

2019

Don't Remind Me: Stereotype Threat in High-Stakes Testing

Arusha Gordon

Follow this and additional works at: <https://scholarworks.law.ubalt.edu/ublr>

Part of the [Law Commons](#)

Recommended Citation

Gordon, Arusha (2019) "Don't Remind Me: Stereotype Threat in High-Stakes Testing," *University of Baltimore Law Review*: Vol. 48 : Iss. 3, Article 3.

Available at: <https://scholarworks.law.ubalt.edu/ublr/vol48/iss3/3>

This Article is brought to you for free and open access by ScholarWorks@University of Baltimore School of Law. It has been accepted for inclusion in University of Baltimore Law Review by an authorized editor of ScholarWorks@University of Baltimore School of Law. For more information, please contact hmorrell@ubalt.edu.

DON'T REMIND ME: STEREOTYPE THREAT IN HIGH-STAKES TESTING

Arusha Gordon*

I. INTRODUCTION

Think back to the day you took the Scholastic Aptitude Test (SAT) or some other standardized test. Like most people, you probably took your seat, put your sharpened Number 2 pencils on the assigned desk and, after listening to instructions from a proctor, started the task of bubbling in your name, address, school, sex, gender, and other identifying information. You likely thought nothing of it—after all, if you are like most students in American schools, you probably grew up taking many tests in which you bubbled in your identifying information right before you got down to business and actually took the test. But what if you were told that the test was not only measuring your knowledge related to the tested material, but also your ability to navigate around mental processes which might tax your ability to focus solely on the test? What if you learned that the test environment, the students in the room, the test administrator, and even the process of bubbling in your identifying information could impact how well you performed?

For years, social psychologists have studied how these factors could impact the performance of members of different social identities through a process known as “stereotype threat.”¹ Stereotype threat refers to a psychological phenomenon in which a member of a negatively stereotyped group underperforms on an activity because of increased anxiety that they may confirm the negative stereotype.² This article examines the role of stereotype threat as it relates to racial and gender identities in high stake testing in educational settings. While there is a significant body of literature,

* Arusha Gordon serves as Counsel on the Voting Rights Project and the Stop Hate Project at the Lawyers' Committee for Civil Rights Under Law.

1. *See infra* Part II.

2. Robert J. Rydell, Allen R. McConnell & Sian L. Beilock, *Multiple Social Identities and Stereotype Threat: Imbalance, Accessibility, and Working Memory*, 96 J. PERSONALITY & SOC. PSYCHOL. 949, 949 (2009); *see also* Claude Steele & Joshua Aronson, *Stereotype Threat and the Intellectual Test Performance of African Americans*, 69 J. PERSONALITY & SOC. PSYCHOL. 797, 797 (1995).

primarily in the field of psychology, discussing the impact of stereotype threat on members of different social groups,³ and while there is a substantial body of literature discussing legal strategies through which advocates have challenged standardized testing,⁴ this article aims to help fill the space between these two bodies of literature. Namely, this article examines the role of stereotype threat in testing,⁵ and what advocates can do to reduce the impact of stereotype threat in high-stake standardized testing.⁶

Part II provides an overview of stereotype threat; what it is,⁷ factors which contribute to stereotype threat,⁸ the impact of stereotype threat,⁹ and who stereotype threat affects most significantly.¹⁰ Part III considers the legal landscape of standardized testing, with specific attention given to how research concerning stereotype threat might play a role in claims brought under the Equal Protection Clause, Title VI of the Civil Rights Act, and Title IX of the United States Education Amendments of 1972.¹¹ Part IV considers strategies addressing stereotype threat,¹² and discusses the science behind techniques to reduce the impact of stereotype threat on testing,¹³ as well as policy and advocacy approaches to addressing the issue.¹⁴

II. STEREOTYPE THREAT

A. *What is Stereotype Threat?*

Stereotype threat refers to the phenomenon in which a member of a negatively stereotyped group underperforms on an activity because of increased anxiety that he or she may confirm the negative stereotype.¹⁵ Many studies over the past few decades have explored

3. *See infra* Part II.

4. *See infra* Part III.

5. *See infra* Part II.

6. *See infra* Section V.B.

7. *See infra* Section II.A.

8. *See infra* Part II.

9. *See infra* Part II.

10. *See infra* Section II.B.

11. *See infra* Part II.

12. *See infra* Part IV.

13. *See infra* Section IV.A.

14. *See infra* Sections IV.B–C.

15. Steele & Aronson, *supra* note 2, at 797–98.

stereotype threat, leading to important findings concerning how the phenomenon operates.¹⁶

At first, the idea that a subliminal reminder of a stereotype—a stereotype you might consciously and vociferously reject—could cause underperformance on a task, may seem implausible. Yet, research shows that stereotype threat can affect people of all different races, ages, language groups, genders, and sexual orientations.¹⁷ To get a better understanding of stereotype threat, let's turn to a few examples:

i. African-Americans and Academic Tests

When taking a difficult academic test, similar to that of the Graduate Record Examination (GRE), Joshua Aronson and Claude Steele found that African-American college students underperformed when they were asked, prior to taking the test, to complete a questionnaire that included a question regarding their race.¹⁸

ii. Women and Math

Women taking a difficult math test underperformed when told that the results generally showed a gender difference, but performed just as well as their male colleagues when told the test did not show any gender difference.¹⁹ In a similar experiment, Asian women underperformed on a math test when they were given a questionnaire emphasizing their gender (i.e. “whether their [dorms] were coed or single sex” and “whether they preferred coed or single-sex [dorms]”) but performed more strongly when given a questionnaire

16. See *infra* Part IV. Although this article's focus is on stereotype threat, researchers have also identified the phenomenon of *stereotype lift*. Gregory M. Walton & Geoffrey L. Cohen, *Stereotype Lift*, 39 J. EXPERIMENTAL SOC. PSYCHOL. 456, 456 (2003). “When a negative stereotype impugns the ability or worth of an outgroup, people [in the ingroup] may experience stereotype lift—a performance boost that occurs when downward comparisons are made with a denigrated outgroup.” *Id.* As Walton and Cohen note in their meta-study on stereotype lift, studies involving SAT modified questions indicate that stereotype lift may produce a “50-point advantage on the SAT for White men—a performance boost that, at the most selective colleges, could make the difference between rejection and acceptance.” *Id.* at 473.

17. See *infra* Part IV.

18. Steele & Aronson, *supra* note 2, at 806–08.

19. Steven J. Spencer et al., *Stereotype Threat and Women's Math Performance*, 35 J. EXPERIMENTAL SOC. PSYCHOL. 4, 12–14 (1999).

emphasizing their race (i.e. “what languages they spoke at home” and “how many generations of their family have lived in America”).²⁰

iii. Athletic Ability and Race

In an experiment involving a task related to golf, white participants performed better when the task was described as being diagnostic of “sports intelligence” and worse when the task was described as testing “natural athletic ability.”²¹ Black participants performed better when the task was described as measuring natural athletic ability and worse when it was described as testing “sports intelligence.”²²

iv. Age

Older adults performing a memory task performed more poorly when they were exposed to a story or priming words reflecting negative stereotypes regarding aging, as compared with a control group of older adults who were exposed to positive stereotypes regarding aging and memory.²³

In short, stereotype threat has been well-established and can impact a range of social identities.²⁴ The question, then, is what factors contribute to or trigger stereotype threat and what can we do about them?

B. Groups Vulnerable to Stereotype Threat

Research indicates that stereotype threat is particularly relevant when individuals who are confident in their abilities in a certain domain are tested at a high level in that domain.²⁵ In the academic setting, this means that stereotype threat “affects confident students more than unconfident ones.”²⁶ This is because for individuals who more strongly identify with the domain, the relevance of the

20. Margaret Shih et al., *Stereotype Susceptibility: Identify Salience and Shifts in Quantitative Performance*, 10 AM. PSYCHOL. SCI. 80, 80–83 (1999).

21. Jeff Stone et al., *Stereotype Threat Effects on Black and White Athletic Performance*, 77 J. PERSONALITY & SOC. PSYCHOL. 1213, 1217–20, 1225–26 (1999).

22. *Id.*

23. Becca Levy, *Improving Memory in Old Age Through Implicit Self-Stereotyping*, 71 J. PERSONALITY & SOC. PSYCHOL. 1092, 1095–96, 1106 (1996).

24. *See supra* notes 18–23 and accompanying text.

25. *See* Claude M. Steele, *A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance*, 57 AM. PSYCHOL. 613, 617 (1997).

26. *Id.*

stereotype plays more of a threat.²⁷ In one experiment, Spencer, Quinn, and Steele selected women who were strong at math and expressed high confidence in their math abilities and gave them a challenging math test.²⁸ Even though these students identified as strong mathematicians, the female students, presumably due to stereotype threat, performed more poorly than their male counterparts; however, where the test focused on advanced literature skills, an area without strong gender-based stereotypes, women performed just as well as their male counterparts.²⁹ Women from this same group who took easier math tests did not underperform.³⁰ Spencer, Quinn, and Steele's research suggests that while individuals who are highly identified with a domain may not feel stereotype threat at lower levels of testing, when they are challenged in the domain with which they self-identify, stereotype threat may cause them to underperform:

For the advanced female math student who has been brilliant up to that point, any frustration she has at the frontier of her skills could confirm the gender-based limitation alleged in the stereotype, making this frontier, because she is so invested in it, a more threatening place than it is for the nonstereotyped. Thus, the work of dispelling stereotype threat through performance probably increases with the difficulty of work in the domain, and whatever exemption is gained has to be won at the next new proving ground.³¹

C. *Other Effects of Stereotype Threat*

In addition to causing underperformance by individuals who are members of traditionally marginalized groups, stereotype threat may also cause people who are subject to a negative stereotype to disassociate from domains in which they feel threatened.³² For instance, women may “disidentify” with math as an important domain, that is, [women may] avoid or drop the domain as an identity

27. *See id.*

28. *Id.* at 619.

29. *See id.*

30. *Id.*

31. *Id.* at 618.

32. *See* Brenda Major et al., *Coping with Negative Stereotypes About Intellectual Performance: The Role of Psychological Disengagement*, 24 *PERSONALITY & SOC. PSYCHOL. BULL.* 34, 34–35 (1998); Spencer et al., *supra* note 19, at 6.

or basis of self-esteem [] to avoid the evaluative threat they might feel in that domain.”³³ In a study exploring stereotypes and disassociation from certain fields, Davies et al. explored the relationship between gender-stereotypic television commercials and female participants’ performance on subsequent math tests and interest in more quantitative domains.³⁴ Participants were shown commercials that either portrayed women in a gender-stereotypic role (e.g., excited about a new kind of brownie mix) or a neutral role (e.g., a woman speaking about automotive engineering).³⁵ Davies et al. found that female participants who viewed the gender-stereotypic commercials were (1) more likely to underperform on a subsequent math test after viewing the commercials, (2) more likely to avoid math problems in favor of verbal problems, and (3) likely to express less interest in fields that had a more quantitative character.³⁶ These and other studies³⁷ illustrate just a few of the effects of stereotype threat on student performance, both in terms of which subjects students pursue as well as how well they perform in certain subject areas.³⁸

III. LEGAL LANDSCAPE – PRIOR CHALLENGES TO STANDARDIZED TESTING

The impact of standardized tests on women and non-Asian minorities has long been the subject of debate.³⁹ Is the fact that women may score lower on standardized math tests indicative of innate inferiority in quantitative skills,⁴⁰ or is it the result of more sociological and cultural factors? Does the fact that, on average, African-Americans score lower than white students on standardized

33. Spencer et al., *supra* note 19, at 6.

34. Paul G. Davies et al., *Consuming Images: How Television Commercials that Elicit Stereotype Threat Can Restrain Women Academically and Professionally*, 28 PERSONALITY & SOC. PSYCHOL. BULL. 1615, 1615 (2002).

35. *Id.* at 1619–20.

36. *Id.* at 1626.

37. *Id.*; Lisa A. Harrison et al., *The Consequences of Stereotype Threat on the Academic Performance of White and Non-White Lower Income College Students*, 9 SOC. PSYCHOL. EDUC. 341, 353–54 (2006); Spencer et al., *supra* note 19, at 21–22.

38. Spencer et al., *supra* note 19, at 21–22; Harrison et al., *supra* note 37, at 354.

39. See Shih et al., *supra* note 20, at 82.

40. Spencer et al., *supra* note 19, at 21–22. In 2005, Harvard President Larry Summers created an uproar when he suggested that innate differences in men and women were the reason for the “under-representation of female scientists at elite universities.” Daniel J. Hemel, *Summers’ Comments on Women and Science Draw Ire*, HARV. CRIMSON (Jan. 14, 2005), <http://www.thecrimson.com/article/2005/1/14/summers-comments-on-women-and-science/>.

tests in the U.S.⁴¹ establish a lack of “work ethic” or is something else at play? The debate over standardized tests has touched the media,⁴² the policy world,⁴³ and the courts.⁴⁴ Advocates have challenged standardized testing procedures using various legal claims, including the Equal Protection Clause⁴⁵ and Title VI and Title IX⁴⁶ of the Civil Rights Act of 1964.⁴⁷ Each of these is discussed further below.

A. *Equal Protection Challenges to Standardized Testing*

The Fourteenth Amendment of the Constitution establishes that no state may “deny to any person within its jurisdiction the equal protection of the laws.”⁴⁸ The Equal Protection Clause, as this section of the Fourteenth Amendment has come to be known,⁴⁹ is meant to help protect citizens from arbitrary and discriminatory state action.⁵⁰ Given the disproportionate and discriminatory effect that certain practices—such as asking for race and gender information before a test—can have on female and minority test takers,⁵¹ the Fourteenth Amendment would seem like a probable vehicle for

-
41. See, e.g., Trip Gabriel, *Proficiency of Black Students Is Found to Be Far Lower than Expected*, N.Y. TIMES (Nov. 9, 2010), http://www.nytimes.com/2010/11/09/education/09gap.html?_r=0 (“Only 12 percent of black fourth-grade boys are proficient in reading, compared with 38 percent of white boys, and only 12 percent of black eighth-grade boys are proficient in math, compared with 44 percent of white boys.”).
 42. See Beverly Ge, *Beware the Stereotype Threat*, GAINESVILLE SUN (Dec. 17, 2015, 12:01 AM), <https://www.gainesville.com/opinion/20151217/beverly-ge-beware-the-stereotype-threat>; see also Amy L. Wax, *The Threat in the Air*, WALL ST. J. (Apr. 13, 2004, 12:01 AM), <https://www.wsj.com/articles/SB108181651872980934>.
 43. See Russell A. McClain, *Helping Our Students Reach Their Full Potential: The Insidious Consequences of Ignoring Stereotype Threat*, 17 RUTGERS RACE & L. REV. 1, 10, 51, 56 (2016).
 44. See *infra* Section III.B.i (discussing *GI Forum v. Tex. Educ. Agency*, 87 F. Supp. 2d 667, 668 (W.D. Tex. 2000) and *Larry P. v. Riles*, 793 F.2d 969, 983 (9th Cir. 1984)); see also *Fisher v. Tucson Unified Sch. Dist.*, 329 F. Supp. 3d 883, 915, 978–79 (D. Ariz. 2018).
 45. See *infra* Section III.A.
 46. See discussion *infra* Sections III.B.i–ii.
 47. Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241, 252–53, 266–67 (1965).
 48. U.S. CONST. amend. XIV, § 1.
 49. See Kendra Johnson, *Racially Bias SAT I/ACT Blocks College Access: Is It Constitutional for College Officials to Condition Admission on a Racially Bias Assessment?*, 33 U. BALT. L.F. 2, 5 (2003).
 50. See *id.*
 51. See *supra* notes 18–23 and accompanying text.

claims against standardized test providers.⁵² However, plaintiffs considering claims under the Fourteenth Amendment Equal Protection Clause are limited in two key ways. First, equal protection claims can only be brought against state actors, which may insulate private institutions from such claims.⁵³ Second, because discriminatory impact alone does not make a law unconstitutional under the Equal Protection Clause, plaintiffs must show that the law was motivated by a discriminatory purpose.⁵⁴ That said, because most teachers, school administrators, and policy makers are unlikely to describe the reason for a policy or practice in terms of racial⁵⁵ or gender animosity,⁵⁶ potential claims under the Equal Protection Clause may be a challenge.⁵⁷

In addition, some circuits require plaintiffs to show a “causal link between the effects of past discrimination and the disproportionate impact the test had upon minority students.”⁵⁸ These barriers limit the effectiveness of the Equal Protection Clause as a means to challenge standardized testing practices that may trigger stereotype threat and which have a disproportionate negative impact on women and minorities.

52. See *Washington v. Davis*, 426 U.S. 229, 232–33 (1976) (showing that plaintiffs argued that a certain written test had the effect of disproportionately excluding African Americans from being promoted within a police department). *Contra id.* at 239 (stating that discriminatory purpose, in addition to discriminatory effect, must be shown to succeed on a discrimination claim brought under the Equal Protection Clause of the Fourteenth Amendment).

53. See Philip J. Faccenda & Kathleen Ross, *Constitutional and Statutory Regulation of Private Colleges and Universities*, 9 VALPARAISO U. L. REV. 539, 544 (1976).

Generally, the actions of a private institution are held to be state actions when the private institution is an integral part of the public purpose or when the state has such an active role in the private institution that the state is deemed to have “so insinuated itself into a position of interdependence” with the private institution that it must be recognized as a joint participant in the challenged activity.

Id.

54. Johnson, *supra* note 49, at 5.

55. See Adam Serwer, *Trump Is Making It Easier to Get Away with Discrimination*, ATLANTIC (Jan. 4, 2019), <https://www.theatlantic.com/ideas/archive/2019/01/disparate-impact/579466/>.

56. *See id.*

57. Betsy A. Gerber, *High Stakes Testing: A Potentially Discriminatory Practice with Diminishing Legal Relief for Students at Risk*, 75 TEMP. L. REV. 863, 879–80 (2002).

58. *Id.* at 875–76.

B. Title VI and Title IX

In addition to the Equal Protection Clause, Title VI and Title IX of the Civil Rights Act of 1964 also provide ways for plaintiffs to challenge standardized testing procedures.⁵⁹ Title VI protects against discrimination based on “race, color, or national origin” in “any program or activity receiving Federal financial assistance.”⁶⁰ Title IX, on the other hand, protects against sex discrimination and states that “[n]o person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance.”⁶¹ In short, for an actor or jurisdiction to be subject to Title VI and Title IX, generally that actor must receive some form of federal financial assistance or otherwise be considered engaged in state action.⁶² Together, these two provisions provide the backbone of much of the important civil rights advocacy for educational equality that has occurred in the past five decades.

i. Title VI

Although under Title VI private plaintiffs can only bring suit in federal courts for intentional discrimination,⁶³ the regulations promulgated pursuant to Title VI “proscribe actions having disparate impact on groups protected by the statute, even if those actions are not intentionally discriminatory.”⁶⁴ This provides an avenue for

59. U.S. DEP’T OF JUSTICE, TITLE VI LEGAL MANUAL, SECTION VI: PROVING DISCRIMINATION – INTENTIONAL DISCRIMINATION (2017). Although discussed separately, the elements of a “Title VI intent claim derive from and are similar to the analysis of cases decided under the Fourteenth Amendment’s Equal Protection Clause.” *Id.*

60. 42 U.S.C. § 2000a(d) (2012).

61. 20 U.S.C. § 1681(a) (2012).

62. For a discussion of regulation of private schools and institutions, see *Facenda & Ross*, *supra* note 53, at 540–48; see also 42 U.S.C. § 2000a (2012); 20 U.S.C. § 1681 (2012).

63. See *Alexander v. Sandoval*, 532 U.S. 275, 280 (2001).

64. Gerber, *supra* note 57, at 876; see *GI Forum Image de Tejas v. Tex. Educ. Agency*, 87 F. Supp. 2d 667, 669 (W.D. Tex. 2000).

[T]his Court has allowed the Plaintiffs to bring a claim pursuant to a regulation adopted in conjunction with Title VI. See 34 C.F.R. § 100.3. That regulation, in clear, unmistakable terms, prohibits a federally funded program from implementing policies that have a disparate impact on minorities. *Id.* While the Court acknowledges that the United States Supreme Court has limited

advocates to use Title VI to challenge practices that are not explicitly based on racial animosity.⁶⁵ However, the utility of Title VI's regulations is still limited, as there is no private right of action for individuals wishing to enforce the regulations.⁶⁶ In other words, although private individuals can sue for *intentional discrimination* under Title VI, they cannot challenge a practice solely because of its disparate impact on a protected class.⁶⁷ Instead, individuals wishing to challenge a practice having a disparate impact must turn to the appropriate enforcement agency.⁶⁸ For cases challenging standardized testing practices that agency is usually the U.S. Department of Education.⁶⁹

Courts often choose to use the three part framework established in *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973) when assessing Title VI cases.⁷⁰ The *McDonnell-Douglas* framework applied in the Title VI context is fairly simple:

[A] plaintiff must first prove that the application of a facially neutral practice caused a disproportionate adverse effect on a particular racial group. If the plaintiff makes

Title VI itself to constitutional parameters (i.e., has required a showing of an intent to discriminate in order to prove a violation), see *United States v. Fordice*, 505 U.S. 717, 722 n.7, 112 S. Ct. 2727, 120 L. Ed. 2d 575 (1992), the Court does not find that this limitation has been clearly and unambiguously extended to its implementing regulations. The Court is not alone in reaching this conclusion.

GI Forum, 87 F. Supp. 2d at 669 (citations omitted).

65. Cf. Gerber, *supra* note 57, at 876 (explaining that plaintiffs suing for discrimination under Title VI must show that a *facially neutral* practice has had a disproportionate adverse effect on a racial group).
66. *Sandoval*, 532 U.S. at 293 (holding that there is no private right of action to enforce Title VI regulations).
67. *Id.* at 285.
68. See, e.g., *id.* at 291–93 (holding that Title VI does not create a freestanding private right of action to enforce disparate impact regulations promulgated under § 602).
69. See Regulations of the Offices of the Department of Education, 34 C.F.R. § 100.1 (2018) (explaining that the purpose of the U.S. Department of Education regulations is to enforce Title VI of the Civil Rights Act of 1364); 34 C.F.R. § 100.7 (2018) (explaining that the Department of Education must investigate and resolve all complaints). Cf. *Sandoval*, 532 U.S. at 289 (discussing how “[s]ection 602 . . . focuses neither on the individuals protected nor even on the funding recipients being regulated, but on the agencies that will do the regulating”).
70. There are a number of analytical frameworks used by courts in assessing intent claims, including examining “express classifications” or circumstantial evidence of discrimination under the *Arlington Heights* analysis. See U.S. DEP’T OF JUSTICE, *supra* note 59.

such a showing, the defendant then bears the burden of showing that the procedure in question is justified by educational necessity. Educational necessity exists when the challenged practice serves a legitimate educational goal of the school or district. Even if the defendant makes a showing of educational necessity, however, the plaintiff may still succeed if he shows that an equally effective alternative is available to meet the educational goal and would result in less racial disproportionality.⁷¹

There have been a handful of cases that have challenged testing practices under the Title VI framework.⁷² For instance, in *Georgia State Conference of Branches of NAACP v. State of Georgia*, plaintiffs claimed that black students were assigned to special education programs and regular classes in a discriminatory way.⁷³ While this case was decided before the U.S. Supreme Court limited Title VI claims to those involving intentional discrimination,⁷⁴ it provides a useful example of how a claim might be assessed. In *Georgia State Conference of Branches of NAACP*, the Eleventh Circuit agreed with the district court's finding that the "plaintiffs had met their burden of establishing a *prima facie* case through statistics showing that the racial composition of many of the local defendants' regular classrooms differs from what would be expected from a random distribution."⁷⁵ However, because the defendants showed that the practice was an educational necessity and plaintiffs did not establish "the existence of equally sound educational alternatives,"⁷⁶ the Eleventh Circuit found that the plaintiffs failed at the second step of the Title VI framework and upheld defendants' practice.⁷⁷

Similarly, in *GI Forum v. Texas Educational Agency*, the court considered "whether the use of the Texas Assessment of Academic Skills (TAAS) examination as a requirement for high school graduation unfairly discriminate[d] against Texas minority

71. Gerber, *supra* note 57, at 876.

72. See, e.g., *Ga. State Conference of Branches of NAACP v. Georgia*, 775 F.2d 1403, 1407 (11th Cir. 1985); *GI Forum Image De Tejas v. Tex. Educ. Agency*, 87 F. Supp. 2d 667, 668 (W.D. Tex. 2000).

73. *Ga. State Conference of Branches of NAACP*, 775 F.2d at 1407.

74. See *Sandoval*, 532 U.S. at 280.

75. *Ga. State Conference of Branches of NAACP*, 775 F.2d at 1417.

76. *Id.*

77. *Id.*

students.”⁷⁸ The plaintiffs successfully proved that the challenged testing practice had an adverse impact on Hispanic students.⁷⁹ The court found that plaintiffs established “an inescapable conclusion that . . . Hispanic and African[-]American students have performed significantly worse on all three sections of the exit exam than majority students.”⁸⁰ However, the court also noted that “it is highly significant that minority students have continued to narrow the passing rate gap at a rapid rate,” that “minority students have made gains on other measures of academic progress,” and that “[t]he number of minority students taking college entrance examinations has also increased.”⁸¹ Because the defendant educational agency was able to show that the test was an educational necessity and that there were no less discriminatory alternatives, plaintiffs, like the plaintiffs in *Georgia State Conference of Branches of NAACP*, failed at the second step and the court upheld the testing practice.⁸²

While plaintiffs in *Georgia State Conference of Branches of NAACP* and *GI Forum* successfully proved a *prima facie* case of discrimination but failed at later stages of the Title VI framework, other plaintiffs prior to the *Sandoval* decision limiting Title VI claims, had more success bringing these types of challenges.⁸³ For instance, in *Larry P. v. Riles*, plaintiffs, six black school children, brought a class action to challenge the use of “IQ tests used by the California school system to place children into special classes for the educable mentally retarded.”⁸⁴ The plaintiffs showed the challenged practice had an adverse impact on minority students by pointing to data establishing “that black children as a whole scored ten points lower than white children on the tests, and that the percentage of black children in [special education] classes was much higher than for whites” and that “these test scores were used to place black

78. *GI Forum*, 87 F. Supp. 2d at 668. Like *Georgia State Conference of Branches of NAACP*, this case was also decided prior to the Supreme Court’s decision that individual plaintiffs could only bring intentional discrimination claims under Title VI. See *Sandoval*, 532 U.S. at 280. It is also important to note that, because this case was decided before the U.S. Supreme Court limited Title VI claims to just those involving intentional discrimination, it provides a useful example of how a claim brought by the U.S. Department of Education—but not individual plaintiffs—might be assessed using a disparate impact analysis. See *GI Forum*, 87 F. Supp. 2d at 668.

79. *GI Forum*, 87 F. Supp. 2d at 675.

80. *Id.*

81. *Id.*

82. *Id.*; *Ga. State Conference of Branches of NAACP*, 775 F.2d at 1420–21.

83. See *supra* notes 72–82 and accompanying text; *Larry P. v. Riles*, 793 F.2d 969, 983 (9th Cir. 1984).

84. *Riles*, 793 F.2d at 972; Gerber, *supra* note 57, at 877–78.

schoolchildren in [special education] classes and to remove them from the regular educational program.”⁸⁵ The burden then shifted to defendants who failed to demonstrate that the tests “were required by educational necessity.”⁸⁶ Finally, the court found that alternatives to the challenged testing procedure—such as “rely[ing] more on observational data”—were less discriminatory than under the IQ-centered standard.⁸⁷

ii. Title IX

Title IX prohibits “any education program or activity receiving Federal financial assistance” from denying their benefits to, or discriminating against, any person on the “basis of sex.”⁸⁸ One of Congress’s main goals in enacting Title IX was “to provide individual citizens effective protection against . . . [discriminatory] practices.”⁸⁹ Again, as with Title VI, generally only entities integrally connected with the state or receiving federal funds can be sued under Title IX.⁹⁰

Courts have consistently interpreted Title IX regulations to prohibit both intentional discrimination and practices which have a disparate impact based on sex.⁹¹ In considering a discrimination claim under Title IX, courts often use a framework similar to that used in employment cases or Title VI cases: a plaintiff must first show that the challenged practice disproportionately affected a sex; the burden then shifts to the defendant to show that the practice is a business necessity; and finally, a plaintiff must show that there are less discriminatory alternatives to the challenged practice.⁹²

In *Sharif v. New York State Education Department* plaintiffs challenged the use of standardized tests which had a disparate impact

85. *Riles*, 793 F.2d at 982–83.

86. *Id.* at 983.

87. *Id.* at 978.

88. 20 U.S.C. § 1681(a) (2012).

89. *Gebser v. Lago Vista Indep. Sch. Dist.*, 524 U.S. 274, 286 (1998) (quoting *Cannon v. Univ. of Chi.*, 441 U.S. 677, 703 (1979)).

90. *See Faccenda & Ross*, *supra* note 53, at 544; *see also Cannon*, 441 U.S. at 729.

91. Katherine Connor & Ellen J. Vargyas, *The Legal Implications of Gender Bias in Standardized Testing*, 7 BERKELEY WOMEN’S L.J. 13, 42–43 (1992); *see also Sharif v. N.Y. State Educ. Dep’t*, 709 F. Supp. 345, 361 (S.D.N.Y. 1989) (interpreting Title IX regulations and previous decisions, the Court found Title IX regulations “prohibit testing practices with a discriminatory effect on one sex,” and that plaintiffs were therefore not required to prove intentional discrimination).

92. Connor & Vargyas, *supra* note 91, at 48–49.

on female students.⁹³ The plaintiffs—a group of ten high school students and two organizations—sued the New York State Education Department and the Commissioner of Education for their practice of strictly relying on SAT scores to determine recipients of New York State merit scholarships.⁹⁴ In ruling for the plaintiffs, the U.S. District Court for the Southern District of New York found that:

As a result of the State's practice of basing scholarship awards solely upon SAT scores, males have consistently received substantially more scholarships than females. In 1987 for example, males were 47 percent of the scholarship competitors, but received 72 percent of the Empire State Scholarships and 57 percent of the Regents Scholarships.⁹⁵

This “persuasive statistical evidence[,]” taken together with “credible expert testimony” was enough for the court to find that plaintiffs “met their burden of establishing a *prima facie* case.”⁹⁶

Turning to whether defendants' reliance on the SAT was rooted in educational necessity, the court found that defendants failed to establish a “relationship between use of the SAT and recognition and award of academic achievement in high school.”⁹⁷ The court noted that “there can be no serious claim that a test given on one single morning can take into account a student's diligence, creativity and social development, and work habits in that student's environment—all part of high school achievement.”⁹⁸

Finally, the court considered whether plaintiffs adequately provided a less discriminatory alternative.⁹⁹ The court found that plaintiffs' alternative—a combination system in which reviewers used both GPA and SAT scores—was “not a perfect alternative” but that it was “the best alternative presently available.”¹⁰⁰

The court's considerations in *Sharif* provide a useful guidebook for investigatory agencies and advocates examining testing practices which may trigger stereotype threat for female students.¹⁰¹

93. *Sharif*, 709 F. Supp. at 348.

94. *Id.*

95. *Id.* at 355.

96. *Id.* at 362.

97. *Id.*

98. *Id.*

99. *Id.* at 362–63.

100. *Id.* at 363.

101. *See supra* text accompanying notes 94–100.

IV. CURRENT STANDARDIZED TESTING PRACTICES IN A TITLE VI AND TITLE IX FRAMEWORK

To determine whether current testing practices which trigger or exacerbate stereotype threat might violate Title VI or Title IX, it is necessary to consider each of the prongs of the burden shifting framework and the relevant research.¹⁰²

A. *Establishing a Prima Facie Case of Discrimination*

First, do testing practices—such as collecting demographic information before testing or having a white or male test proctor—trigger stereotype threat and impact performance to such an extent that a *prima facie* case can be established? The impact of differing test practices on minority and female students, differs from situation to situation, of course, but research has identified a few trends that raise serious Title VI and Title IX concerns.¹⁰³ For instance, numerous studies have established that asking test takers to identify their race or gender, or otherwise priming test takers with their racial or gender identity, immediately before taking a test, can cause minority and female students to underperform to an extent that is statistically significant.¹⁰⁴ In 2008, for example, Danahar and Crandall examined the results of an earlier study concerning the impact of stereotype threat on girls taking the AP Calculus examination.¹⁰⁵ Danahar and Crandall concluded that simply moving demographic questions to the end of an AP Calculus examination produced a significant increase in female performance on this test.¹⁰⁶ They estimated that “[t]his simple, small, and inexpensive change could increase U.S. women receiving AP Calculus AB credit by more than 4,700 every year.”¹⁰⁷

102. See *infra* Sections IV.A–C.

103. See Gerber, *supra* note 57, at 877–78 (explaining how the Ninth Circuit held that a school district violated Title VI in its testing for special education classes); Andrea Silverstein, *Standardized Tests: The Continuation of Gender Bias in Higher Education*, 29 HOFSTRA L. REV. 669, 683, 686–89 (2000) (explaining that courts have recognized a Title IX concern with certain discriminatory testing practices).

104. See, e.g., Steele & Aronson, *supra* note 2, at 806–07 (finding that black students did significantly worse on the SAT when they were asked to list their race right before they took the test).

105. Kelly Danahar & Christian S. Crandall, *Stereotype Threat in Applied Settings Re-Examined*, 38 J. APPLIED SOC. PSYCHOL. 1639, 1639 (2008).

106. *Id.*

107. *Id.*

Similarly, as discussed above, a study by Aronson and Steele found that soliciting information regarding students' racial identification immediately before issuing a test containing tough questions from the verbal section of the GRE (Graduate Record Examination) suppressed the scores of African-American test takers but not white test takers.¹⁰⁸ When asked to identify their race before the test, African-American test takers performed "significantly worse" than African-Americans test takers who were not asked about their race immediately before the test.¹⁰⁹

Research also indicates that other common testing practices have a discriminatory impact that could be challenged under a Title VI or Title IX framework.¹¹⁰ Marx and Goff found that black students performed better on a difficult verbal test when the test was administered by a black proctor, rather than a white proctor, while white students were unaffected by the race of the test administrator.¹¹¹

In other studies, researchers have examined how the way a task is described potentially triggers stereotype threat and causes underperformance for minority and female students.¹¹² Huguet and Régner found that girls aged eleven to thirteen performed better on a task when it was described as a "memory game" or as testing their drawing skills, than when it was described as a "geometry test" or as testing math skills.¹¹³ Brown and Day found that African-Americans under-performed on a test when it was described as "an IQ test" measuring "individuals' intelligence and ability," but performed as well as white students when it was described as just a "series of puzzles."¹¹⁴ As the researchers conclude, their results "suggest that just the implication that a test is intellectually evaluative is enough to diminish performance among African[-]American respondents."¹¹⁵ The difference in performance in these studies have been shown to be

108. Steele & Aronson, *supra* note 2, at 802.

109. *Id.*

110. See *infra* notes 111–16 and accompanying text.

111. David M. Marx & Philip Atiba Goff, *Clearing the Air: The Effect of Experimenter Race on Target's Test Performance and Subjective Experience*, 44 BRITISH J. SOC. PSYCHOL. 645, 651 (2005).

112. See Ryan P. Brown & Eric Anthony Day, *The Difference Isn't Black and White: Stereotype Threat and the Race Gap on Raven's Advanced Progressive Matrices*, 91 J. APPLIED PSYCHOL. 979, 979 (2006); see also Pascal Huguet & Isabelle Régner, *Stereotype Threat Among Schoolgirls in Quasi-Ordinary Classroom Circumstances*, 99 J. EDUC. PSYCHOL. 545, 545 (2007).

113. Huguet & Régner, *supra* note 112, at 549–50.

114. Brown & Day, *supra* note 112, at 981–82.

115. *Id.* at 984.

statistically significant—for instance, in Brown and Day’s study they noted that “the differences within African-American participants were not only statistically significant, they were also substantial—approximately three fourths of a standard deviation.”¹¹⁶

The fact that the results in these experiments are recognized as statistically significant within the relevant scientific community is an important factor in considering whether potential plaintiffs in a Title VI or Title IX case would be able to establish a *prima facie* case challenging a testing practice as having a discriminatory impact.¹¹⁷ As discussed above, a simple showing “through statistics” that “the racial composition” or impact of a practice “differs from what would be expected from random distribution” was sufficient to establish a *prima facie* case in *Georgia State Conference of Branches of NAACP*.¹¹⁸ Similarly, in *Larry P.*, the court found that “plaintiffs met their burden of establishing a *prima facie* case of discrimination as it was “undisputed that black children as a whole scored” lower than white children on the test being challenged and that “the percentage of black children” in classes for the “mentally retarded” “was much higher than for whites.”¹¹⁹ Still other courts have looked at whether a challenged educational practice passes the “Four-Fifths” rule used in employment cases¹²⁰ or the *Shoben* rule.¹²¹ Although there is “no rigid mathematical threshold of disproportionality,”¹²² courts considering challenged testing practices that may trigger stereotype threat should consider whether statistical evidence shows the results of the challenged practice “differ[] from what would be expected from a random distribution”¹²³ or otherwise satisfies one of

116. *Id.* at 982.

117. *See supra* notes 72–101 and accompanying text.

118. *Ga. State Conference of Branches of NAACP v. Georgia*, 775 F.2d 1403, 1417 (11th Cir. 1985). In considering the next step in the burden shifting framework, however, the court went on to find that defendants “successfully rebutted the plaintiffs’ *prima facie* case by establishing the educational necessity for grouping students” *Id.*

119. *Larry P. v. Riles*, 793 F.2d 969, 982–83 (9th Cir. 1984).

120. *GI Forum Image de Tejas v. Tex. Educ. Agency*, 87 F. Supp. 2d 667, 675 n.7 (W.D. Tex. 2000) (discussing whether the challenged practice has an “adverse impact where the pass rate for the minority group is less than 80 percent of the passing rate for the majority group”).

121. *Id.* at 675 n.8 (finding that the *Shoben* rule “seeks to assess the statistical significance of observed numerical disparities by determining differences between independent proportions”).

122. *Id.* at 679 (citing *Cureton v. Nat’l Collegiate Athletic Assoc.*, 37 F. Supp. 2d 687, 697 (E.D. Pa. 1999)).

123. *Ga. State Conference of Branches of NAACP*, 775 F.2d at 1417.

the tests courts have used to find a *prima facie* case of discrimination.¹²⁴

B. Testing Practices as an Educational Necessity

Second, attorneys, judges, and others considering a testing practice within the Title VI or Title IX framework must ask if the challenged practice is an educational necessity.¹²⁵ As mentioned above, the inquiry into whether a challenged practice is an educational necessity mirrors the inquiry under Title VII into whether a challenged practice in the employment setting is a business necessity.¹²⁶ In order to establish that a testing practice is an educational necessity, “the defendant must show a rational relationship between the practice . . . and the purpose of using the [practice].”¹²⁷

As the court noted in *GI Forum*, “[t]he word ‘necessity’ . . . is somewhat misleading; the law does not place so stringent a burden on the defendant as that word’s common usage might suggest. Instead, an educational necessity exists where the challenged practice serves the *legitimate* educational goals of the institution.”¹²⁸ In that case, the court specifically looked at whether the defendant could show that there was a “manifest relationship” between the challenged test and “a legitimate educational goal.”¹²⁹ The court considered the “alleged deficiencies” of the challenged test and weighed them against the “articulated goals” of the test, including holding “schools, students, and teachers accountable for education and to ensure that all Texas students receive the same, adequate learning opportunities.”¹³⁰

124. See *supra* notes 117–21 and accompanying text.

125. See *supra* notes 73–101 and accompanying text.

126. See *supra* notes 73–101 and accompanying text.

127. Andrea Silverstein, *Standardized Tests: The Continuation of Gender Bias in Higher Education*, 29 HOFSTRA L. REV. 669, 693 (2000). Although this framework mirrors that used by courts considering employment practices, at least one court has acknowledged, the concept of “necessity” in the education setting is fundamentally different as compared with the business setting. See Connor & Vargyas, *supra* note 91, at 46 (quoting *Larry P. v. Riles*, 495 F. Supp. 926, 969 (N.D. Cal. 1979)).

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

Riles, 495 F. Supp at 969.

128. *GI Forum Image de Tejas v. Tex. Educ. Agency*, 87 F. Supp. 2d 667, 679 (W.D. Tex. 2000).

129. *Id.*

130. *Id.*

The court found that the “test accomplishes what it sets out to accomplish, which is to provide an objective assessment of whether students have mastered a discrete set of skills and knowledge.”¹³¹

Similarly, in *Georgia State Conference of Branches of NAACP*, the court considered whether the challenged practice—the assignment of black children to special education programs—had a “manifest demonstrable relationship” to classroom education.¹³² The Eleventh Circuit found that there was no “direct evidence” showing that students were assigned to classrooms on anything but criteria “related to the subject matter taught in the specific class.”¹³³ Furthermore, because there was evidence showing the “validity of certain testing procedures” and an improvement in student scores, the court found defendants successfully established the necessity of the challenged practice.¹³⁴

In challenging a testing practice that may trigger stereotype threat, such as collecting demographic data before a test or describing the purpose of a test in language that triggers stereotypes, plaintiff attorneys will need to be able to rebut arguments that the challenged practice is an educational necessity.¹³⁵ Arguments that a practice is an educational necessity are likely to range from practice to practice, and will depend on how defendants describe the goal of the practice.¹³⁶ For instance, exams including passages for reading comprehension might unintentionally trigger stereotype threat if the reading passages are loaded with obvious, or even subtle, words priming stereotypes.¹³⁷ If a test provider is able to show that a specific reading passage is a necessity, and that a different passage could not be substituted in, and if that passage bears a manifest relationship with the goal of the test, it is likely a court will find a plaintiffs’ *prima facie* case of discrimination rebutted.¹³⁸

131. *Id.* at 680.

132. *See* *Ga. State Conference of Branches of NAACP v. Georgia*, 775 F.2d 1403, 1418 (11th Cir. 1985).

133. *Id.* at 1420.

134. *Id.*

135. *See* *GI Forum*, 87 F. Supp. 2d at 677.

136. *See id.* at 679.

137. *Cf.* *Steele & Aronson*, *supra* note 2, at 808 (explaining that even a person’s “mere cognitive availability of the racial stereotype [present in an exam] is enough to depress . . . intellectual performance”).

138. *See, e.g.,* *GI Forum*, 87 F. Supp. 2d at 670–71 (holding that regardless of whether a test is found to cause an adverse impact on minority students, if a manifest relationship with the goal of the test exists and the plaintiff has not proven that an adequate alternative existed, the plaintiff’s case of discrimination has been rebutted).

C. *Alternatives Having a Less Discriminatory Impact*

Third, in considering a challenge to a practice that might trigger stereotype threat on testing, courts will consider whether plaintiffs have “demonstrate[d] that an alternative practice results in smaller racial/ethnic disparities but is nonetheless equally effective in meeting the institution’s educational goals.”¹³⁹ A court will examine the relationship between the stated goal of defendant and the challenged practice and proposed alternative.¹⁴⁰ In *GI Forum*, the court found that, even though there were alternatives to the challenged testing practice that had a less discriminatory impact, none were sufficient as the alternatives failed to “sufficiently motivate students [and teachers] to perform to their highest ability” and “provide an adequate and fair education.”¹⁴¹ In *Larry P. v. Riles*, discussed above, the court found that alternatives had “been in effect” and that those alternatives—taking “more time and care with their assessments” and “rely[ing] more on observational data” in determining which students had educational problems—were “less discriminatory than under the IQ-centered standard.”¹⁴²

In the context of a challenge concerning a practice triggering stereotype threat, arguments that there are less discriminatory alternatives are likely to range from practice to practice.¹⁴³ For instance, some studies have found that African-American students perform better when black proctors administer tests, rather than white proctors.¹⁴⁴ Yet, in a challenge concerning the lack of diversity in test proctors, a court may find that, if the pool of potential test administrators is small due to geographic or other limitations, there may be no alternative to test proctors being mostly white.¹⁴⁵ However, other practices triggering stereotype threat may have ready alternatives available which sufficiently meet the goals of the defendant jurisdiction or party.¹⁴⁶ For example a ready alternative, with little cost, exists to the practice of gathering demographic

139. William C. Kidder & Jay Rosner, *How the SAT Creates “Built-in Headwinds”: An Educational and Legal Analysis of Disparate Impact*, 43 SANTA CLARA L. REV. 131, 200 (2002).

140. *See, e.g., GI Forum*, 87 F. Supp. 2d at 679.

141. *Id.* at 681.

142. *Larry P. v. Riles*, 793 F.2d 969, 976 (9th Cir. 1984).

143. *See infra* notes 151–67 and accompanying text.

144. Marx & Goff, *supra* note 111, at 651.

145. *Cf. GI Forum*, 87 F. Supp. 2d at 684.

146. *See infra* notes 175–76 and accompanying text.

information before a test: collect demographic information *after* a test.¹⁴⁷

V. STRATEGIES FOR REDUCING STEREOTYPE THREAT

Given the significant impact that stereotype threat can have on the performance of women and minorities in high stake testing, it is imperative that stakeholders explore means through which to lessen the role of stereotype threat.¹⁴⁸ This Section first discusses research offering insight into how to reduce the impact of stereotype threat¹⁴⁹ and then discusses policy actions for implementing strategies to reduce stereotype threat.¹⁵⁰

A. *The Science of Reducing Stereotype Threat*

Researchers have identified various strategies for reducing the impact of stereotype threat.¹⁵¹ As previously discussed, studies have found that by asking test takers for their gender or racial identity after a test, rather than before, underperformance for female and non-Asian minority test takers can be reduced.¹⁵² Researchers believe this is because “reminding” a test taker of their racial or gender identity—an identity which may carry negative stereotypes relevant to the task at hand—can distract the test taker from the task at hand by increasing anxiety.¹⁵³ One key strategy for reducing the impact of stereotype threat is to lessen conditions that may prime a subject with social identities that are negatively stereotyped in ways relevant to the task at hand.¹⁵⁴

A second, but related, strategy for reducing the impact of stereotype threat is to change contextual clues in how a task is described.¹⁵⁵ Researchers have found that simply adjusting how a test or task is described can have a serious impact on how well a subject performs—for instance, African-American participants perform better on a logic test when the task is described as a series of puzzles than when the same task is described as a high level IQ

147. See *supra* notes 105–07 and accompanying text.

148. See Marx & Goff, *supra* note 111, at 645.

149. See *infra* Section V.A.

150. See *infra* Section V.B.

151. See *infra* notes 152–71 and accompanying text; see also *infra* Section V.B.2.

152. See *supra* notes 104–07 and accompanying text.

153. Steele & Aronson, *supra* note 2, at 808.

154. See *id.*

155. *Id.* at 109.

test.¹⁵⁶ Thus, terms which may be commonly used in stereotyping a certain group should be avoided when describing activities in which certain social groups traditionally underperform to reduce stereotype threat.¹⁵⁷ For instance, describing an activity as “competitive” may subconsciously deter women from participating.¹⁵⁸ However, given the fluidity of language, a comprehensive strategy eliminating the use of words that may trigger stereotype threat is impossible.¹⁵⁹

A related, more practical strategy, involves assuring test takers that a test is fair.¹⁶⁰ Good, Aronson and Harder found that female students taking a difficult math test performed better if they were told that “this mathematics test has not shown any gender differences in performance or mathematics ability.”¹⁶¹

Still another means of lowering the impact of stereotype threat is to provide a positive representative of an individual from the stereotyped group.¹⁶² This might help “reverse” any negative primes or stereotypes.¹⁶³ Researchers have found providing diverse test proctors can help reduce stereotype threat for female and minority students.¹⁶⁴ Marx and Goff found that black test takers performed as well as white test takers when the test administrator was a black individual than when the test administrator was white.¹⁶⁵ Similar results have been found for women taking a difficult math test: when the test was administered by a woman, women and men performed at the same level.¹⁶⁶ However, when the test was administered by a man, women underperformed.¹⁶⁷

Finally, because stereotype threat triggers anxiety, which may then impede mental focus, researchers have found that providing test takers with an alternate explanation for their sense of anxiety can

156. Brown & Day, *supra* note 112, at 981, 984.

157. Steele & Aronson, *supra* note 2, at 109.

158. Muriel Niederle & Lise Vesterlund, *Explaining the Gender Gap in Math Test Scores: The Role of Competition*, 24 J. ECON. PERSP. 129, 133–35 (2010). Author interviews with law school staff also show that, anecdotally at least, describing classes and other law school activities—such as classes concerning oral argument skills—as “competitive” may suppress interest and participation by women. *Cf. id.*

159. *Cf.* Brown & Day, *supra* note 112, at 984.

160. Catherine Good et al., *Problems in the Pipeline: Stereotype Threat and Women’s Achievement in High-Level Math Courses*, 29 J. APPLIED PSYCHOL. 17, 26 (2008).

161. *Id.* at 22.

162. Marx & Goff, *supra* note 111, at 647.

163. *See supra* Part IV.

164. Marx & Goff, *supra* note 111, at 647.

165. *Id.* at 651.

166. David M. Marx & Jasmin S. Roman, *Female Role Models: Protecting Women’s Math Test Performance*, 28 PERSONALITY & SOC. PSYCHOL. BULL. 1183, 1186 (2002).

167. *Id.*

help reduce the impact of stereotype threat.¹⁶⁸ The performance gap for seventh-grade boys and girls disappeared when they were told that “many students tend to experience difficulty when they move to a new educational situation (such as junior high school) but then bounce back after they become accustomed to their new environment.”¹⁶⁹ This result is likely because girls had an external source to which to attribute their feelings of anxiety.¹⁷⁰ In other words, a “nonpejorative attribution[] for difficulties may render standardized testing more equitable for students who must contend with stereotypes impugning their intellectual abilities.”¹⁷¹

These and other experiments¹⁷² demonstrate that researchers have identified concrete means for reducing the impact of stereotype threat in the learning environment.¹⁷³ To best transfer the strategies identified by psychologists to the classroom and testing environment, policy reform must occur.¹⁷⁴

B. Policy Reform

Systemic reform, aimed at reducing practices that may exacerbate stereotype threat, is an important step in ensuring that everyone, including female and minority students, have an equal shot in the classroom and when taking tests. Policies that may (unintentionally) prime students with their race, gender, or other social identity that might carry negative stereotypes related to a specific task should be identified and assessed. Low-cost methods of minimizing the impact of stereotype threat should be explored and implemented when possible.

i. Federal and State Guidelines

Guidelines issued by the U.S. Department of Education and state departments of education regarding stereotype threat and standardized testing would play an important role in educating stakeholders about the psychological phenomenon. These guidelines could also offer best practices for reducing the impact of stereotype

168. Catherine Good et al., *Improving Adolescents' Standardized Test Performance: An Intervention to Reduce the Effects of Stereotype Threat*, 24 APPLIED DEVELOPMENTAL PSYCHOL. 645, 658–59 (2003).

169. *Id.* at 654.

170. *See id.* at 659.

171. *Id.*

172. *See id.* at 647.

173. *See id.*

174. *See infra* notes 175–79 and accompanying text.

threat in the educational environment. For instance, such guidelines might recommend that test providers—such as the College Board and Law School Admissions Council—collect demographic data on test takers *after* the test, rather than before.¹⁷⁵ Similarly, guidelines might encourage school administrators and test providers to hire diverse exam proctors for high-stake tests.¹⁷⁶

ii. Advocacy with Educational Stakeholders

In addition to pushing for official guidelines issued by the U.S. Department of Education, advocates wishing to reduce the impact of implicit bias on standardized tests can also work with private testing companies and state associations to reform testing practices. Many higher education programs are operated by private organizations, such as the Law School Admissions Corporation, Educational Testing Services,¹⁷⁷ or the College Board.¹⁷⁸ State-level organizations also create and administer standardized tests; for example, state bar associations are often responsible for operating a state bar exam which lawyers must pass in order to practice in a state, placing these associations in a good position to leverage widespread reform.¹⁷⁹

A comprehensive survey of standardized tests, ranging from exams students must take at the K-12 level, to tests required for admission to higher education programs, to exams related to licensing and other professional advancement, is necessary. Gathering information regarding the specific testing practices of each exam, as well as test performance broken down by race and gender, will allow advocates

175. See Danaher & Crandall, *supra* note 105, at 1640–41. Collecting such information serves an important role in furthering various civil rights goals and measuring whether a practice may have other unintended consequences. See *id.*

176. Cf. *supra* notes 104–07 and accompanying text (evidencing that waiting until a test taker has finished their exam would still allow for demographic information to be collected without interfering with any student's test score—regardless of their gender or racial identity).

177. Educational Testing Service operates the GRE (Graduate Record Examinations or Revised General Test), TOEFL (Test of English as a Foreign Language), and other standardized tests. See *Resources for Higher Education*, EDUC. TESTING SERV., <https://www.ets.org/highered> (last visited Feb. 9, 2019).

178. The College Board operates the SAT (Scholastic Aptitude Test), PSAT (Preliminary Scholastic Aptitude Test), Advanced Placement exams, and other standardized tests. See *About the College Board*, THE COLL. BD., <https://about.collegeboard.org/overview> (last visited Feb. 9, 2019).

179. See The Times Editorial Board, *Ease Up on California's Bar Exam to Achieve More Diversity Among Lawyers*, L.A. TIMES (Oct. 6, 2017, 5:00 AM), <https://www.latimes.com/opinion/editorials/la-ed-california-bar-exam-20171006-story.html>.

to better target their outreach to those testing programs which include practices that may make it more likely that stereotype threat is triggered.

V. CONCLUSION

Stereotypes permeate society at all levels, influencing children as young as four.¹⁸⁰ Students who are members of social groups which are subject to negative stereotypes concerning a certain subject or area of expertise may have to navigate a mental “tax” when taking a standardized test in that subject.¹⁸¹ This mental “tax”—otherwise known as stereotype threat—raises questions regarding whether common practices used in standardized testing might create an unequal playing field.¹⁸²

Lawyers and other advocates invested in ensuring students have equal opportunities must continue to educate themselves about the unconscious impact stereotypes can have on students.¹⁸³ A range of civil rights laws provide the means to challenge practices which may trigger a psychological response and disadvantage women and minority students on standardized tests.¹⁸⁴

Attorneys and investigatory agencies can rely on previous case law challenging standardized testing practices to pressure stakeholders to change practices which trigger stereotype threat and may cause underperformance by women and minority students.¹⁸⁵ Policy advocates can also rely on social science research and civil rights law to put pressure on education departments and testing companies to adopt practices and guidelines which reduce the impact of stereotype threat in their testing procedures.¹⁸⁶

180. Cf. Perri Klass, *Breaking Gender Stereotypes in the Toy Box*, N.Y. TIMES (Feb. 5, 2018), <https://www.nytimes.com/2018/02/05/well/family/gender-stereotypes-children-toys.html>.

181. See *supra* notes 19–20, 26–31 and accompanying text.

182. See *supra* note 51 and accompanying text.

183. See *supra* notes 117–24 and accompanying text.

184. See *supra* Section III.B.

185. See *supra* notes 72–87, 93–101 and accompanying text.

186. See *supra* Part III.

