

Testing Whiteness: No Child or No School Left Behind?

Helen A. Moore*

I began my study of schooling and critical pedagogy as a research assistant for a program called PRIME: Program Research in Multicultural Education. One of our challenges in the mid-1970s was to assist the Principal Investigator, Dr. Jane Mercer, in constructing educational materials for her role as an expert witness in *Larry P. v. Riles*.¹ Through that project, I learned a great deal about testing issues as they relate to the whiteness of educational, political, legal and policy perspectives. Today, as a sociologist, I commit my research to an understanding of the processes and consequences of testing policy and the whiteness of evaluation paradigms as they play out in education on issues of social inequality.

Omi and Winant² argue that racial formation is a “fundamental organizing principle” for all macro-social relationships, including schooling in the United States.³ At micro-levels, we interact with others in a variety of social and educational settings, including the conditions under which we, or our students or children, take tests, and how we use those tests to shape individual educational opportunities.⁴ At the macro-level of collectivity, Omi and Winant encourage us to understand the complex relationships of whiteness to economic, cultural and ideological structures.⁵ Today, the category of “whiteness” itself is an unstable and “decentered” complex set of

* Professor of Sociology, University of Nebraska-Lincoln.

1. 343 F. Supp. 1306 (N.D. Cal. 1972) (preliminary injunction), *aff'd*, 502 F.2d 963 (9th Cir. 1974), 495 F. Supp. 926 (N.D. Cal. 1979) (decision on merits), *aff'd in part, rev'd in part*, 793 F.2d 969 (9th Cir. 1984).

2. MICHAEL OMI & HOWARD WINANT, RACIAL FORMATION IN THE UNITED STATES (1994); see also Henry A. Giroux, *Rewriting the Discourse of Racial Identity*. 67 HARV. EDUC. REV. 285 (1997).

3. OMI & WINANT, *supra* note 2, at 56.

4. *Id.* at 59.

5. *Id.* at 74–75.

social meanings constantly being transformed by political struggle.⁶ These and other sociologists encourage us to examine that struggle because today, as in the past, racial minorities “bear the heavy burden in human suffering as a result of their categorization as ‘other’ in dominant practices and ideologies.”⁷ They argue that we are at a new stage of socially based politics of racial formation. I draw on their assessment of race to consider testing and whiteness as efforts to center and cement racial categories of privilege through testing policies.

In this paper, I examine the No Child Left Behind Act of 2001 (NCLB)⁸ as a manifest social policy that imbeds racial formation practices centered around whiteness into a national movement of standardized testing. The first section of this article provides background information on the NCLB, the construction of school testing, and potential policies that advantage whiteness. The second section describes key elements of “testing whiteness” that invalidate the assumptions of NCLB and raise legal, social, and policy questions for the courts and our communities. I draw on *Larry P.* and the critiques of racialized testing patterns as instructive of potential judicial frameworks for these ongoing educational issues. Our legal and educational responses to NCLB may become the frame for educational policy challenges that confront public education and race dynamics over the coming decades.

I. WHITENESS, TESTING, AND NCLB

The No Child Left Behind Act was signed by President Bush on January 8, 2002.⁹ It represents the first major shift in federal education policy since the war on poverty that produced Head Start¹⁰

6. Giroux, *supra* note 2, at 287.

7. OMI & WINANT, *supra* note 2, at 62.

8. See Pub. L. No. 107-110, 115 Stat. 1425 (2002) (codified in scattered sections of 20 U.S.C.).

9. *Id.*; see also Elizabeth Bumiller, *Focusing on Home Front, Bush Signs Education Bill*, N.Y. TIMES, Jan. 9, 2002, at A1.

10. Head Start was created under the Economic Opportunity Act of 1964, Pub. L. No. 88-452, 78 Stat. 508 (1964) (codified as amended in scattered sections of 42 U.S.C.). It was reauthorized under the Head Start Act, Pub. L. No. 97-35, 96 Stat. 499 (1981) (codified as amended in scattered sections of 42 U.S.C.).

and the Elementary and Secondary Education Act of 1965¹¹ that created Title I funding.¹² Today, the federal government provides approximately seven percent of all funding to public and private schools—a drop from a high of ten percent under administrations of the 1990s.¹³

Unlike the United Kingdom and other nations around the world that have centralized curricular and budget control of educational policy, the U.S. has little or no history of federal educational control.¹⁴ Our Constitution makes no explicit mention of education as a federal role, and thus the Tenth Amendment relegates that control to the states.¹⁵ While the courts have constructed important decisions such as *Brown v. Board of Education*¹⁶ and *Swann v. Charlotte-Mecklenburg Board of Education*,¹⁷ it was not until the 1990s, specifically centered around America 2000 and Goals 2000—that we saw systematic federal efforts to shape educational standards.

NCLB sets an ambitious educational goal of closing the achievement gap among student groups with a particular focus on racial/ethnic (and some other groups such as English Language Learners (ELLs) and Special Education) test scores.¹⁸ Mandating a twelve-year timetable for equality in testing outcomes across groups,¹⁹ Congress optimistically assumes that bureaucratic standards encapsulated in “high-stakes testing” and school report cards will generate the motivations and pedagogies to accomplish this equality.

11. Pub. L. No. 89-10, 79 Stat. 27 (codified in scattered sections of 20 U.S.C.).

12. *See id.* tit. I, 79 Stat. at 27–36 (codified as amended at 20 U.S.C. §§ 6301–578).

13. WILLIAM J. MATHIS, THE FEDERAL “NO CHILD LEFT BEHIND ACT”: WHAT WILL IT COST STATES? 5 (2004), at http://www.aasa.org/NCLB/What_Cost_States-Mathis.pdf (last visited Mar. 22, 2005).

14. *See* James Serenity Boutwell, *A Case of Unconstitutional Immigration: The Importation of England’s National Curriculum to the United States*, 34 VAND. J. TRANSNAT’L L. 333, 345–55 (2001) (discussing centralized national curriculum in England and history of federal policy in education in United States).

15. U.S. CONST. amend. X.

16. 347 U.S. 483 (1954).

17. 402 U.S. 1 (1971).

18. *See* 20 U.S.C. § 6311(b)(2)(C)(v)(II) (2000) (discussing achievement of economically disadvantaged students, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency).

19. *Id.* para. (2)(F) (“[N]ot later than 12 years after the end of the 2001–2002 school year, all students in each group described in subparagraph (C)(v) will meet or exceed the State’s proficient level of academic achievement . . .”).

After identifying the NCLB criteria and sanctions, I will then consider the lack of attention to the whiteness of pedagogical practices, resources, learning, and testing models that frame this legislation.

Under NCLB, standardized test scores for each school, by student subgroups, by district and by state are to be reported annually.²⁰ Every state must test all students annually in math and reading in grades three through eight by the 2005–06 school year.²¹ In 2007, students must also be tested in science.²² States must demonstrate “Adequate Yearly Progress” (AYP) toward specified learning standards for *all* groups of students, as well as across all categorized sub-groups as follows: economically disadvantaged students, racial/ethnic groups, disabled students and limited English proficient students.²³ (Note that the NCLB racial typology fails to recognize the existence of racially mixed children in its categories.) The goal of NCLB is for all students to “meet or exceed the State’s proficient level of academic achievement” on tests for all categories of students by the 2013–14 school year.²⁴ Legislators must assume that public report cards and public awareness of test scores will pressure schools and teachers and students to improve.

Federal dollars are used as positive sanctions for demonstrated “success.”²⁵ For schools that continually increase test scores and reach benchmark levels over the twelve years for all groups in a total of twenty-two categories, budgets will increase.²⁶ Increased federal dollars are included as bonuses in teacher salaries and for various educational expenses.²⁷ Importantly, note that if *one* student group in *one* area fails to meet the testing standard, then the entire school fails

20. *Id.* para. (3).

21. *Id.* subpara. (C)(vii).

22. *Id.* subpara. (C)(v)(II).

23. *Id.* subpara. (C) (defining criteria for AYP).

24. *Id.* subpara. (F).

25. ARTHUR COLEMAN & SCOTT PALMER, NIXON PEABODY LLP, STANDARDS REFORM, FEDERAL LAW, AND THE AMERICAN DIPLOMA PROJECT 17, available at <http://www.edroundtable.state.in.us/pdf/adp/adplegalstudy.pdf> (last visited May 22, 2005).

26. THE WHITE HOUSE, FACT SHEET: NO CHILD LEFT BEHIND ACT, at <http://www.whitehouse.gov/infocus/compassionate/education.html> (last visited Mar. 21, 2005).

27. Elizabeth R. Hinde, *Tyranny of the Test: Elementary Teachers’ Conceptualizations of the Effects of State Standards and Mandated Tests on Their Practice*, 6(10) CURRENT ISSUES IN EDUC. 2003, at <http://cie.ed.asu.edu/volume6/number10> (last visited Mar. 21, 2005).

and federal dollars are invested elsewhere.²⁸ This is a “high stakes test” model for schools. From the outset, the burden of proof is set at the individual and group level of students, and they and their teachers are “responsible” for educational outcomes as demonstrated through tests. The macro-level structures of whiteness and the racial formation of testing are completely obscured in this equation, while individual test scores are disingenuously focused on racial sub-categories for measurement comparisons.

States also can ratchet up overall testing standards by including criteria for attendance, teacher certification rates, student graduation rates, reduction of violence, etc. In the Missouri state standards, a small school (fewer than 250 students) with more than five expulsions per year, or a large school (more than 1,000 students) with more than fifteen expulsions per year, is defined as failing to meet standards.²⁹ NCLB creates this “all or nothing” threshold for “success” in a variety of dimensions. Policy makers and educators have taken the concept of “high-stakes testing” to an illogical and irresponsible extreme that likely will not hold up under scrutiny by educators, parents, students or, I argue, the courts.

If one group in any one area fails, the school fails.³⁰ If the school fails to make AYP in a second year, immediate sanctions result.³¹ All schools that receive Title I funds from the government (two thirds of all public schools)³² and that do not make AYP will face a hierarchy of consequences, beginning with reduced federal dollars.³³ In the second year, the school is classified as “in need of improvement.”³⁴ Every student in the school will be given the option to transfer to a better-performing school (public or public charter school) in the

28. THE WHITE HOUSE, FACT SHEET: NO CHILD LEFT BEHIND ACT, at <http://www.whitehouse.gov/infocus/compassionate/education.html> (last visited Mar. 21, 2005).

29. 50 C.S.R. 50-355.100 (2005).

30. 20 U.S.C. § 6311(b)(2)(I)(i) (2000) (stating the general proposition that “each group of students . . . must meet or exceed the objectives set by the State . . .”).

31. Lance D. Fusarelli, *The Potential Impact of the No Child Left Behind Act on Equity and Diversity in American Education*, 18 EDUC. POL’Y 71, 72 (2004).

32. MYRA SADKER & DAVID SADKER, *TEACHING SCHOOLS AND SOCIETY* 241 (2005).

33. Fusarelli, *supra* note 31, at 73.

34. 20 U.S.C. § 6316(b)(1)(E) (2000 & Supp. III 2004). This subdivision does not specifically use the phrase “in need of improvement,” but rather notes schools that have been “identified for school improvement.” *Id.*

district, with free transportation funded by the public school district.³⁵ NCLB requires that priority in providing school choice will be directed to low-achieving children from low-income families,³⁶ and school districts have some flexibility in meeting this requirement. Districts may restrict which schools are available for transfer and transfers may only be permitted during a certain time period.³⁷ They may sign contracts with neighboring districts to accept students from failing schools, contract with “cyber” schools, create schools within schools, offer supplemental services a year early, hire more teachers, add portable classrooms or build new classrooms at more successful schools.³⁸

After three years of not meeting AYP, the school must also provide “supplemental education services” to those children who remain at the school.³⁹ Those services can include tutoring, remedial classes, after-school services and summer programs (as federal funds are redirected to successful schools).⁴⁰ In the fourth year, the district must take corrective actions such as replacing certain staff or fully implementing a new curriculum.⁴¹ After five consecutive years, the school is identified for restructuring.⁴² Consistent with state law, under restructuring the local educational agency must implement at least one of five alternative governance options: reopening as a charter school; replacing all or most of the school staff; entering into a contract with an entity, such as a private management company, to operate the public school; turning the operation of the school over to the state; or other “major restructuring” of school governance.⁴³

A provision that is not well known and less-often communicated to parents is that parents who oppose these testing practices may “opt out.”⁴⁴ Schools are supposed to inform parents of their right to opt out

35. *Id.* para. (1)(E)–(F) (transfer); *id.* para. (9) (transportation).

36. *Id.* para. (1)(E)(ii).

37. THE WHITE HOUSE, FACT SHEET: NO CHILD LEFT BEHIND ACT, at <http://www.whitehouse.gov/infocus/compassionate/education.html> (last visited Mar. 21, 2005).

38. *Id.*

39. § 6316(b)(5).

40. *Id.* para. (12)(C) (definition of “supplemental educational services”).

41. *Id.* para. (7)(C).

42. *Id.* para. (8).

43. *Id.* subpara. (B).

44. Anita Ramasastry, *No Child Left Unrecruited?*, FINDLAW’S WRIT, Dec. 4, 2002, at

by the act of individuals signing a waiver.⁴⁵ In the rush to comply with the new regulations, many schools have not informed parents of this option.⁴⁶

Fusarelli highlights a series of Catch-22 situations in NCLB testing generated by the guidelines.⁴⁷ Student classifications will shift in ways that preclude careful tracking of score increases for schools over time.⁴⁸ For example, once an ELL student becomes proficient, he or she is removed from the ELL classification.⁴⁹ Thus, scores in successful schools cannot continue to rise because of this ceiling effect. Additionally, NCLB standards require that ninety-five percent of each group in each grade in each school be tested, or by definition the school fails to meet AYP.⁵⁰ In 2003, sixty-three percent of schools in Georgia failed to make AYP because they had fewer than ninety-five-percent test-taking in a particular sub-group.⁵¹ School districts and schools with historically high absenteeism rates are systematically most likely to fail this standard.⁵²

Schools with historically high turnover rates for teachers and students will also be difficult to track authentically for AYP. Lincoln, Nebraska, has elementary schools ranging from less than five percent to fifty-six-percent turnover per year.⁵³ That is, students who are tested in one year are substantially unlikely to comprise the group that is tested in a second year. Furthermore, while Nebraska's average teacher retention is approximately twelve years,⁵⁴ public

<http://writ.news.findlaw.com/ramasastry/20021204.html>.

45. *Id.*

46. Terry Nagel, GreatSchools.net, What No Child Left Behind Means for Your Child (2005), at <http://greatschools.net/cgi-bin/showarticle/az/205>. A little-known provision in NCLB has stirred up additional political controversy. The military has the right to obtain lists from high schools of students' names, addresses and phone numbers for recruiting purposes, and must be granted the same access to schools that is given to college and business recruiters. Schools that do not comply risk losing federal aid. Ramasastry, *supra*, note 44.

47. Fusarelli, *supra* note 31, at 78–79 (2004).

48. *Id.* at 78.

49. *Id.*

50. 20 U.S.C. § 6311(b)(2)(I)(ii) (2000 & Supp. III 2004).

51. Fusarelli, *supra* note 31, at 78.

52. *Id.* at 78–79.

53. LINCOLN PUBLIC SCHOOLS, 2004–2005 STATISTICAL HANDBOOK 32–34 (2005), available at <http://www.lps.org/about/statistics/documents/stats04-05.pdf>.

54. See NEBRASKA DEP'T OF EDUC., STATISTICS AND FACTS ABOUT NEBRASKA SCHOOLS 1998–99 tbl. 25 (1999).

schools on reservation lands have historically had average teacher tenure rates of fewer than five years.⁵⁵ The NCLB provisions that allow students to move and teachers to be replaced may only exacerbate this uprooting process and will certainly invalidate measures of school “success” or “failure.” It will come as no surprise that educational research shows time and again that high student mobility rates are linked to lower test scores.

As a consequence, the more schools fail, the fewer resources and the less advantage they accumulate over time. But note this parallel pattern: that as more schools fail within a state, the more these states reduce test score standards. Ohio, Louisiana, Michigan, Texas, and Arizona have each done so.⁵⁶ Michigan redefined its AYP cutoff, from a seventy-five-percent passage rate on standardized tests, to a forty-two-percent passage rate.⁵⁷ In one fell swoop, Michigan reduced its number of failing schools from 1513 to 216.⁵⁸

Another Catch-22 is the assumption that a state’s standardized tests are aligned to the actual curriculum taught in local schools. English and Steffy argue that this is seldom true and conclude that the latent effect of high stakes testing is to “flatten the curriculum,” reduce diversity, reward minimum performance with commendations, and reduce initiatives to engage in reform.⁵⁹ As the five largest NCLB testing agencies overlap with major textbook publication houses, their corporate connections will likely narrow the curriculum to those texts that support the tests (and vice versa).⁶⁰ Omi and Winant remind us that these macro-relationships can obscure the long-term effects of whiteness on reducing the multicultural education curriculum movement and constraining critical and radical pedagogies in the classroom.⁶¹

55. HELEN A. MOORE, *THE EDUCATIONAL STATUS OF NATIVE AMERICANS IN NEBRASKA* 27 (1999).

56. Fusarelli, *supra* note 31, at 82.

57. *Id.*

58. *Id.*

59. FENWICK ENGLISH & BETTY STEFFY, *DEEP CURRICULUM ALIGNMENT* 13 (2001).

60. See Public Broadcasting Service, *The Testing Industry's Big Four*, at <http://www.pbs.org/wgbh/pages/frontline/shows/schools/testing/companies.html> (last visited Mar. 21, 2005).

61. OMI & WINANT, *supra* note 2, at 74–75.

Finally, note that only half of the funds Congress identified as necessary have been allocated to NCLB.⁶² William Mathis estimates that NCLB could cost states about ten times as much as they receive from the federal government to implement the testing and programs required under NCLB.⁶³ This means that each school “failure” shifts an ever-restricting budget base toward other “successful” schools. This is the free market model of education and school choice at its financial meanest.

II. SCHOOL TESTING AND WHITENESS

Padilla and Lindholm argue that standardized quantitative testing in educational research relies on Eurocentric social and physical science models that assert paradigms that are nomothetic or universal.⁶⁴ Research conclusions in these models about school testing assume that using similar measures to compare groups of people across culture, ethnic groups or languages is legitimate and can uncover universal principles.⁶⁵ The problem for testing “arises when ‘biased’ instruments that favor white middle class males are used in a comparative research framework to examine differences between racial or ethnic groups.”⁶⁶

62. THOMAS W. FAGAN & NANCY L. KOBER, CTR. ON EDUC. POLICY, WHO’S GAINING, WHO’S LOSING AND WHY (2004):

While both sides in the funding debate are clearly being selective in their use of data, estimates from various sources do suggest that funds for implementation fall substantially short of the necessary ones, especially for Title I. The current Title I appropriation of 12.3 billion is only about half of the 24.7 billion it would take to serve all children counted under the law’s basic formula, using the law’s own expenditure factors. . . . President Bush and Congress must substantially increase funding for NCLB if the goals of the Act are to be realized.

63. William J. Mathis, *No Child Left Behind: Costs and Benefits*, 84 PHI DELTA KAPPAN 679, 682 (2003).

64. Amado Padilla & Kathryn Lindholm, *Quantitative Educational Research and Ethnic Minorities*, in HANDBOOK OF RESEARCH ON MULTICULTURAL EDUCATION 98 (Banks & Banks eds., 1995).

65. *Id.*

66. *Id.*

In the case of high stakes testing, Amrein and Berliner document the disproportionate use of high stakes tests in states with higher proportions of African-American and Latino enrollments.⁶⁷ Indeed, none of the states with the ten lowest proportions of African-American students have instituted high stakes testing or high school graduation exams.⁶⁸ In contrast, states with the top ten populations of African-American students have high stakes testing and all but one of those ten states have established high school graduation exams.⁶⁹

Sociologists have pointed out the latent functions of using standardized testing to generate educational inequality on numerous occasions.⁷⁰ In essence, the pervasive use of standardized tests, including IQ tests, Graduate Record Exams (GREs), and LSATs, provides educators, policy makers and admissions committees with the tools to tell parents that their individual students are failures. Linda McNeil asserts that Texas's accountability system views high dropout rates among minority students (largely Latino, black, poor and new immigrant children) as the manifest outcome of the process, rather than an unintended consequence.⁷¹

Rather than youths failing schools, schools are failing our minority youths through the TAAS system of testing. . . . There has been much analysis of the test score numbers, including analyses disaggregated by race. However, such studies have relied primarily on numbers provided by the state education agency, and/or school districts . . . by analysts employed by, or on contract to, the state or employed by

67. AUDREY L. AMREIN & DAVID C. BERLINER, ARIZ. STATE UNIV., THE IMPACT OF HIGH-STAKES TESTS ON STUDENT ACADEMIC PERFORMANCE (2002), available at <http://asu.edu/educ/epsil/EPRU/documents/EPSSL-0211-126-EPRU.pdf>.

68. *Id.* at 12.

69. *Id.*

70. See, e.g., Jane R. Mercer, *Latent Functions of Intelligence Testing in the Public Schools*, in THE TESTING OF BLACK STUDENTS 77 (Lamar P. Miller ed., 1974).

71. LINDA MCNEIL, CONTRADICTIONS IN SCHOOL REFORM (2000); see also Steven P. Klein et al., *What Do Test Scores in Texas Tell Us?*, 8 EDUC. POL'Y ANALYSIS ARCHIVES 49 (2000), at <http://epaa.asu.edu/epaa/v8n49>.

organizations with continuing state contracts . . . for TAAS implementation, TAAS consulting or training⁷²

NCLB has the latent function of empowering the federal government to tell school districts that their schools are failures when their students differ from the testing norms, and that the government will no longer invest in them as such. Sociologists have challenged the scientific racism of *The Bell Curve* and the use of biased tests to label students as “hyphenated learners” (gifted, educable mentally retarded, developmentally disabled, etc.).⁷³ We should ask the journalistic questions: Whose tests? Whose science? What knowledge? I would add: Where will “test score equality” take us educationally? James Traub of the New York Times investigated the frequent NCLB policy assertion that scientific testing is valid, reliable, fair, achievable, and that the consequences will lead to racial equality in achievement.⁷⁴ Traub found the phrase “scientifically based research” to occur more than one-hundred times in NCLB.⁷⁵ Apparently repetition yields truth.

Christine Sleeter argues that Euro-Americans in all spheres of social life (schools, the economy, the political arena) avoid examining white supremacy and the privileges we gain from white racism.⁷⁶ “We screen out what people of color try to tell us about white supremacy and our own role in reproducing it, because we fear losing material and psychological advantages that we enjoy.”⁷⁷

In systems of dominant whiteness, policy makers and educators learn to talk about race-related issues in ways that represent the status

72. Linda McNeil & Angela Valenzuela, *The Harmful Impact of the TAAS System of Testing in Texas*, in RAISING STANDARDS OR RAISING BARRIERS? 148 (Gary Orfield & Mindy L. Kornhaber eds., 2001). McNeil and Valenzuela conclude that there is a need to redress the “upside down system of accountability” and generate independent research on the economic and political forces that influence these systems of testing. *Id.* at 149.

73. See DAVID LEVINE & GARY PAINTER, *THE NELS CURVE: REPLICATING THE BELL CURVE* (1998), available at <http://repositories.cdlib.org/iir/iirwps/iirwps-066-98> (last visited Mar. 21, 2005); MEASURED LIES: THE BELL CURVE EXAMINED (Kinchloe et al. eds., 1997).

74. James Traub, *No Child Left Behind: Does it Work?*, N.Y. TIMES, Nov. 10, 2002, at 4A.

75. *Id.*

76. Christine E. Sleeter, *White Silence, White Solidarity*, in RACE TRAITOR 257, 258 (Noel Ignatiev & John Garvey eds., 1996).

77. *Id.* at 261.

quo as “neutral.” In doing so, they remove themselves from complicity and secure approval from other whites (voters, parents, school boards, etc.). Peggy McIntosh critiques these “neutral” conditions in which white privilege is unacknowledged, denied and protected by many whites.⁷⁸ These privileges include an invisible “package of unearned assets” that remain hidden.⁷⁹ In terms of high stakes testing, entire states with the highest proportions of white enrollments have ducked the costs and concerns of high stakes graduation testing.⁸⁰ What other “unearned assets” in the testing arena undermine the NCLB “scientific” approach to racial equality in education?

III. TESTING WHITENESS “OBJECTIVELY”

Testing and test scores generate an aura of “fairness” of race and gender neutrality. But the deep debates following the publication of *The Bell Curve*⁸¹ unmask the strategies of scientific racism in intelligence testing.⁸² In the processes of testing, our society creates a language of “objectivity” that obscures the privileges that systematically transfer to whites. In terms of race, we assume that who is tested, what is tested, and how tests are administered and interpreted have all been bathed by neutrality through “testing conditions” that include sterile classrooms, “expert” test givers, and the use of inanimate computers and “color blind” standards. Any slight variations in these testing conditions are trumped by our core belief in a free market meritocracy.

Sleeter reminds us that whites have the power to transmute many issues of racism into depoliticized questions of “cultural difference.”⁸³ In testing, if a culturally “loaded” item can be individually identified and removed or altered, then the test as a whole is considered “improved” and de-culturated. This approach

78. Peggy McIntosh, *White Privilege and Male Privilege, in RACE, CLASS AND GENDER* (Margaret L. Andersen & Patricia Hill Collins eds., 1992).

79. *Id.* at 71.

80. See AMREIN & BURLINER, *supra* note 67.

81. RICHARD J. HERRNSTEIN & CHARLES A. MURRAY, *THE BELL CURVE* (1994).

82. See *supra* text accompanying note 76.

83. Sleeter, *supra* note 76, at 259.

assumes that we normalize tests across equally weighted cultural vectors. That is, test takers assume that every cultural dimension in the test-taking population is equally likely to be represented in the normative sample of test items. This has not been the history of test development.

In *Other People's Children*, Lisa Delpit identifies multiple dimensions of the powers that are enacted in the classroom around testing and standards.⁸⁴ It begins with the power of teachers over students in their everyday lives as teachers reinforce dominant themes in curricula and testing.⁸⁵ The power of publishers of textbooks, curriculum developers, and test makers is less immediately visible in determining the worldview of the school and the norms enforced for students.⁸⁶ Indeed, the government first enforced compulsory schooling and now constructs testing standards that reaffirm the textbooks and tests and carefully screens teachers for accountability to these standards. Many individuals and political groups then have power to assess and determine a student's "intelligence" or "normalcy" and the outcomes for individual students and their schools.⁸⁷

When schooling prepares people for jobs, and the kind of job a parent has determines her or his child's access to schools, graduation from high school, admission to college and resulting economic statuses, then schooling is intimately related to that power. Delpit argues that access to institutional power is predicated upon acquisition of the culture of those groups who are in power.⁸⁸ Children from middle-class homes tend to do better in school because the culture of the school and the format of testing are based on the culture of those in power (in terms of both class and race).⁸⁹ Children from non-powerful families "operate within perfectly wonderful and viable cultures but not cultures that carry the codes or rules of

84. LISA DELPIT, *OTHER PEOPLE'S CHILDREN* (1995).

85. *Id.* at 181.

86. *Id.* at 132.

87. See LEON KAMIN, *THE SCIENCE AND POLITICS OF I.Q.* (1974).

88. DELPIT, *supra* note 84, at 28–29.

89. See Kathy Hytten & John Warren, *Engaging Whiteness: How Racial Power Gets Reified in Education*, 16 INT'L J. QUALITATIVE STUD. EDUC. 65 (2003).

power.”⁹⁰ Finally, those with power are frequently least aware of, or willing to acknowledge, its existence. Those with less power are often most aware of its existence. When de-emphasizing power, Delpit points to a group movement toward indirect communication.⁹¹ This means that relying on test scores, state standards, and NCLB will distort our discussions and debates about the role of whiteness in setting standards for “fair and objective” treatment.

For all of these reasons involving the creation and maintenance of power, standardized testing has gained the favor of policy makers in both state and federal government. Kornhaber argues that we are in the third decade of exam-driven education reform driven by national political momentum to develop new testing processes that will generate “world class standards.”⁹² Policy makers equate standard business approaches to education: they value quantitative, bottom-line figures that establish clear targets and rely on disincentive systems as motivators.⁹³ Many business leaders came on the bandwagon during the 1970s and early 1980s when “minimum competency tests” were established.⁹⁴ The 1983 *A Nation At Risk* study from the National Commission on Excellence in Education criticized low educational standards as a source of weakness for the national economy.⁹⁵ Today, twenty-three state legislative bodies demand that test scores be used in considering whether to award the high-school diploma.⁹⁶ In other states, these test scores are used in retention considerations for individual students along with the array of sanctions outlined by NCLB.

Mathis concludes that “whether education is or is not a market commodity, business model education reforms are ascendant. The theory is that an emphasis on efficiency, outcomes, the bottom-line

90. DELPIT, *supra* note 84, at 25.

91. Sleeter, *supra* note 76, at 259.

92. Mindy L. Kornhaber, *Appropriate and Inappropriate Forms of Testing, Assessment and Accountability*, 18 EDUC. POL’Y 45, 48 (2004).

93. *Id.* at 47–48.

94. *Id.* at 48.

95. See THE NATIONAL COMM’N ON EXCELLENCE IN EDUC., UNITED STATES DEP’T OF EDUC., *A NATION AT RISK* 1–3 (1983).

96. JAY P. GREEN & MARCUS A. WINTERS, *PUSHED OUT OR PULLED UP?* (Manhattan Institute, Education Working Paper No. 5, 2004), available at http://www.manhattan-institute.org/html/ewp_05.htm.

and tough minded business management will result in more efficient schools.”⁹⁷ School-level and state-level test scores are now publicly available as the educational “bottom line” for funding, and class and race are ignored as signals of the racial formations Omi and Winant warn against. The narrower use of test assessments (avoiding more qualitative paradigms of authentic assessment models) is cheaper and more efficient because standardized testing is corporately packaged, promoted and published.

IV. LEGAL TESTS, LEGAL SIGNS OF WHITENESS

So why engage law students, legal scholars, and current or potential members of the judiciary in these schooling debates? In my past work with *Larry P. v. Riles*,⁹⁸ it became clear that courts have been nibbling around the edges of the biases of whiteness and testing. While, in earlier decades, *Lau v. Nichols* focused on linguistic biases in tests,⁹⁹ *Larry P.* challenged the cultural bias of tests and the invalid application of standardized tests for placement (or displacement) of minority students.¹⁰⁰ Court involvement in the current debates over college admissions and test scores underline the legal elements of these educational debates.

As Arthur Coleman and Scott Palmer summarize, there are two overarching federal doctrines that apply to the test standards reform movement and policies such as No Child Left Behind.¹⁰¹ The first is due process,¹⁰² and the second is nondiscrimination.¹⁰³ As Coleman and Palmer point out, “[f]ederal due process protections apply only to public entities, such as states, school districts, and public universities,

97. Mathis, *supra* note 63, at 679.

98. 343 F. Supp. 1306 (N.D. Cal. 1972) (preliminary injunction), *aff'd*, 502 F.2d 963 (9th Cir. 1974), 495 F. Supp. 926 (N.D. Cal. 1979) (decision on merits), *aff'd in part, rev'd in part*, 793 F.2d 969 (9th Cir. 1984).

99. 414 U.S. 563 (1994).

100. 495 F. Supp. at 965–66.

101. ARTHUR COLEMAN & SCOTT PALMER, STANDARDS REFORM, FEDERAL LAW, AND THE AMERICAN DIPLOMA PROJECT 4–8, at <http://www.edroundtable.state.in.us/pdf/adp/ADPLegalStudy.pdf> (last visited Feb. 24, 2005).

102. *Id.* at 4–5, 7–8.

103. *Id.* at 5–8.

while federal nondiscrimination laws, through various mechanisms, may apply to both public and private actors”¹⁰⁴

Federal laws such as Title VI of the 1964 Civil Rights Act, alongside the Fourteenth Amendment, generally prohibit the “different treatment” of persons based on race, national origin or other suspect classifications.¹⁰⁵ These statutes and their implementing regulations “also prohibit policies or programs that are neutral on their face but have the effect of discriminating in that they have an unjustified ‘disparate impact’ by race, national origin or other protected classification.”¹⁰⁶

As these authors note, the “disparate impact” standard in an educational context is “designed to distinguish discrimination from disparity alone.”¹⁰⁷ When we ponder the NCLB requirements for sanctioning schools, or the resulting patterns of dropout rates, the questions we ask should focus on the disproportionate impact on minority students. As Coleman and Palmer state, the disparate impact standard requires that we ask: (1) Does the policy result in a significant disparity in the distribution of benefits by race or national origin? (2) If so, is the policy educationally justified? (3) If so, is there an alternative policy that would equally serve the institution’s goals with lesser disparity?¹⁰⁸

Larry P. raised the question of whether test scores standardized on white, middle-class populations can be used as an educational justification and whether these are valid indicators of learning for poor and minority students.¹⁰⁹ The district court also noted that, historically, testing experts have revised standardized intelligence tests when sample tests yielded different test scores for boys and

104. *Id.* at 4.

105. U.S. CONST. amend. XIV; Title VI of the Civil Rights Act of 1964 prohibits discrimination on the basis of race, color, or national origin by federal fund recipients (which includes school districts and all public and nearly all private universities). 42 U.S.C. § 2000d (2000).

106. COLEMAN & PALMER, *supra* note 101, at 6.

107. *Id.*

108. *Id.*

109. 495 F. Supp. 926, 979 (N.D. Cal. 1979) (describing discriminatory effect of IQ test for particular placement), *aff’d in part, rev’d in part*, 793 F.2d 969, 980 (9th Cir. 1984) (discussing the “difference of IQ predictability for black as opposed to white students”).

girls.¹¹⁰ No such systematic modifications have been made by testing companies on the basis of racial categories.¹¹¹ Test companies are privately held and nationally unregulated and unaudited; largely, they lack accountability.

I bring to this discussion of whiteness and testing four criticisms of standardized tests. In each criticism, I examine the technicalities of test construction, the court issues in *Larry P.*, as well as the political framework provided by critical race theorists, who consider tests as white (middle class and male) purveyors of privilege.

A. Test Scores Are Mutable

Test scores, even IQ tests and other achievement scores, can and do change substantially for individual students over time and across testing conditions.¹¹² All intellectual, standardized test scores, including IQ tests scores, are mutable (i.e., changeable, inconsistent, unsettled, capricious, alterable). Why else would some higher education campuses encourage, and others discourage, potential applicants to retake standardized exams such as the GRE and LSAT? Indeed, how could major test preparation corporations even exist were it not for the potential “retake” of standardized tests?

How do we technically account for this? Most test specialists recognize that the conditions under which tests are administered may vary or contain stress. But generic stress is considered a reasonable condition for testing. There is a language for “test anxiety” and it is assumed that it is up to the individual to control the stressors.¹¹³ The ability to control stress and reduce its impact on our tests may be a good predictor of the ability to “work under fire”—which is probably why some professional schools rely on cutoffs for the MCAT, LSAT, and GRE. Perhaps we want to know something about the individual’s ability to “gut it out” in a high stakes situation.

Motivation is often considered a screening factor at the point of taking a test—particularly a college or professional school

110. *Id.* at 955, 957.

111. *See id.* at 956, 957 & n.64.

112. KAMIN, *supra* note 87, at 175 (1974) (stating that “[t]he data have repeatedly demonstrated profound environmental effects on IQ scores”).

113. That is, get a good night’s sleep, practice the test conditions and formats, etc.

admissions test. That is, if you are not motivated to apply, you will not take the test. However, we know that certain forms of test motivation become well patterned across testing even at the occasion of our first standardized test, typically taken in the first grade. NCLB requires testing by the third grade.¹¹⁴ In some states (Missouri, for example), you are required to participate even if you are a recently immigrated ELL student who arrived in school yesterday.¹¹⁵

We have a lengthy and contradictory research history on the issue of tests and motivation for students. Fordham and Ogbu assert that African-American students in general develop a collective identity that rejects school activities (including attendance and testing) that they regard as characteristic of a “whiteness that effaces their identity and culture.”¹¹⁶ Achievement in schools and on tests may be viewed as threatening identity with social peers and may significantly reduce motivation and achievement for racial/ethnic minority groups.¹¹⁷

Sociologists understand that motivation can be suppressed by conditions of threat (e.g., a threat to identity, to your school, to your ability to be credentialed). Claude Steele’s classic research on “stereotype threat” traces the consequences for minority students who take tests in a threat setting in which their social/racial identity (and gender) significantly reduces test performance outcomes.¹¹⁸ After adjusting for differences in skill levels, he found that black Stanford University students solved fewer verbal problems than white students when the test was presented as diagnostic of group ability.¹¹⁹ But blacks performed just as well as whites when the same test was presented as not diagnostic of ability.¹²⁰ These controlled experiments were conducted with high-functioning college students matriculating

114. See 20 U.S.C. § 6311(b)(3)(C)(v).

115. See, e.g., MO. DEPT OF ELEMENTARY & SECONDARY EDUC., SERVING UNDOCUMENTED ENGLISH LANGUAGE LEARNERS, available at <http://www.dese.state.mo.us/divimprove/fedprog/discretionarygrants/bilingual-esol/svngundocell.pdf> (last visited May 26, 2005) (“No ELL student should be exempted from taking the MAP [Missouri Assessment Program Test] after enrolling in a Missouri school. The NCLB Act of 2001 requires that all ELL students be given the MAP.”).

116. Signithia Fordham & William Ogbu, *Black Students and School Success*, 18 URB. REV. 176, 176–206 (1986).

117. *Id.* at 176.

118. Claude M. Steele, *A Threat in the Air*, 52 AM. PSYCHOLOGIST 613, 616 (1997).

119. *Id.* at 620.

120. *Id.*

at Stanford University. What would be the outcomes for students in poorly funded public schools with high teacher- and student-turnover rates and under the threat of federal sanctions?

Despite recognizing this threat condition, students are generally still encouraged by most schools to participate in standardized testing whether for assessment or admissions purposes. The Law School Admissions Council (LSAC) website encourages minority students to participate in the LSAT test process because:

Standardized Tests are Everywhere

It's unlikely that standardized tests will be eliminated in the near future. Look at the large role such tests play in society, from teacher accreditation to other professional certifications. You are likely to see more standardized tests in your professional lifetime.

Rather than fearing testing, you should develop test-taking strategies now that may serve you well in the future. If you know you have difficulty in testing, talk to your college advisors or counselors about your problem. Many schools hire professionals specifically to help students develop better test-taking skills.¹²¹

Given findings that test scores are changeable over time for individuals under varying conditions of threat, motivation and stress (or the systematic removal of threats, motivators or stressors) the American Educational Research Association (AERA) argues that *no* test should be a single-stake—or high-stake—test for individuals or schools.¹²²

B. Test Content Can Be Biased and Invalid

Most of us are familiar with the concept of test bias, the notion that individual items on a test may be systematically less familiar to some cultural groups in a given society. As early as the 1930s,

121. LSAC, *Minority Perspectives* (2005), at <http://lsac.org/lsac.asp?url=/lsac/minorities-in-legal-education-selected-articles.asp>.

122. AERA, *Position Statement, High-Stakes Testing in PreK-12 Education* (July 2000), at <http://aera.net/policyandprograms/?id=378>.

George Sanchez provided a systematic critique of the cultural bias in tests.¹²³ Kamin also identifies elements of IQ tests that are imbedded in the experiences of children whose schools and families provide systematic access to cultural artifacts that are valued by the makers of tests.¹²⁴

Other biases come from unequal educational opportunities and access to curricula that prepare for positive testing outcomes. High proportions of low-income and minority students are enrolled in urban or rural public schools that disproportionately lack adequate facilities, enough teachers, and proper teaching materials, in addition to having a host of other socio-economic problems that their suburban counterparts may not have. A joint study performed by Columbia University and Michigan State University raises concerns about high stakes tests and their potential to further exacerbate already substantial inequities in schooling outcomes.¹²⁵ One author of the study concludes: “I am opposed to only holding kids accountable and not holding anyone else accountable. Right now, only kids—and particularly poor and minority kids—are bearing the burden.”¹²⁶

After several decades of criticism, major testing organizations have moved to two solutions to counter claims of test bias. The first is to remove specific test items that over time are deemed culturally “biased” due to distinctive response sets in some sub-groups.¹²⁷ The second approach is to inject items that are more “abstract and conceptual”—i.e., that provide analytic forms of standardized tests relying on spatial relationships or logical consequences.¹²⁸ In the end, testing organizations have not been held accountable for creating tests that are normed on the full diversity of school matriculants. Such a test, by statistical definition, would not yield divergent average test scores and standard deviations by categories of students.

123. GEORGE SANCHEZ, *THE FORGOTTEN PEOPLE* (1940).

124. KAMIN, *supra* note 87, at 176–77.

125. Gary Natriello & Aaron M. Pallas, *The Development and Impact of High-Stakes Testing*, in *RAISING STANDARDS OR RAISING BARRIERS?*, *supra* note 72, at 19–38.

126. Katy Anthes, Casenote, *Competency Testing for High School Graduation*, ECS NOTES, May 2, 2000, <http://ecs.org/clearinghouse/13/88/1388.htm>.

127. GUILLERMO SOLANO-FLORES & SHARON NELSON-BARBER, WESTED, *ASSESSING THE CULTURAL VALIDITY OF SCIENCE AND MATHEMATICS ASSESSMENTS* (2000), at http://edgateway.net/cs/cvap/print/docs/cvap/pub_3.htm.

128. *Id.*

Instead of placing the burden of proper test development on testing corporations, NCLB sanctions have been placed on schools that are already underfunded.

Additionally, critics charge that the culturally specific content is set according to the standards of external test “experts” who do not know the day-to-day curriculum goals in a specific classroom.¹²⁹ These universal “standards” can bias results for individual students or groups of students. Kamin argues that test items are normed on the culture of those who construct the test rather than the school curriculum as represented reliably and validly by the test items.¹³⁰ Students whose family background and experiences diverge from those test-defining norms will reduce the probability that the student “passes” for graduation, or contributes a “proficient” level score to his or her school.¹³¹

When standards are set far away (and far from the local classroom or community culture), Kornhaber cautions that these do not necessarily “get into the classroom or into the curriculum in the expected ways.”¹³² These standards may be “slow to seep in” and may not mesh with the concerns of local schools and teachers and communities.¹³³ Bartolome and Macedo argue that policymakers and public opinion shapers reward “the demonizing of other cultural subjects” and promote an “a-critical education.”¹³⁴ They challenge the notion of testing one “American common culture . . . in view of the quasi-apartheid conditions that have relegated American Indians to reservations, created ghettos, and supported the affirmative action of red-lining”¹³⁵ Thus, testing is an arena of cultural dominance and racism that hides behind the equity issues asserted by NCLB.

The validity of state achievement tests such as the Texas Assessment of Academic Skills (TAAS) test is questionable, given

129. See, e.g., KAMIN, *supra* note 87; SADKER & SADKER, *supra* note 32, at 248–49.

130. KAMIN, *supra* note 87.

131. *Id.*

132. Mindy L. Kornhaber, *Appropriate and Inappropriate Forms of Testing, Assessment and Accountability*, 18 EDUC. POL’Y 45, 52 (2004).

133. *Id.*

134. Lilia I. Bartolome & Donaldo P. Macedo, *Dancing with Bigotry*, 67 HARV. EDUC. REV. 222, 230 (1997).

135. *Id.* at 229–30.

that increases in state scores are not reflected in increases in the National Assessment of Educational Progress or on college entrance exams or advanced placement tests.¹³⁶ Although test scores on state-administered tests usually increase somewhat after high stakes testing policies are implemented,¹³⁷ evidence suggests that students are learning the content of the state-administered test and perhaps little else.¹³⁸ This learning does not appear to have any meaningful carryover effect. Amrein and Berliner's assessment of a range of studies suggest that after high school graduation exams are implemented, state scores on the SAT and ACT actually decrease for those students.¹³⁹ Thus, preparing for one set of exams does not correlate with preparation for a different set of exams.

NCLB may serve the purpose of narrowing course curriculum content to "teach to the test." Potentially, few states under NCLB will ever test in student skills in foreign language, art, music, poetry, political science, history, psychology, family studies, etc. "Teaching to the test" will be considered in more detail below, but the problem of rote instruction to the test has cropped up in numerous cities as teachers respond to the threat of "reconfiguration." Teachers have returned to rote drill in Chicago to ensure that test materials are highlighted.¹⁴⁰ In Texas, several schools serving poor and minority students changed their instruction on writing essays to parallel the TAAS.¹⁴¹ Students in those schools learned to write only one form of essay, short passage reading and responses in line with the test requirements.¹⁴² Their study found that students were less able to read extended literature and the classroom content was lost.¹⁴³

Amrein and Berliner conclude that early research on high stakes tests demonstrates mostly negative consequences on student achievement, including a growth in the gap between minority and

136. SADKER & SADKER, *supra* note 32, at 247.

137. *Id.*

138. See AMREIN & BURLINER, *supra* note 67, at 56–57.

139. *Id.* at 57

140. Brian Jacob & Steven Levitt, *Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating*, 118(3) Q.J. OF ECON. 843–77 (2003).

141. Linda McNeil & Angela Valenzuela, *The Harmful Impact of the TAAS System of Testing in Texas*, in RAISING STANDARDS OR RAISING BARRIERS?, *supra* note 72, at 127, 134.

142. *Id.* at 133.

143. *Id.*

non-minority students in school retention and graduation.¹⁴⁴ Other researchers have noted that standardized testing exerts negative effects on the academic achievement of poor and minority youth in particular.¹⁴⁵ This operates primarily through increased drop-out rates and lack of retention—that is, voluntary or involuntary removal of large numbers of students from the pool of test takers.¹⁴⁶

In my own research on school success and the predictive ability of standardized tests such as the GRE, I found that test scores do not predict beyond a minimal set of educational outcomes. Across the nation, GRE scores for sociology graduate students predict only their first-year theory and methods-course grades. These scores specifically did *not* predict: (1) grades beyond first year; (2) who was accepted as a graduate teaching or research assistant; or (3) who completed the degree. Educators all have stories of students with high admission-test scores and disappointing performance. But these test scores are used with a discretionary outcome not explicit to the exam. Minority students with test scores equivalent to or higher than those of their peers are significantly less likely to be invited onto a research team, funded as a graduate research assistant, or credited as a co-author in professional papers.¹⁴⁷

It is important to remember that in *Larry P.* the court cited the cultural bias in tests, and then raised concerns about the biases in other more subjective evaluations as well.¹⁴⁸ Even when adding multiple indicators such as teacher evaluations, or perhaps building a team with social workers or psychologists, you cannot erase the potential biases in a single evaluation.

144. See generally AMREIN & BERLINER, *supra* note 67.

145. See Edgar Epps, *Situational Effects in Testing*, in THE TESTING OF BLACK STUDENTS 41–51 (LaMar Miller ed., 1974); LEVINE & PAINTER, *supra* note 73.

146. MARGUERITE CLARKE ET AL., BOSTON COLLEGE, HIGH STAKES TESTING AND HIGH SCHOOL COMPLETION (2000), available at <http://bc.edu/research/nbetpp/publications/v1n3.html>.

147. Helen Moore & Bruce Keith, *Human Capital, Social Integration and Tournament Models: A Test of Graduate Student Success*, 23 AM. SOC. 52, 52–71 (1992).

148. 495 F. Supp. 926, 959–60 (N.D. Cal. 1979).

C. Testing Itself Is a Culturally Specific Process

Testing, as a process, may intersect with mutable test scores and biased and invalid test content because test-taking itself (not the content) is an acquired individual skill. When and how are we taught to take tests and to be motivated to test well? We know that most students begin taking standardized tests as early as the first grade, but some students are prepared by their families to “answer” and quickly recall facts (whether to learn their numbers, or to identify non-English words for objects in their homes). We do not know all of these answers, but we do know that test-taking techniques can be taught and manipulated. At least one company guarantees it.¹⁴⁹

Delpit highlights these codes or rules for participating in power in schools. She describes in rich detail a “culture of power” that includes testing, communicative strategies, ways of talking, dressing and interacting.¹⁵⁰ The rules of the “culture of power” reflect the culture of those who have power—it is their language and their valued cultural objects that are represented and tested.¹⁵¹ Delpit is concerned that students and parents who are not already participants in the culture of power are told explicitly that following the rules of that dominant culture makes acquiring power easier.¹⁵²

Each of these power dynamics applies to “tests” and high stakes evaluations in the classroom as well. Numerous demonstrations have shown that test-score gaps can be reduced substantially by intensive one-on-one coaching. Edgar Epps and others have consistently reduced the gap between minority and non-minority scores by teaching underlying test processes, not content.¹⁵³ First, they increase student motivation with positive material rewards (in contrast to the looming negative sanctions of NCLB).¹⁵⁴ Second, they teach test-taking techniques (techniques similar to those in corporate test

149. See Kaplan Test Prep and Admissions, Law (2005), at http://www.kaptest.com/repository/templates/LevMInitDroplet.jhtml?_levMParent=/www/KapTest/docs/repository/content/Law/Law_Main/Law_Main.

150. DELPIT, *supra* note 84, at 25.

151. *Id.* at 45.

152. *Id.* at 32.

153. Edgar G. Epps, *Situational Effects in Testing*, in *THE TESTING OF BLACK STUDENTS* 41, 41–51 (Lamar Miller ed., 1974).

154. *Id.* at 46.

preparation classes that are available to students in families that can afford the tuition).¹⁵⁵ Finally, as role models for their racial/ethnic minority groups, these minority researchers provided same-ethnicity leadership and confidence that the materials were knowable.¹⁵⁶ They “raised” test scores to close those gaps that NCLB targets. Why are these research-based strategies ignored in federal policy?

D. Invalidity of Tests as a Marker of “Learning” and “Teaching” at the School Level

The AERA states that tests that are validated for one use in one setting or for one purpose may be invalid for another.¹⁵⁷ That is, each separate use of a high stakes test (for individual graduation, school evaluation, curricular reform, teacher assessment) “requires a separate evaluation of the strengths and limitations of both the testing program and the test itself.”¹⁵⁸ Validation requires a high number of cases, over time, with long-term outcome indicators. This standard has not been applied in NCLB.

Kornhaber discusses the potential for teachers, schools, and districts to engage in NCLB “gaming.”¹⁵⁹ That is, test scores can be changed by manipulating conditions unrelated to teaching a curriculum or improving generalized learning.¹⁶⁰ These strategies might include “changing the pool of test takers to weed out those that are struggling.”¹⁶¹ She points to the noticeable increase in the proportions of students grade-retained in the year prior to taking a high stakes test in the states of Texas and Massachusetts.¹⁶² Retaining a student provides a second chance to cover materials and remove the student from the test for a particular grade level.¹⁶³ Educators know

155. *Id.*

156. *Id.* at 44.

157. AERA, Position Statement, High-Stakes Testing in PreK-12 Education (2000), at <http://aera.net/policyandprograms/?id=378>.

158. *Id.*

159. Kornhaber, *supra* note 132, at 56.

160. *Id.*

161. *Id.*

162. *Id.*

163. *Id.*

that retention in a grade is the variable most highly correlated to dropping out, rather than with increased learning.¹⁶⁴

A second form of “gaming” is to reset the bar. As more schools “fail,” more states reduce standards.¹⁶⁵ Michigan redefined its cutoff and reduced its number of “failing schools.”¹⁶⁶

Less obvious (and more desperate) are “gaming” efforts by teachers and administrators who coach their children directly to test items, or who cheat on testing procedures in order to increase school test scores.¹⁶⁷ Cizek cites examples of school principals who give vocabulary words in their morning announcements.¹⁶⁸ This is not a bad idea, but immoral when the principal holds the vocabulary section of the state-mandated test in hand.¹⁶⁹ Cheating by teachers might involve failing to monitor students during tests, or encouraging students who test poorly to be absent on test days.¹⁷⁰ Others “pre-screen” answer sheets for “stray marks.”¹⁷¹ In Austin, Texas, school administrators entered incorrect ID numbers on the answer sheets of low-scoring students, which invalidated those scores.¹⁷² The largest cheating scandal occurred in the New York City school district.¹⁷³ One teacher had students write answers on a separate sheet of paper and then corrected answers before they bubbled their official answer sheets.¹⁷⁴ Others wrote test questions and answers on the chalkboard.¹⁷⁵ Ten teachers were ultimately recommended for termination.¹⁷⁶

164. *Id.*

165. *See supra* text accompanying note 56.

166. *See supra* text accompanying notes 54–58.

167. *See* Kornhaber, *supra* note 132, at 57; Gregory J. Cizek, *Cheating to the Test*, EDUC. NEXT (2001), available at <http://www.educationnext.org/2001sp/40.html> (“The practices included giving extra time on timed tests, changing students’ answers, suggesting answers to students, and directly teaching specific portions of a test. More flagrant examples included teachers’ giving their students dictionaries and thesauruses for use on a state-mandated writing test.”).

168. Cizek, *supra* note 167.

169. *Id.*

170. *Id.*

171. *Id.*

172. *Id.*

173. *Id.*

174. *Id.*

175. *Id.*

176. *Id.*

Important consequences for teachers include undermining their professional autonomy and commitment. Hinde found that elementary school teachers recognized that the widening disparity between the privileged and non-privileged students was a result of biased assessments and contradictory to sound education practices.¹⁷⁷ Teachers in her research had strong objections to a testing system that undermined their professional control of the curriculum and the classroom.¹⁷⁸ These participating teachers were amenable to recognizing the testing standards as valid tools for lesson and unit planning, but they were offended by the threat posed to their instructional autonomy by the standardized tests.¹⁷⁹ Like other professionals, teachers feel that they alone can understand fully the inner workings of their professional turf: the classroom. Likely they view non-practitioners (especially legislators) with suspicion for imposing their lay philosophies and methods on them. One participant teacher in the study summarized the mixed feelings about the standards and accompanying tests: "I think they [standards] give you a starting point, but there's more to it than just the standards. I mean it all has to go with your school and your principal and your teachers and your team."¹⁸⁰

The national AERA High Stakes Testing policy states:

[I]f high stakes testing programs are implemented in circumstances where tests lack sufficient reliability and validity for their intended purposes, there is potential for harm. Policy makers and the public may be misled by the spurious test score increases unrelated to any fundamental educational improvement; students may be placed at increased risk of educational failure and dropping out; teachers may be blamed or punished for inequitable resources over which they have no control; and curriculum and instruction may be severely

177. Elizabeth R. Hinde, *The Tyranny of the Test*, 6(10) CURRENT ISSUES IN EDUC. (2003), <http://cie.asu.edu/volume6/number10>.

178. *Id.*

179. *Id.*

180. *Id.*

distorted if high test scores per se, rather than learning, become the overriding goal of classroom instruction.¹⁸¹

In summary, we have learned from prior court cases that standardized tests have clear biases.¹⁸² Researchers continue to demonstrate that test score gaps are not necessarily indicators of reduced individual motivation or intelligence, and that schools have the capacity to “game” test score results to their favor.¹⁸³ The discriminatory educational outcomes from NCLB for minority students (and perhaps for schools) can be challenged on a constitutional basis.

V. FUTURE CHALLENGES FOR LEGAL/POLICY CONCLUSIONS

The overuse and abuse of biased assessments in NCLB becomes one more obstacle that blocks the path of vulnerable students and communities. As Omi and Winant caution us, the racial formation of NCLB has been submerged under the twin agendas of business accountability and invalid science of assessment.¹⁸⁴ Over the long term, standardized testing creates inequities, widening the gap between the quality of education for poor and minority youth and that of more privileged students.

This debate over the construction of whiteness and testing should involve communities, parents, teachers, school administrators, school boards and politicians. Natriello argues that the goals, costs and processes of high stakes testing have not been clearly thought through in order to create a valid high stakes system.¹⁸⁵ Stakes need to be kept high for politicians and educators, not just students and their communities.¹⁸⁶ “Those who promote high stakes testing should be held accountable for providing proper analysis as well as the necessary educational opportunities to ensure student success.”¹⁸⁷ Teacher training and professional development relating to the new

181. AERA, *supra* note 157.

182. *See supra* Part IV.B.

183. *See supra* note 167 and text accompanying notes 159–76.

184. OMI & WINANT, *supra* note 2, at 75.

185. Natriello & Pallas, *supra* note 125, at 19–38.

186. *See supra* text accompanying note 126.

187. Anthes, *supra* note 126.

tests, administrative experience in analyzing and applying school and individual test data, and extra funding for remedial tutoring and programs are necessary conditions of high stakes testing programs.

Bartolome and Macedo find we have politically mandated:

a pedagogy of entrapment that makes it undemocratic to argue against school choice . . . a discourse that brooks no dissension or argument, for to argue against it is to deny democracy. The hidden curriculum of school choice consists of taking resources from poor schools that are on the verge of bankruptcy to support private or well-to-do public schools.¹⁸⁸

Hunter and Bartee further conclude that closing the achievement gap among racial/ethnic groups does not require NCLB's testing, competition for resources and forcible implementation of "school choice."¹⁸⁹ These researchers argue that the last time we saw systematic reduction in test score gaps across racial/ethnic groups was during those decades following the war on poverty, accomplished with initiatives to reduce discrimination in the larger society.¹⁹⁰ These policies were based on the civil rights perspective that schools and society "were interrelated, not separate, functioning domains."¹⁹¹

Instead, NCLB policies propose to equalize outcomes for all students primarily through sanctions that redistribute rewards to those who test well. Schools and their students that fail to meet these standards will be punished through sanctions such as replacing staff, transplanting students, and transforming schools into magnets without increasing resources. These "quick fixes" may not erode the privileges of whiteness. Rather, they may indeed exacerbate inequalities among communities and their schools and students.

I am not convinced, nor should anyone be, that NCLB can withstand sharp legal and social questioning. First, we must look to those careful assessments of testing set out in early court cases and

188. Bartolome & Macedo, *supra* note 134, at 233.

189. Richard Hunter & RoSusan Bartee, *The Achievement Gap*, 35 EDUC. & URB. SOC'Y 151, 158-59 (2003).

190. *Id.* at 158.

191. *Id.*

learn from these legal challenges to racialized test biases. Policy makers have not demonstrated that today's tests are less biased, nor that the consequences of testing will not re-segregate students and entire schools and communities on the basis of these biases.

If by 2013–14, these test biases are overcome and all students are indeed “proficient”—with no categorical achievement gaps—how has the larger economy or even higher education been prepared to meet the challenge of a fully employable and educable citizenry? A disconnect exists between the proposed educational outcomes of full equality in test scores and our economic capacity to fully employ our students.

The short-term latent effect may be to manipulate students and teachers in primarily low-income/disadvantaged schools and populations to support vouchers and move away from their communities. Their alternative will be to cede to federal definitions of the curriculum and testing in order to compete for ever-scarcer resources. These national standards, as envisioned by profit-making text authors and test makers, will narrow everyone's educational autonomy, and devolve distressed schools toward basic education and reduced state standards. In *Larry P. v. Riles*, the appellate court found that “[E.M.R.] classes are conceived of as ‘dead-end classes’ and a misplacement in E.M.R. causes a stigma and irreparable injury to the student.”¹⁹² We may find the same outcome on a national scale for those schools and their students systematically left behind by NCLB.

192. 793 F.2d 969, 973 (9th Cir. 1984).