation, tracking, or within-class ability grouping of an individual student . . ."⁵⁵ The legislation further provides that such tests must be reliable and valid, based on a functional performance level rather than on performance relative to that of other students, aligned with the curriculum and classroom instruction, offered on multiple occasions, administered in compliance with guidelines provided by the test developer or publisher, appropriate for the intended purposes, and modified to accommodate disabled and limited-English-proficient students.⁵⁶ As a list of supporters shows, the bill responds to reservations about high-stakes testing shared by teachers, educational experts, and advocates on behalf of low-income, minority students. Wellstone's legislation has been endorsed by educational organizations, such as the

⁵³ Julie Blair, *OCR Issues Revised Guidance on High-Stakes Testing*, 19 EDUC. WEEK ON THE WEB 25 (Jan. 12, 2000) <<http://www.edweek.org/ew/ewstory.cfm>>.

⁵⁴ Ben Wildavsky, *Achievement Testing Gets Its Day in Court*, U.S. NEWS AND WORLD REPORT, Sept. 27, 1999, at 30, 32; Gail Russell Chaddock, *Adverse Impact*, CHRISTIAN SCIENCE MONITOR, Nov. 30, 1999, at 14; Rosalind Rossi, *Complaints Hit School Promotion Rule*, CHICAGO SUN-TIMES, Oct. 22, 1999, at 34.

⁵⁵ S. 2348, 106th Cong. § 1(b) (2000); *see also* H.R. 4333, 106th Cong. § 1(b) (2000).

⁵⁶ S. 2348, at § 1 (b); H.R. 4333, at § 1(b).

National PTA, the Council of Great City Schools, the National Education Association, the American Psychological Association, and the National Association of School Psychologists, as well as by racial and ethnic organizations, such as the National Association for the Advancement of Colored People, the Mexican American Legal Defense and Education Fund, and the National Urban League.⁵⁷ In offering the legislation, Wellstone revealed that he was motivated in part by his own personal history. He himself has a learning disability that hampers his performance on standardized tests.⁵⁸ As a result, Wellstone felt compelled to sponsor the legislation, even though it faces an uphill battle in Congress.⁵⁹

IV. LEGAL CHALLENGES TO HIGH-STAKES TESTING

Concerned parents, students, and community leaders are not waiting for Congress or the Office for Civil Rights to save them from the perils of high-stakes testing. In several states, parents have protested the use of tests as graduation requirements, and students have boycotted standards testing.⁶⁰ In Massachusetts this year, hundreds of teachers, students, and parents protested the Massachusetts Comprehensive Assessment System (MCAS) by rallying on the capitol steps. At the rally, students and parents carried signs that read “High incomes=High MCAS scores” and “Hey [Governor] Celluci, kiss my MCAS.”⁶¹ Students also delivered petitions and packets of information to legislators and the governor. Moreover, hundreds of students around the state boycotted the MCAS on the day it was given.⁶² Similar boycotts and protests have taken place in Arizona, Florida, Georgia, Illinois, Louisiana, Ohio, and Wisconsin.⁶³

⁵⁷ Rob Hotakainen, *Calling Graduation Tests Harsh, Wellstone Offers Alternative Plan*, MINNEAPOLIS-ST. PAUL STAR TRIBUNE, Apr. 4, 2000, at 1A.

⁵⁸ Paul D. Wellstone, *Even a Senator Could Fail This School Test*, USA TODAY, Jan. 13, 2000, at 18A.

⁵⁹ Hotakainen, *supra* note 57.

⁶⁰ Whitmire, *supra* note 45; Coeyman, *supra* note 45.

⁶¹ Anand Vaishnav & Doreen Iudica Vigue, *Hundreds of Parents, Teachers, Students Rally Against MCAS*, BOSTON GLOBE, May 16, 2000, at B3; Erica Thesing, *Teens Tell Governor He Is Failing Them; MCAS Foes Deliver 800 Letters of Protest*, BOSTON GLOBE, June 6, 2000, at B2.

⁶² John Gehring, *Students Boycott Tests in Mass. To Protest Emphasis on Exams*, 19 EDUCATION WEEK ON THE WEB 8 (Apr. 19, 2000) <<http://www.edweek.org/ew/ewstory.cfm>>; Doreen Iudica Vigue & Tara Yaekel, *Hundreds of Students Boycott MCAS Test Sanctions Are Issued in Scattered Protests*, BOSTON GLOBE, Apr. 13, 2000, at A1.

⁶³ Kelly Pearce, *Court Unmasks AIMS; State Ordered to Disclose Questions*, ARIZ. REPUBLIC, Apr. 22, 2000, at A1; Robbie Sherwood & Kelly Pearce, *Students Organize Protests Demanding Better Schools*, ARIZ. REPUBLIC, Apr. 6,

Some students and advocacy organizations have turned to the courts for protection from high-stakes examinations. The leading case on high school exit examinations is *Debra P. v. Turlington*.⁶⁴ During the push for minimum competency requirements, the Florida legislature passed the Educational Accountability Act of 1976. The Act established three requirements for graduation from Florida public high schools: (1) a minimum number of credits; (2) mastery of basic skills; and (3) satisfactory performance in functional literacy as determined by the State Board of Education. The Act also provided for remediation and mandated creation of a comprehensive testing program to evaluate basic skills.⁶⁵ In 1978, the legislature amended the Act to require that students pass a functional literacy examination as a condition of receiving a high school diploma. Those who failed the test but had sufficient credits to graduate would receive a certificate of completion.⁶⁶ The distinction between a diploma and certification of completion was significant: only those who received a diploma were eligible for certain state jobs and admission to the state university system.⁶⁷

When the test was administered, black students were disproportionately barred from receiving a diploma because they could not pass it. On the test's first administration, 78% of black students failed the communications or mathematics section or both; by contrast, only 25% of white students failed one or both sections. The pattern was the same during two subsequent administrations of the examination. As a result, 20% of black students remained ineligible for a diploma compared to only 1.9% of white students.⁶⁸ Because of these disparate outcomes, a class action was filed in federal district court against the state and local officials responsible for adopting and implementing the test. Brought on behalf of present and future twelfth-grade students who had failed or would fail

2000, at A1; Sophia Lezin Jones, *Gwinnett Parents Protest Gateway; All-or-Nothing Test Draws Fire From Those Who Say It Should Not Be Only Criterion for Promotion*, ATLANTA JOURNAL AND CONSTITUTION, Oct. 23, 1999, at 3B; Jacques Steinberg, *Blue Books Closed, Students Protest State Tests*, N.Y. TIMES, Apr. 13, 2000, at A28.

⁶⁴ *Debra P. v. Turlington*, 474 F. Supp. 244 (M.D. Fla. 1979), *vacated and remanded*, 644 F.2d 397 (5th Cir. 1981), *on remand*, 564 F. Supp. 177 (M.D. Fla. 1983), *aff'd*, 730 F.2d 1405 (5th Cir. 1984).

⁶⁵ *Debra P.*, 474 F. Supp. at 247-48.

⁶⁶ *Id.* at 248.

⁶⁷ *Id.* at 249.

⁶⁸ *Id.* at 248-49.

the test, the suit alleged that the test was discriminatory and violated due process.⁶⁹

The discrimination claims were based on equal protection, Title VI, and the Equal Educational Opportunities Act. The plaintiffs contended that the test contained biased items, perpetuated the vestiges of past discrimination, and resegregated black students in remedial classes.⁷⁰ The trial court did not find any evidence of discrimination in the test's construction because Florida officials had used methods of test design that were generally accepted as valid and reliable.⁷¹ To address whether the test perpetuated past discrimination, the court reviewed the history of public elementary and secondary education in Florida. The state's entire school system had been subject to *de jure* (that is, state-mandated) segregation on the basis of race from 1890 to 1967. Race-based pupil assignments were declared unconstitutional in 1967, and the state's schools were still under a federal desegregation decree.⁷² The dual school system had created grave educational inequities, and students taking the exit examination in 1978 had spent at least some time in a system that had yet to undo the effects of past discrimination. Even if the test had been adopted with the best of intentions, it could not add insult to constitutional injury by denying diplomas to students victimized by officially sanctioned racism. Therefore, the use of the test as a graduation requirement had to be postponed until the senior class had been through twelve years of schooling in a desegregated system.⁷³ The trial court rejected the claim that disproportionate assignment of blacks to remediation led to impermissible segregation. The judge held that these assignments were a constitutionally acceptable way to address the lingering effects of unequal educational opportunity.⁷⁴

In addition, the test violated both procedural and substantive due process. The district court first found that students had protected property and liberty interests in a diploma. The property right arose because the state of

⁶⁹ *Id.* at 246-47. There were three classes of plaintiffs, one composed of all students who had failed or would fail the test, and two composed of black students in this situation. One class of black students was statewide, and one was limited to Hillsborough County, Florida. *Debra P.*, 474 F. Supp. at 246.

⁷⁰ *Id.* at 246-47, 250-51, 267-68.

⁷¹ *Id.* at 254, 260-62.

⁷² *Id.* at 250-52.

⁷³ *Id.* at 255-57.

⁷⁴ *Id.* at 268.

Florida compelled students to attend school and, in turn, offered them a diploma with economic value. The liberty interest derived from a student's right to be free of the stigma of being labeled "functionally illiterate" by state officials. As a result, Florida officials could not divest high school pupils of these interests without following the dictates of due process.⁷⁵

With respect to procedural due process, state and local agencies had to provide students with adequate notice. The court held that two years' advance warning was insufficient, and that Florida had to allow students more time to prepare for and meet the new requirement.⁷⁶ On appeal, the Fifth Circuit basically agreed with the trial court's conclusions, but added an additional requirement. Under substantive due process, the state had to demonstrate that all students had a fair opportunity to learn; that is, the curriculum had to offer them a chance to master the skills necessary to pass the test.⁷⁷ On remand, Florida officials introduced surveys of teachers, students, and administrators to establish the curricular validity of the exit examination. In addition, officials visited selected classrooms to observe the material that was covered.⁷⁸ The trial court held, and the Fifth Circuit agreed, that this evidence was sufficient to satisfy substantive due process.⁷⁹ The federal judiciary was not willing to impose the onerous burden of showing that every student in every classroom actually had the same opportunity to learn.⁸⁰

Recently, several new legal challenges have been filed. The most significant is *GI Forum v. Texas Education Agency*,⁸¹ a class action that relied directly on the legacy of *Debra P.* Individual students and several advocacy organizations sued Texas officials, alleging that the use of the Texas Academic Assessment System (TAAS) as a graduation requirement discriminated against black and Hispanic students and violated due process. Adopted in 1990, the TAAS test certainly produced disparate outcomes for students based on race and ethnicity. In

⁷⁵ *Debra P.*, 474 F. Supp. at 266.

⁷⁶ *Id.* at 267.

⁷⁷ *Debra P. v. Turlington*, 644 F.2d 397, 404-07.

⁷⁸ *Debra P. v. Turlington*, 564 F. Supp. 177, 179-82.

⁷⁹ *Id.* at 185-86, *aff'd*, 730 F.2d at 1411.

⁸⁰ *Id.* at 183-84, *aff'd*, 730 F.2d at 1409.

⁸¹ *GI Forum v. Texas Educ. Agency*, 87 F. Supp.2d 667 (W.D. Tex. 2000).

1991, 67% of black tenth-graders and 59% of Hispanic tenth-graders failed the examination, while only 31% of whites did.⁸²

Citing the Equal Protection Clause, the plaintiffs claimed that Texas officials had intentionally discriminated against non-white students in adopting and implementing the examination. They based this argument on the fact that the Texas Education Agency adopted cut-off scores of 70%, even when it knew that they would seriously harm black and Hispanic students. Moreover, as circumstantial evidence of animus, the plaintiffs cited procedural irregularities in the rush to adopt the test.⁸³ The defendants countered that they had adopted and implemented the TAAS as a way to ensure that all students received a meaningful education, regardless of race or ethnicity. Far from acting with discriminatory intent, legislators and state education officials sought to improve the academic performance of low-income, minority students.⁸⁴ On a motion for summary judgment, the district court concluded that there was insufficient evidence of discriminatory intent to merit a trial and dismissed the plaintiffs' claim.⁸⁵

The plaintiffs argued in the alternative that state officials had violated Title VI because the test had a discriminatory effect, regardless of whether it was adopted with a benign motivation. The defendants responded that there was no private cause of action under Title VI and that wrongful intent was a prerequisite under both Title VI and the Equal Protection Clause.⁸⁶ The trial court rejected both of these contentions and allowed the plaintiffs to proceed to trial on the Title VI claim.⁸⁷ Citing the gap in test scores between white, black, and Hispanic students, the plaintiffs asserted that the test had an impermissible disparate impact. In making this argument, they asked that the four-fifths rule used in employment discrimination be applied in an educational setting—at least when diploma requirements were at issue. That is, if the passage rate for non-whites was less than 80 percent of that for whites, the

⁸² *Id.* at 671, 673.

⁸³ Plaintiffs' Response and Brief in Opposition to Defendants' Motion for Summary Judgment, at 7, 17-18, 25-26, *GI Forum v. Texas Educ. Agency*, 87 F. Supp.2d 667 (W.D. Tex. June 22, 1999) (No. SA-97-CA-1278 EP).

⁸⁴ Defendants' Motion for Summary Judgment and Brief in Support, at 10-14, *GI Forum v. Texas Education Agency*, 87 F. Supp.2d 667 (W.D. Tex. June 4, 1999) (No. SA-97-CA-1278 EP).

⁸⁵ *GI Forum*, 87 F. Supp.2d at 670.

⁸⁶ Defendants' Motion for Summary Judgment and Brief in Support, *supra* note 84, at 20-28.

⁸⁷ *GI Forum*, 87 F. Supp.2d at 669.

plaintiffs would make out a *prima facie* case of impermissible disparate impact.⁸⁸

The defendants responded that the court should look at cumulative, not first-time, pass rates because only when students had failed the test after a number of attempts were they denied a diploma. According to state officials, remediation efforts steadily closed the gap between whites and non-whites as students took the tests on successive occasions.⁸⁹ The defendants also questioned the wisdom of importing employment standards into the field of education, but pointed out that non-whites' cumulative pass rates were slightly over 80 percent of those for whites.⁹⁰ Ultimately, the trial court agreed that cumulative pass rates were the relevant scores to compare but rejected rigid reliance on a four-fifths rule. Instead, the judge concluded that the disparity was wide enough to qualify as substantial, and the burden shifted to defendants to show that the TAAS test was educationally necessary.⁹¹ The defendants satisfied this burden by demonstrating that the test was scientifically constructed and that the high stakes attached to it created important incentives for students to learn.⁹² The plaintiffs were unable to persuade the district court that their proffered alternatives would be equally effective in pushing students to master basic concepts.⁹³

The plaintiffs also alleged that the TAAS test violated equal protection and the Equal Educational Opportunities Act because it perpetuated the vestiges of past discrimination in a state with a history of unconstitutional segregation, wrongful discrimination against limited-English-proficient students, and a grossly unequal system of school finance.⁹⁴ The defendants argued, and the district court agreed, that these historical

⁸⁸ Plaintiffs' Response and Brief in Opposition to Defendants' Motion for Summary Judgment, *supra* note 83, at 12-14.

⁸⁹ Defendants' Motion for Summary Judgment and Brief in Support, *supra* note 84, at 24-25; Reply Brief in Further Support of Defendants' Motion for Summary Judgment, at 5, *GI Forum v. Texas Educ. Agency*, 87 F. Supp.2d 667 (W.D. Tex. July 2, 1999) (No. SA-97-CA-1278 EP).

⁹⁰ Defendants' Motion for Summary Judgment and Brief in Support, *supra* note 84, at 22-25.

⁹¹ 87 F. Supp.2d at 677-79.

⁹² *Id.* at 679-81.

⁹³ *Id.* at 681-82.

⁹⁴ Plaintiffs' Response and Brief in Opposition to Defendants' Motion for Summary Judgment, *supra* note 83, at 39-42.

wrongs had been sufficiently cured and that the test did not unfairly penalize students who had been victimized by prior governmental misconduct.⁹⁵ Relying on the Equal Educational Opportunities Act, the plaintiffs also contended that the TAAS test had the effect of excluding limited-English-proficient students from the instructional program. The trial court did not permit this claim to go to trial, presumably because it accepted the state's argument that a diploma is not a form of instruction.⁹⁶

Following *Debra P.*, the plaintiffs asserted that students had both protected liberty and property interests in a diploma. Although the defendants argued that there was no constitutionally safeguarded interest whatsoever, the trial judge accepted the plaintiffs' position that due process protections were triggered by the TAAS requirement.⁹⁷ With respect to procedural due process, both plaintiffs and defendants conceded that notice of the TAAS test generally had been adequate. However, plaintiffs asserted that limited-English-proficient students had not received sufficient warning, in part because they sometimes entered high school at an advanced age and did not have enough time to prepare to take the test in English. The court rejected this claim, presumably accepting the defendants' view that general notice, not the plight of individual students who voluntarily entered the district late in their high school careers, was the relevant concern.⁹⁸

With respect to substantive due process, the plaintiffs contended that curricular validity had to be judged based on the regular curriculum. The defendants insisted, and the district court agreed, that remedial programs should be considered in assessing whether students had an opportunity to learn that prepared them for the test.⁹⁹ The

⁹⁵ *GI Forum*, 87 F. Supp.2d at 670; Defendants' Motion for Summary Judgment and Brief in Support, *supra* note 84, at 28-29; Reply Brief in Further Support of Defendants' Motion for Summary Judgment, *supra* note 89, at 12.

⁹⁶ Defendants' Motion for Summary Judgment and Brief in Support, *supra* note 84, at 29 n.6.

⁹⁷ *GI Forum*, 87 F. Supp.2d at 682; Plaintiffs' Response and Brief in Opposition to Defendants' Motion for Summary Judgment, *supra* note 83, at 35-37; Defendants' Motion for Summary Judgment and Brief in Support, *supra* note 84, at 14-17; Reply Brief in Further Support of Defendants' Motion for Summary Judgment, *supra* note 89, at 9-11.

⁹⁸ *GI Forum*, 87 F. Supp.2d at 682; Defendants' Motion for Summary Judgment and Brief in Support, *supra* note 84, at 17-18; Plaintiffs' Response and Brief in Opposition to Defendants' Motion for Summary Judgment, *supra* note 83, at 37-38.

⁹⁹ *GI Forum*, 87 F. Supp.2d at 682; Plaintiffs' Response and Brief in Opposition to Defendants' Motion for Summary Judgment, *supra* note 83, at 38-39; Defendants' Motion for Summary Judgment and Brief in Support, *supra* note 84, at 18-20.

defendants distinguished Texas's situation from that in Florida because the state had adopted the Texas Essential Knowledge and Skills to guide curricular development in the public schools. The TAAS test was designed to evaluate mastery of these skills, and consultations with teachers during the construction of the test confirmed that the examination measured skills taught in the classroom.¹⁰⁰ Although the Texas Education Agency had not conducted surveys and site visits like those in *Debra P.*, the trial court found that it had discharged the obligation of establishing general curricular validity.¹⁰¹

There are several notable points about the *GI Forum* decision that could be important in future litigation. First, with respect to equality issues, the district court focused exclusively on the testing process itself and largely refused to explore how the test affects a student's overall educational experience. In addressing claims of bias, the judge relied heavily on the scientific methods used to construct the TAAS test. Although disproportionate failure rates for blacks and Hispanics were nonetheless cause for concern, the court accepted the defendants' argument that a high-stakes test was uniquely beneficial in motivating students and therefore was educationally necessary. The judge did not require elaborate statistical proof that the threat of being denied a diploma effectively prodded at-risk students to learn. The court deferred to the defendants' account of the test's impact, even though the plaintiffs had argued that the test demoralized black and Hispanic students and increased the likelihood that they would drop out of school.¹⁰² The trial court refused to accept the plaintiffs' evidence that high rates of failure on the TAAS test correlated with increased drop-out rates. The court wanted proof of a causal relationship, not a mere correlation. In the court's view, students might drop out for a number of reasons that had nothing to do with the TAAS test.¹⁰³

Here, the allocation of the burden of proof was critical. Even when the plaintiffs had made out a *prima*

¹⁰⁰ Dr. S.E. Phillips, The Texas Assessment of Academic Skills Exit Level Test, Defendant's Exhibit D316, at 10-11, *GI Forum v. Texas Educ. Agency*, 87 F. Supp.2d 667 (W.D. Tex. Jan. 1999) (No. SA-97-CA-1278 EP); Professor Philip Uri Treisman, Preliminary Expert Witness Report, at 14-15, *GI Forum v. Texas Educ. Agency*, Defendant's Exhibit D33, 87 F. Supp.2d 667 (W.D. Tex. Feb. 25, 1999) (No. SA-97-CA-1278 EP).

¹⁰¹ *GI Forum*, 87 F. Supp.2d at 682-83.

¹⁰² See Dr. Walter M. Haney, Preliminary Report on Texas Assessment of Academic Skills Exit Test (TAAS X), Plaintiff's Exhibit P 36, *GI Forum v. Texas Educ. Agency*, 87 F. Supp.2d 667 (W.D. Tex. Dec. 11, 1998) (No. SA-97-CA-1236 EP); Dr. Walter M. Haney, Supplementary Report on Texas Assessment of Academic Skills Exit Test (TAAS-X), Plaintiff's Exhibit, *GI Forum v. Texas Educ. Agency*, 87 F. Supp.2d 667 (W.D. Tex. July 30, 1999; editorial revisions Aug. 5, 1999) (No. SA-97-CA-1278 EP).

¹⁰³ *GI Forum*, 87 F. Supp.2d at 681.

facie case of disproportionate impact, the court insisted that they demonstrate that the TAAS test was not a good motivator. However, because the defendants bore the burden of establishing that the test was educationally necessary, the court logically should have required proof of a causal relationship between its administration and improved student performance. The defendants did offer evidence that, on successive administrations, the scores for black and Hispanic students improved.¹⁰⁴ However, this evidence did not plainly establish a causal relationship between the test and student motivation. Most importantly, students who dropped out of school would not be among the those who took the test multiple times. As a result, the multiple test-takers might constitute a pool of self-selected students who were already highly determined to succeed despite repeated obstacles. In addition, even if the test had no effect on student motivation, pass rates might improve because of the practice effect of taking the test over and over. In short, the defendants' evidence that the test was an educationally necessary incentive was probably no more convincing than the plaintiffs' evidence that it was detrimental.

The court also accepted the defendants' definition of what was at stake as a result of TAAS testing. According to the trial judge, the only relevant inequity related to the denial of a diploma, not the impact on the students' general educational experience. For instance, the plaintiffs contended that the disparity in first-time pass rates was the relevant criterion for determining whether the test had a discriminatory impact. Assuming that the primary concern was that students were stigmatized by this failure, were assigned to remediation, and perhaps even dropped out of school, then first-time pass rates were the proper focus of a Title VI inquiry. Failure could have a negative effect on a student's education, regardless of whether the district ever formally denied a student a diploma. The court, however, emphasized the gap in cumulative pass rates because, only after multiple failed attempts, did the state of Texas penalize students by withholding a degree. As a result, the sole question was how substantially the gap in pass rates had narrowed, even if the cost was that blacks and Hispanics failed the test many more times than did whites, lost many more hours of regular instruction than did whites, and perhaps decided to drop out of school more often than did whites. This narrow focus enabled the court to defer to "the State's right to pursue educational policies that it legitimately believes are in the best interests of Texas students," even if "the policies are debated and debatable among learned people."¹⁰⁵

¹⁰⁴ Phillips, *supra* note 100, at 13-15; Treisman, *supra* note 100, at 21-33.

¹⁰⁵ *GI Forum*, 87 F. Supp.2d at 671.

The court's unwillingness to look at the test in context is further evidenced by its treatment of substantive due process issues. In *Debra P.*, the state of Florida could not simply assert that because the exit examination measured basic skills, it necessarily drew on material taught in public school classrooms. Nor could officials point to the steps they had taken to ensure that the test was reliable and valid. Instead, the state had to offer survey evidence and site visits to convince the court that children had a fair opportunity to prepare for the test. The trial court in *GI Forum* refused to impose the same evidentiary burden on the state of Texas. Instead, it relied on the state's adoption of a formal curriculum, the Texas Essential Knowledge and Skills, and its consultation with select classroom teachers in developing the test. Moreover, the judge allowed state officials to demonstrate that remediation prepared students to pass the TAAS test eventually, although a fair opportunity to learn arguably requires that all students receive appropriate instruction in the regular curriculum.

The proof of curricular validity offered by the state of Texas does not compare favorably with the evaluation of classroom teaching offered in *Debra P.* A formally mandated curriculum cannot ensure a common classroom experience for all children. There must be some evidence that the paper requirements are in fact being implemented. Teachers who are consulted when a test is designed are hardly likely to be a random sample. Instead, they are probably among the unusually successful and engaged instructors chosen for high-profile assignments. Even if these teachers were not self-selected, their number is so small that representativeness would still remain a serious problem. Again, by deferring to the state's view that curricular validity could be assured without much direct evidence of the classroom experience, the court in *GI Forum* adopted a deferential stance, assuming that officials acted in good faith and "would agree with the proposition that unequal education is a matter of great concern and must be eradicated."¹⁰⁶

Faced with defeat on both equal protection and due process grounds, the plaintiffs chose not to appeal the district court's ruling in *GI Forum*, presumably because counsel feared that a conservative Fifth Circuit Court of Appeals might overturn the helpful precedent in *Debra P.*¹⁰⁷ Other cases also have met with limited success in the

¹⁰⁶ *GI Forum*, 87 F. Supp.2d at 683.

¹⁰⁷ Cecilia Balli, *MALDEF Won't Appeal Ruling; Group will Eye Texas' New Test*, SAN ANTONIO EXPRESS-NEWS, Feb. 8, 2000, at 1B. In the wake of the state's successful defense of TAAS testing, the Board of Regents has proposed standardized testing for college students at the University of Texas. Chris Vaughn, *UT Regents Consider Standardized Tests; Faculty Reaction 'Anger and Disbelief'*, FORT WORTH STAR-TELEGRAM, Sept. 20, 2000, at 5; Linda K. Wertheimer, *UT Schools Weigh Systemwide Testing; Faculty Question TAAS-Like Accountability Plan*,

courts. In *Erik V. v. Causby*,¹⁰⁸ the plaintiffs challenged a North Carolina county's Student Accountability Policy, which provided that students who failed to obtain a designated score on a state standardized test would be held back, regardless of their academic performance during the school year. Students who failed the test had to undergo remediation in the summer and retake the test. If they failed, they could take the test one more time. If they had not passed after three attempts, they were denied promotion to the next grade. The testing requirement was waived for special education and limited-English-proficient students.¹⁰⁹

The plaintiffs alleged that minority students' pass rates were about half of those for whites. Based on these disparate outcomes, the lawsuit asserted that the accountability program violated federal equal protection, Title VI, and due process, as well as North Carolina's own constitutional and statutory protections. The trial court denied the plaintiffs' request for a preliminary injunction. Although the judge did not reach the merits of the case, he signaled that "federal courts have no business substituting their judgment for that of the local school board when it comes to qualitative achievement standards for promotion."¹¹⁰ The parties eventually settled the case out of court, and the district instituted a policy that makes students who fail the test but have better than a C average eligible for a waiver. In addition, school officials created a parental notification program and a system for developing individualized remediation plans. As a result, the number of students held back dropped substantially.¹¹¹

In Louisiana, parents opposed to the state's educational assessment program filed suit to obtain copies of the test questions. The court held that although the public records act did not explicitly exempt the questions from disclosure, releasing them would be "nonsensical" because the test's validity would be compromised.¹¹² In addition, the plaintiffs sought to enjoin Louisiana officials from using test results in making promotion decisions, but the district court threw out the case as premature since no results had yet been released. In doing so, the judge indicated

DALLAS MORNING NEWS, Sept. 19, 2000, at 1A.

¹⁰⁸ *Erik V. v. Causby*, 977 F. Supp. 384 (E.D.N.C. 1997).

¹⁰⁹ *Id.* at 386-87; *see also* Elul, *supra* note 22, at 501-02.

¹¹⁰ *Erik V.*, 977 F. Supp. at 388.

¹¹¹ Elul, *supra* note 22, at 503.

¹¹² Anand Vaishnav, *Parents Cannot Get Copies of LEAP Test; State District Judge Calls Request 'Nonsensical'*, NEW ORLEANS TIMES-PICAYUNE, Feb. 10, 2000, at A5.

that she had serious doubts that students had a protected property interest in promotion to the next grade.¹¹³ After the litigation, the state board of education recommended that local officials retain students in fourth grade no more than one time, regardless of their test scores.¹¹⁴ Despite these setbacks, litigation over high-stakes testing continues. Recently, in Indiana, students with disabilities filed suit, alleging that the state's exit examination fails to accommodate their unique needs.¹¹⁵

V. THE FUTURE OF HIGH-STAKES TESTING

The *GI Forum* decision and other recent cases demonstrate that in a conservative era, federal courts are increasingly unwilling to probe the workings of state educational systems. The judiciary's reformist impulse has largely been spent after decades of struggle over desegregation. In the late 1960s and 1970s, federal judges embarked on a lengthy campaign to desegregate public schools. The result was protracted, and sometimes intractable, litigation that produced at best a tenuous racial balance, when the goal of integration was not wholly thwarted by white flight.¹¹⁶ Beginning in the 1980s and continuing into the 1990s, district courts began to terminate these cases and withdraw from oversight of desegregation decrees.¹¹⁷

With these weighty and pervasive obligations at an end, federal judges are undoubtedly reluctant to take up the task of scrutinizing ongoing racial inequities through the lens of high-stakes testing. Instead, as in *GI Forum*, courts would prefer to define the issues narrowly so that the entire system of public elementary and secondary education is not on trial. By focusing on the scientific methods used to construct a high-stakes test, courts can require the state to be race-neutral in devising and administering questions without insisting that other inequities in the educational system be remedied. If curricular validity can be shown by formally mandated standards and

¹¹³ Parents for Educ. Justice v. Picard, No. 00-633, 2000 U.S. Dist. LEXIS 6382 (E.D. La. Apr. 20, 2000); Randy McClain, *Suit Attacks LEAP Use in Promotions*, BATON ROUGE ADVOCATE, Mar. 2, 2000, at 11A; *Sensible Decision on LEAP*, NEW ORLEANS TIMES-PICAYUNE, Apr. 26, 2000, at B6.

¹¹⁴ Will Sentell, *BESE Changes Rule on Holding Back 4th-Graders*, BATON ROUGE ADVOCATE, May 1, 2000, at 1A.

¹¹⁵ See *Rene v. Reed*, 726 N.E. 2d 808 (Ind. App. 2000) (class certification); Lynn Olson, *Indiana Case Focuses on Special Ed.: Suit Challenges High-Stakes Tests*, EDUC. WEEK, May 31, 2000, at 1.

¹¹⁶ Rachel F. Moran, *Milo's Miracle*, 29 CONN. L. REV. 1079, 1084-87 (1997).

¹¹⁷ *Id.* at 1087-90.

consultation with select teachers, then courts can affirm that students have an equal opportunity to learn, even when classroom experiences in fact differ dramatically. In any event, testing cases are unlikely to lead to systemic reform, even if they succeed. Instead, the test will be thrown out, and inequalities in educational opportunity will persist.

Court-generated reform is more likely to come about in state than in federal court, and this reform is more likely to prompt – rather than derive from – high-stakes testing. For instance, in both Kentucky and Texas, the high courts held that the system of financing public schools was inequitable enough to violate the state constitution.¹¹⁸ As a result, the Kentucky and Texas legislatures undertook comprehensive reform initiatives, in which accountability and high-stakes testing were one component. Far from driving systemic change, testing programs were used to measure the success of independent efforts to improve the public schools. These tests were incidental, not essential, to reform. For example, in Texas, if the TAAS test had been successfully challenged in federal court, the state would nevertheless have remained committed to school finance equalization as mandated by its own supreme court.

In states such as Kentucky and Texas, the schools historically have been funded at such low levels that even with court-mandated school finance reform, classrooms continue to confront the dilemma of limited resources. For example, Kentucky ranked forty-third in per-pupil spending and thirty-seventh in average teacher's salary when the state high court held the school system unconstitutional. Only 68.2% of ninth-graders finished high school, and in poor Appalachian districts, 48.4% of the population was functionally illiterate.¹¹⁹ Texas, too, has long ranked in the bottom tier of states with respect to per-capita student spending on schools and other services for children.¹²⁰ Its school finance system also produced gross disparities in the resources available to children in low-wealth and high-wealth districts. For some districts, the ratio of available resources was as large as 700 to 1.¹²¹ In addition, there was a significant correlation between district wealth and the racial make-up of the student body.¹²² Although

¹¹⁸ See *supra* notes 43-44 and accompanying text.

¹¹⁹ Trimble & Forsaith, *supra* note 44, at 601.

¹²⁰ Robert Reinhold, *Smile When You Say 'School Reform,'* *Pardner*, N.Y. TIMES, Jan. 4, 1987, § 12, at 60; J. Michael Kennedy, *Doomsday Budget Shows Texas the Cost of Not Raising Taxes*, L.A. TIMES, May 22, 1991, at A5; Jennifer Bradley, *Tex-Mix*, NEW REPUBLIC, Mar. 25, 1996, at 27.

¹²¹ Farr & Trachtenberg, *supra* note 43, at 615.

¹²² *Id.* at 642-43.

disparities in available resources have been moderated, Texas remains a state that invests relatively little in its schools. The state still ranks well below the national average in per-capita student spending, and its teachers receive comparatively low wages and few benefits.¹²³ In fact, Texas school districts have confronted such difficulties in staffing their classrooms that some have begun to recruit high school students to teach.¹²⁴

Where testing programs have been adopted legislatively without the impetus of court-mandated reform, they have sometimes been justified as independent levers for systemic reform. Unfortunately, there is little evidence that test results alone can motivate students and schools to perform better. Most states that have adopted high-stakes testing requirements did not appropriate substantial amounts of money to enhance classroom instruction. Indeed, some legislatures did not even want to pay to devise new tests; instead, officials tried to rely on existing tests, even when they were inappropriate for use in decisions about graduation, grade promotion, or tracking. Overall, then, the tests have been another bureaucratic requirement for schools to satisfy, but they have not led to an infusion of funds to handle the extra demands.¹²⁵

High-stakes testing is popular because it offers a way to identify and blame individuals without acknowledging a collective unwillingness to invest in public schools, particularly those in low-income, often minority areas. The tests have taken on the dimensions of a morality tale.¹²⁶ Failure is located in low-performing schools and failing pupils, not in systems of public education that suffer from serious inequities in the allocation of resources. The wealthiest students, who are enrolled in private schools, are spared the need to take high-stakes tests to demonstrate their bona fides. Middle-class and upper middle-class students in suburban communities expect to pass the test, and indeed, some of them view it as an annoyance that diverts them from more important work, such as

¹²³ Tyce Palmaffy, *The Gold Star State*, 3/1/98 POL'Y REV. 30, 30 (1998).

¹²⁴ Bob Banta, *Courses Prepare Students to Fill Teacher Gap*, AUSTIN AMERICAN-STATESMAN, May 25, 2000, at 1.

¹²⁵ Sirotnik & Kimball, *supra* note 51, at 214; George F. Madaus, *The Effects of Important Tests on Students*, 73 PHI DELTA KAPPAN 226, 230 (1991); Eisner, *supra* note 12, at 764; Reigeluth, *supra* note 40, at 206; Linda Darling-Hammond & Beverly Falk, *Using Standards and Assessments to Support Student Learning*, 79 PHI DELTA KAPPAN 190, 196-97 (1997); Rhona S. Weinstein, *High Standards in a Tracked System of Schooling: For Which Students and with What Educational Supports?*, 25 EDUC. RESEARCHER 16, 16-17 (1996).

¹²⁶ Orfield & Wald, *supra* note 41 (describing the popular perception that high-stakes tests are “a necessary form of tough love for struggling schools”).

preparing for advanced placement tests or college entrance examinations.¹²⁷

In low-income schools that serve large numbers of at-risk students, however, the tests occupy a major place in the educational process. Schools risk being labeled low-performing and losing their autonomy; teachers risk being accused of neglect and losing their credibility; and students risk being identified as inadequate and losing their respectability. To prove both institutional and individual worth, principals must emphasize the importance of the test, teachers must revise their lesson plans to ensure that children are prepared, and children must be given incentives such as parties and prizes to keep up lengthy and sometimes tedious drills and practice examinations. So much effort is diverted to the test that it becomes not a minimum standard, but the primary standard. The school's goal is to ensure that each child meets the minimum rather than to ensure that each child's potential is maximized. As a consequence, children in low-income communities of color who can readily pass become captives of test-driven instruction directed at their at-risk peers.¹²⁸

The current push for standards testing has enjoyed a long period of popularity, certainly longer than that of the minimum competency movement of the 1970s. For one thing, the rhetoric of holding students, teachers, and schools accountable sounds more upright and less derogatory than labeling them incompetent. For another, the country continues to enjoy an unprecedented run of economic prosperity. In a global economy, the United States has acquitted itself quite well, although the gap between the rich and the poor grows ever larger and the status of the middle-class seems increasingly fragile and tentative.¹²⁹ To fuel America's economic engine, it has been necessary to sort individuals into appropriate niches. The sorting at the top has seemingly been perfected through college and

¹²⁷ See, e.g., Melanie Markley, *Regardless of Where They Live, Parents Ask the Same Question: How Good Are the Schools?*, HOUSTON CHRONICLE, Oct. 13, 1996, at 2 (focusing on minimum standards diverts Texas public schools from "tak[ing] bright, privileged students to a higher level of learning"); Reinhold, *supra* note 120 (TAAS tests do little to benefit affluent schools and simply add "a layer of restrictions on creativity"); Richard Lee Colvin, *Crusader Argues School Reforms Hinder Learning*, L.A. TIMES, Feb. 22, 2000, at A1 (a suburban parent complained that test preparation displaced her daughter's writing workshop).

¹²⁸ McLaughlin, *supra* note 51, at 248.

¹²⁹ DANIEL H. WEINBERG, CENSUS BUREAU, CURRENT POPULATION REPORTS, HOUSEHOLD ECONOMIC STUDIES 60-191, A BRIEF LOOK AT POSTWAR U.S. INCOME INEQUALITY 3-4 (1996) <<http://www.census.gov/prod/1/pop/p60-191.pdf>>. (economic stratification has grown steadily beginning in 1968 and continuing through the 1990s); TERESA A. SULLIVAN, ET AL., THE FRAGILE MIDDLE CLASS: AMERICANS IN DEBT (2000) (documenting that the middle class is now more vulnerable than ever to going bankrupt as a result of divorce, job loss, injury, and illness).

graduate admissions processes, which have relied on standardized testing to a significant degree in the post-World War II era.¹³⁰ Now that a high school diploma is seen as a bare minimum for most entry-level jobs, sorting individuals at the bottom of the job queue also has become imperative.¹³¹ If a diploma is not a reliable marker of basic skills, employers will squander time and resources on training the untrainable. The accountability movement offers a way to distinguish not just between high school dropouts and high school graduates, but also between those who attended class and those who mastered the material.

The standards movement so far has not been derailed by the forces that hampered the minimum competency movement. In the 1970s, the country was still grappling with the legacy of officially mandated racial segregation. Racial unrest was very real, and a movement that seemed to reinstate racial hierarchy, even through a scientifically constructed test, was suspect.¹³² Today, however, there is a growing clamor for formal colorblindness. According to this view, individuals should be judged on their own merits, regardless of race. A properly designed and administered test is seen as a perfectly appropriate way to hold individuals accountable, whether or not it results in racial disparities.¹³³ In addition, many of the social reform movements of the 1970s were overshadowed by the Vietnam War, Watergate, and economic troubles. So far, persistent problems of this magnitude have not come along to displace the popular fascination with standards and accountability. Peace and prosperity have given Americans the luxury of tinkering with reform, although record budget surpluses at the federal and state levels have

¹³⁰ See LEMANN, *supra* note 17, at 343-45.

¹³¹ See Gerald W. Bracey, *The Forgotten 42%*, 80 PHI DELTA KAPPAN 711 (1999) (noting that in 1996, half of high school dropouts lived below the poverty line compared to 24% of high school graduates, 16% of those with some college, and 2.5% of college graduates); Gerald W. Bracey, *The Eighth Bracey Report on the Condition of Public Education*, 80 PHI DELTA KAPPAN 112 (1998) (noting economic stratification in a robust economy and a widespread and overstated belief that schools are tightly linked to the economy).

¹³² See John A. Powell & Marguerite L. Spencer, *Remaking the Urban University for the Urban Student: Talking About Race*, 30 CONN. L. REV. 1247, 1248 (1998) (describing how the use of standards testing abated to some degree in the post-*Brown* era but reemerged in the late 1970s); Note, *Teaching Inequality: The Problem of Public School Tracking*, 102 HARV. L. REV. 1318, 1318 (1989) (noting short period in the late 1960s and 1970s when courts struck down tracking based on standardized test scores as racially discriminatory); David Kirp, *Intelligent Policy*, 2 PSYCH. PUB. POL'Y & L. 603, 610 (1996) (describing campus protests in 1969 when Professor Arthur Jensen published an article on racial differences in intelligence).

¹³³ See Powell & Spencer, *supra* note 132, at 1286-90 (describing increased reliance on standardized testing in college and university admissions in the push for formally colorblind decision-making).

not led legislatures to open the purse strings for public education.¹³⁴ Perhaps, when the United States faces a fiscal downturn or international unrest, the interest in refining educational standards will subside. Unfortunately, the opportunity for meaningful change through public investment in the school system will have passed as well. Tests used for sorting will become a distraction from, and even a substitute for, genuine reform.

VI. CONCLUSION

High-stakes testing has become a pervasive part of the American educational scene. Indeed, prime-time television has joined the craze with the hit show of the season, “Who Wants to Be a Millionaire.” Just as that show demonstrates our fascination with the power of multiple choice tests, it also reveals our cynicism and ambivalence. The questions require contestants to recall trivia, spouting facts that have little or nothing to do with their ability to be valuable and productive persons. At times, the audience secretly wonders why knowing inconsequential information should make such a huge difference in a person’s financial fortunes. After all, the outcomes seem to turn on a combination of blind luck and random bits of knowledge, not on the contestant’s inherent worth as a person, worker, or citizen. Even so, the show starkly reveals the consequences of failure. Someone who once had the potential to win a million dollars can go home with a pittance. A person who held the interest of millions of viewers, whose struggle was high drama, disappears during a commercial and is promptly forgotten.

The drama of high-stakes testing in the public schools takes place every year for millions of children across the nation. Despite some legal challenges and lobbying efforts, the accountability movement remains a powerful part of the American political scene. However well-developed the science behind assessment and measurement may be, high-stakes tests cannot work miracles in failing systems. The examinations are a yardstick and, yes, even a stick, but they are not a magic wand. And, unlike a television game show, America cannot conveniently arrange for those who fail the tests to disappear, nor can its responsibilities to these children be promptly forgotten.

¹³⁴ See Rachel F. Moran, *Bilingual Education, Immigration, and the Culture of Disinvestment*, 2 J. GENDER, RACE & JUST. 163, 205-06 (1999).