Stereotype Threat or a Lack of Control(s)? Evaluating Robustness Against Changes in Specifications

Share Your Story

Stereotype Threat or a Lack of Control(s)? Evaluating Robustness Against Changes in Specifications

Remington Hill

4/30/2018

1 Abstract

Affirmative action in higher education has been controversial since 1961. *The Source of the River* (Douglas Massey and Fischer (2003)) uses the National Longitudinal Survey of Freshmen—a demographic and socioeconomic survey of nearly 4,000 students entering 28 highly selective colleges in 1999—to evaluate explanations of minority underachievement in college. The authors develop two measures of vulnerability to stereotype threat. The authors focus on the stereotype threat hypothesis, and conclude that stereotype threat undermines minority achievement. We replicate their results, and investigate the robustness of their specification. We find that their main conclusion is fragile: controlling for other demographic characteristics, the effect of stereotype threat becomes statistically insignificant.

2 Introduction

In 1961, President Kennedy signed affirmative action into law with Executive Order 10925, and in 1968, the Office of Federal Contract Compliance (OFCC) defined its parameters. As West (1998) explains, the OFCC mandated that employers "develop specific goals and timetables for the prompt achievement of full and equal employment opportunity." This also applied to higher education. Corporations and universities alike had an affirmative responsibility to employ and admit underrepresented minorities. Less than a decade later, *Regents of the University of California v. Bakke* was argued before the Supreme Court. *Bakke* challenged a "special admissions program" at the Medical School of the University of California at Davis. With an opinion delivered by Justice Powell (1978), the Court ruled that the program operated as a quota system, and was thus unconstitutional, but that "the goal of achieving a diverse student body was sufficiently compelling to justify [the] consideration of race in admissions decisions." Affirmative action was upheld. Since *Bakke*, debate has abound regarding the merits of affirmative action in college admissions. Some critics contend that affirmative action compels universities to admit otherwise unqualified students, creating a mismatch between affirmative action beneificiaries and the academic rigor of their universities. Sander (2004) explores this hypothesis in law schools, finding that "the admission preferences extended to blacks [...] do not successfully identify students who will perform better than one would predict based on their academic indices, [and] consequently, most black law applicants end up at schools where they will struggle academically and fail at higher rates than they would in the absence of preferences." However, David Chambers and Lempert (2005) conclude that Sander's findings arise out of "a series of statistical errors, oversights, and implausible assumptions," and that "in neither the most elite schools, nor the least elite schools does the mismatch theory find support." Others, like West (1998), invoke ethical defenses, contending that affirmative action is necessary to remedy "the continuing impact of [...] notions of the racial superiority of whites and the racial inferiority" of people of color. Others still allege that affirmative action is a form of reverse-discrimination, and functions takes spots at these elite universities away from "more-qualified candidates."[1] Douglas Massey and Fischer (2003) explore these critiques and others in *The Source of the River: The Social Origins of Freshmen at America's Selective Colleges and Universities.*

The Source of the River was born out of the National Longitudinal Survey of Freshmen (NLSF), a "systematic, nationally representative study that investigated the determinants of college success for different racial and ethnic groups by surveying representative samples of white, Asian, Latino, and black freshmen entering a set of twenty-eight selective colleges and universities in the fall of 1999." As D. Massey and Charles explain, the

---

[1] Leonard (1974)

NLSF was designed to provide a "basis for linking pre-college experiences to behaviors and psychological states emerging in the course of higher education, and for sorting out the direction of causality between determinants and outcomes." The NLSF asked thirty-five schools to participate, and determined target sample sizes at each institution from the relative size of its black population. At schools with a black student body in excess of 1,000, the NLSF targeted 280 students for interviews—70 from each of four racial groups: white, black, Asian, and Latino. At schools with smaller black populations, the study targeted fewer students. Included in the call for participation were five historically black colleges and universities, but only Howard University was able to participate. In all, 3,924 students completed the survey, with 959 Asian respondents, 998 white respondents, 1,051 African American respondents, and 916 Latino respondents. The survey collected information regarding respondents' background, including information about their upbringing, childhood, school experience, future aspirations, academic history, and attitudes. Douglas Massey and Fischer (2003) use this data to assess theoretical explanations for minority underachievement. They consider: the theory of capital deficiency, which posits that minority students lack the resources necessary to succeed; the theory of oppositional culture, which, per Ogbu (1986), argues that academic success burdens minorities, because it is associated with "acting white," and confers feelings of racelessness; and the theory of stereotype threat, which, per Claude M. Steele (1995), contends that the stigmatization of minorities as intellectually inferior induces negative psychological pressure that impairs minority student achievement.

To conduct this analysis, Douglas Massey and Fischer (2003) regress academic outcomes on race, and then add indicators of college preparation along four dimensions—academic, financial, social, and psychological—and controls for three psychological vulnerabilities—overconfidence, stereotype threat, and racelessness—to their specification. They examine the differences in intergroup variation in outcomes with and without controls.

As originally posited by Claude M. Steele (1995), the "threat" in stereotype threat arises out of the fear of "confirming, as a self-characteristic, a negative stereotype about one's group." Levied as a criticism of affirmative action, this hypothesis suggests that in higher education, affirmative action stereotypes minorities as "academically challenged and intellectually weak," inducing negative psychological pressure on minority students and undermining their ability to succeed. Douglas Massey and Fischer (2003) evaluate this hypothesis among the cohort of freshmen studied in the NLSF by constructing two indices of vulnerability to stereotype threat. The first index pinpoints black and Latin students that self-identified as extremely self-conscious about their teachers' perceptions of them, and did not regard themselves as very good students (i.e they said it was "not true" or only "somewhat true" that they were good students). The second index identifies black and Latino students who reported the belief that being American ought to be more important to minorities than being black, Latin, or Asian, and also gave their own racial group an unintelligence rating of five or more on a ten-point scale. The authors find that stereotype vulnerability—as measured by the first index—has a negative and statistically significant effect on first semester grade point average. Indeed, those "black and Latino students who doubted their own abilities and were simultaneously sensitive to the views of their teachers earned a GPA that was 0.122 point lower than that earned by other minority students."

In addition to regressing academic outcomes on these indicators of preparation, Douglas Massey and Fischer (2003) also regress each indicator of preparation on demographic characteristics, ranging from family structure to degree of school segregation to peer environment. They find that "differences in background and upbringing documented [in the NLSF] translate into differences in college preparation." Along all four preparatory dimensions, controlling for background differences mitigates, if not eliminates, intergroup variation. Despite this, Douglas Massey and Fischer (2003) do not include these background characteristics in their specifications for academic outcomes. We investigate the robustness of their findings regarding stereotype threat, and find that changes in the regression specification change their results. Controlling for background differences, intergroup variation in first-semester grade point average is further reduced, and the effect of stereotype threat—measured by the first index—becomes statistically indistinguishable from zero.

3 Data and Measurement

Our data come from the National Longitudinal Survey of Freshmen, a "probative sample of students who entered 28 selective U.S. colleges and universities as freshmen in the Fall of 1999." A complete description of the sampling methodology and the Wave 1 questionnaire itself can be found in *The Source of the River*

Table 1: Dimensions and Indicators of Preparation for Higher Education

Dimension and Indicator	Replicated				Published			
	Whites	Asians	Latinos	Blacks	Whites	Asians	Latinos	Blacks
Academic								
Number of AP Courses Taken	3.55	4.19	3.20	2.59	3.01	4.24	2.62	2.03
High School GPA	3.78	3.79	3.70	3.55	3.77	3.83	3.70	3.53
Self-Rated Preparation	6.59	6.04	5.66	6.13	6.73	6.28	6.15	6.49
Financial								
Share of Cost Paid by Family	0.74	0.69	0.52	0.38	0.74	0.68	0.54	0.42
Social								
Susceptibility to Peer Influence	15.31	15.14	15.52	15.35	9.82	10.55	9.22	10.38
Social Distance from Minorities	14.26	15.03	11.94	10.88	14.18	15.25	11.81	10.69
Social Distance from Whites	8.62	10.87	10.87	15.07	8.17	11.23	10.44	15.75
Psycholoigcal								
Self-Esteem	32.00	30.31	32.27	33.76	32.15	29.78	32.55	33.86
Self-Efficacy	18.90	18.03	18.99	19.08	18.96	17.88	18.94	19.04
Self-Confidence	34.63	36.04	35.43	36.35	34.57	36.01	35.42	36.65
Freshman Fall GPA	3.34	3.31	3.07	2.97	3.31	3.29	3.05	2.95

[a] This table replicates the results from page 314 of The Source of the River. The data come from the National Longitudinal Survey of Freshmen: Waves 1, 2, 3, and 5. First-semester college grade point average comes from self-reported grades, collected in the Spring of 2000 in Wave 2 of the NLSF. Each indicator of preparation is based on a summated rating scale, the construction of which is detailed in appendix A. As MCLF describe on page 153, each item in the each index was assigned a numerical value to indicate its relative weight, and then the values were across items to construct the summated ratings scale. The 'Published' columns come from The Source of the River. The key result is that we are able to near-perfectly replicate 9 of 11 calculations from The Source of the River.

(Douglas Massey and Fischer (2003)), but we note that the baseline survey was conducted via face-to-face interviews. It gathered extensive information about respondents' lives, environments, and upbringings prior to college, and provides most of our independent variables. Almost all of the variables used to construct the ten indicators of preparation Douglas Massey and Fischer (2003) define are recorded in this first survey. The dependent variable that we focus on, first semester grade point average, was defined in the second wave of the NLSF, in Spring 2000. We calculated first semester GPA on a 4.00 scale from the self-reported grades of respondents. Qualitative grades (i.e. Pass, Satisfactory, Fail) were excluded from consideration. We were able to calculate GPA's for 3,640 of 3,924 respondents.

3.1 Indicators of Academic Preparation

Douglas Massey and Fischer (2003) conceptualize indicators of college preparation along four dimensions: academic, financial, social and psychological. They construct three indicators of academic preparation: number of AP courses taken in high school, high school grade point average, and the respondent's own assessment of his/her level of preparation. High school GPA is estimated on a 4.00 scale, from self-reported grades collected in the baseline survey. Preparation is scaled from 0—10, with 10 being the most prepared, and 0 the least.

In Wave 3 of the NLSF, students reported the amount of money they needed to pay for college that year, and how much of that money they could expect to come from their parents, other family members, personal savings, and work-study earnings. To control for financial preparation, we calculate the fraction of money needed that the student reported would come from one of those four sources (i.e. would come from his/her family).

Douglas Massey and Fischer (2003) specify three indicators of social preparation: susceptibility to peer influence, social distance from minorities, and social distance from whites. To index susceptibility to peer influence, they code respondents' degree of agreement with statements such as "I thought and acted like others," "I felt comfortable with others," and "I did things so that others would like me," and then construct a summated ratings scale. Each of item was coded from 0—4, such that higher scores indicated greater

sensitivity to peer influence. Social distance from minorities was measured using "0—10 closeness ratings" with respect to six targets: blacks in general, young black men, young black women, latinos in general, young latin men, and young latin women. Higher values indicated a greater degree of perceived closeness. To tabulate social distance from whites, we codified perceived closeness to whites in general, young white men, and young white women. Delpit (1988) finds that the codes, rules and norms enacted in classrooms are of a "culture of power" dictated by middle-class white society, such that success in schools is "predicated upon acquisition of [this] culture of power." As such, social distance from whites could have academic implications in multiple forms. Beyond preparing students to interact with the white peers they undoubtedly encountered at these elite universities, it's possible that there's a relationship between exposure to white people and acquisition of the culture of power Delpit (1988) investigates.

Finally, Douglas Massey and Fischer (2003) create three metrics of psychological preparation: self-esteem, self-efficacy, and self-confidence. To assess self-esteem and self-efficacy, the NLSF developed statements from standard measures of self-esteem and self-efficacy developed by Rosenberg and Simmons (1971), and asked respondents to report their degree of agreement. From this data, the authors constructed two summated rating scales. The index for self-esteem was based on 10 statements, including: "I am a person of worth, equal to others," "I am inclined to feel that I am a failure," and "I do not have much to be proud of." The index for self-efficacy was based on responses to six statements, including "I don't have control over my life" and "I feel left out of things going on around me." Responses were coded from 0—4, with higher values indicating higher degrees of self-esteem or self-efficacy. The index for self-confidence was constructed from student responses to four hypotheticals. Respondents evaluated the likelihood that they would: finish two years of college, graduate from college, begin postgraduate education, and finish a graduate and/or professional degree. Students' responses were scaled from 0—10, with 10 indicating "very likely" and 0 indicating "very unlikely." As before, values were summed across items to create a summated ratings scale for self confidence, with a higher value corresponding to a higher degree of self-confidence.

3.2 Psychological Vulnerabilities

In addition to controlling for indicators of college preparation, Douglas Massey and Fischer (2003) control for three psychological vulnerabilities: overconfidence, stereotype threat, and racelessness, with five dichotomous variables. In this paper, we focus on stereotype threat. Douglas Massey and Fischer (2003) hypothesize that "blacks [and] Latinos who expressed reservations about their academic abilities and were unusually self-conscious of teachers would be at elevated risk of stereotype threat, along with [minority] students who expressed doubts about their [own] group's ability while adhering to a mainstream "American" identity."

The authors' first index identifies black and Latino students of the first camp. Students that reported being "extremely self-conscious about their teachers' perceptions of them, and second, did not consider themselves very good students" (i.e they said it was "not true" or only "somewhat true" that they were good students) were considered particularly vulnerable to stereotype threat, as measured by this first index. The second index identifies black and Latino students in the second camp. Respondents that report believing that being American ought to be more important to minorities than being black, Latin, or Asian, and also gave their own racial group an unintelligence rating of five or more on a ten-point scale, were designated vulnerable to stereotype threat per the second index. Each index was represented by a dichotomous variable. These variables were coded 1 for Black and Latinos with vulnerability to stereotype per that specific index, and 0 otherwise. Of the 1051 black respondents, 58 displayed a vulnerability to stereotype threat per the first index. Among the 916 Latin respondents, 38 displayed this same vulnerability. 39 Latino respondents, and 37 black respondents were identified as vulnerable per the second index. Very few students were displayed vulnerability per both indices. As Douglas Massey and Fischer (2003) note, a vast majority of blacks and Latinos are "not at serious risk of succumbing to stereotype vulnerability," at least as crudely defined by these indices.

As Claude M. Steele (1995) explains, stereotype threat arises out of the "immediate situational threat that derives from the broad dissemination of negative stereotypes about one's group." The authors assume that affirmative action propagates no negative stereotypes regarding the academic capabilities of white and Asian students, and as such, both indices for stereotype vulnerability are coded 0 for Asian and white respondents,

in line with the later practices of M. J. Fischer and Massey (2006).

3.3 Outcome Measures

Douglas Massey and Fischer (2003) consider the effects of college preparation and psychological vulnerability on three measures of academic outcomes: first semester grade point average, likelihood of dropping a course during first semester of freshman year, and likelihood of failing a course during first semester of freshman year. We focus on the first measure: fall GPA, which is derived from "retrospective self-reports of courses taken and grades earned." Douglas Massey and Fischer (2003) match self-reported grades to data from official transcripts, and find a "high degree of reliability in self-reporting." In fact, 82.8% of self-reported grades matched their recorded counterparts perfectly.

These self-reports reflected grades earned in individual courses. As such, we calculated fall GPA manually, on a 4.00 scale. We did not distinguish between A+'s, A's, and A-'s; B+'s, B's, and B-'s, and so forth. In our calculations, A's equalled 4.00 grade points, B's equalled 3.00 points, C's equalled 2.00 points, D's equalled 1.00 point, and E's and F's equalled 0.00 points. We weighted all courses as the same, and excluded qualitative grades (i.e. Pass/Satisfactory, Pass, or Fail) from consideration. We were able to calculate 3,640 of the 3,924 respondents fall GPA's.

3.4 Demographic and Background Characteristics

To assess the robustness of the findings in *The Source of the River,* we regress first semester grade point average on indicators of preparation, psychological vulnerabilities, and other background characteristics. We specify the same thirty-nine variables Douglas Massey and Fischer (2003) specify when they regress indicators of preparation. As do Douglas Massey and Fischer (2003), we create sets of dichotomous variables for family structure, immigrant origins, household income, parental education, poverty status, kind of school, and degree of school segregation. These variables are coded 1 when the condition is true, and 0 otherwise. For parental child-rearing strategies, indicators of school quality, and peer environment, we create summated rating scales following the indices detailed in Appendix B of *The Source of the River.*

As shown in Table 1, we are successful in replicating nine of the authors' eleven calculations. We are unable to perfectly replicate the number of AP courses taken in high school, and susceptibility to peer influence. Most importantly, with rounding to the nearest tenth, we are able to perfectly replicate the authors' calculations for first semester college grade point average, our academic outcome of study.

4 Analysis

As detailed earlier, Douglas Massey and Fischer (2003) explore intergroup variation in three academic outcomes: first semester grade point average, likelihood of dropping a class, and likelihood of failing a class. They first regress these outcomes on a set of "dichotomous variable indicating [racial] group membership," and then add controls for indicators of preparation and psychological vulnerabilities. Their conclusions are based on the differences in variation observed before and after the addition of controls. We successfully replicate these regressions. Our coefficients are nearly identical to those shown on page 453, with the exception of the coefficient on share of cost paid by family. MCLF find that this indicator has a positive and statistically significant effect on first-semester GPA at the 0.05 level. Our result is statistically indistinguishable from 0.

Douglas Massey and Fischer (2003) find that "during the very first semester of college there are clear and significant differences in academic performance that emerge between groups." Indeed, their results suggest that while grades earned by Asians and Whites are, on average, statistically indistinguishable from one another, blacks earn grades a third of a grade point lower than their white peers, and latinos earn grades over a quarter of a grade point lower. Even with controls for indicators of preparation and psychological vulnerabilities, grades earned by blacks were, on average, over a fifth of a grade point lower than those of whites, and grades earned by latinos were still one sixth of a grade point lower. Douglas Massey and Fischer (2003) also find that vulnerability to stereotype threat—as defined by their first index—has a negative and

statistically significant effect on first semester GPA. Black and Latino respondents with this vulnerability earned grades, on average, 0.13 grade points lower than their counterparts without such a vulnerability.

Table 2: Effect of Indicators of Preparation on First Semester Grade Point Average

	Dependent variable:	
	Group Effects	MCLF Full Model
	(1)	(2)
Black	−0.363*** (0.025)	−0.253*** (0.034)
Latin	−0.269*** (0.026)	−0.162*** (0.029)
Asian	−0.029 (0.026)	−0.009 (0.027)
White		
AP Courses Taken		0.016*** (0.005)
High School GPA		0.424*** (0.033)
Self-Rated Preparation		0.047*** (0.003)
Share of Cost Paid by Family		0.034 (0.028)
Susceptibility to Peer Influence		−0.013*** (0.002)
Social Distance from Minorities		0.005** (0.002)
Social Distance from Whites		0.0003 (0.002)
Self-Esteem		0.003 (0.002)
Self-Efficacy		−0.004 (0.004)
Self-Confidence		0.004* (0.002)
Overconfidence		0.034 (0.040)
Stereotype Threat Index 1		−0.175*** (0.063)
Stereotype Threat Index 2		0.077 (0.061)
Racelessness Index 1		0.133 (0.101)
Racelessness Index 2		0.107** (0.041)
Constant	3.337*** (0.018)	1.307*** (0.155)
Observations	3,640	2,595
R^2	0.076	0.241
Adjusted R^2	0.075	0.236
Residual Std. Error	0.546 (df = 3636)	0.475 (df = 2576)
F Statistic	99.565*** (df = 3; 3636)	45.409*** (df = 18; 2576)

Note: *p<0.1; **p<0.05; ***p<0.01

This regression uses data from the National Longitudinal Survey of Freshmen (NLSF), and replicates Table 9.1 from The Source of the River, found on page 453. These results are nearly identical to those shown on page 453, with the exception of the coefficient on share of cost paid by family, which MCLF find is statistically significant at the 0.05 level. The dependent variable, first semester college GPA, was calculated on a 4.00 scale from respondents' retrospective self-reports of courses taken and grades earned. Qualitative grades were not considered in these calculations. Two indicators of vulnerability to stereotype threat were developed. The first identifies black and Latino students of the first camp that simultaneously reported being "extremely self-conscious about their teachers' perceptions of them, and second, did not consider themselves very good students" (i.e they said it was "not true" or only "somewhat true" that they were good students). The second identifies black and Latino students that reported believing that being American ought to be more important to minorities than being black, Latin, or Asian, and also gave their own racial group an unintelligence rating of five or more on a ten-point scale. Each index was represented by a dichotomous variable, which was coded 1 if the respondent displayed the vulnerability, and 0 otherwise. Of the 1051 black respondents, 58 displayed a vulnerability to stereotype threat per the first index. Among the 916 Latin respondents, 38 displayed this same vulnerability. 39 Latino respondents and 37 black respondents were identified as vulnerable per the second index. Very few students were displayed vulnerability per both indices.

The authors' first index identifies black and Latino students of the first camp. Students that reported being "extremely self-conscious about their teachers' perceptions of them, and second, did not consider themselves very good students" (i.e they said it was "not true" or only "somewhat true" that they were good students) were considered particularly vulnerable to stereotype threat, as measured by this first index. The second index identifies black and Latino students in the second camp. Respondents that report believing that being

6

American ought to be more important to minorities than being black, Latin, or Asian, and also gave their own racial group an unintelligence rating of five or more on a ten-point scale, were designated vulnerable to stereotype threat per the second index. Each index was represented by a dichotomous variable. These variables were coded 1 for Black and Latinos with vulnerability to stereotype per that specific index, and 0 otherwise. Of the 1051 black respondents, 58 displayed a vulnerability to stereotype threat per the first index. Among the 916 Latin respondents, 38 displayed this same vulnerability. 39 Latino respondents, and 37 black respondents were identified as vulnerable per the second index. Very few students were displayed vulnerability per both indices.

Earlier in their paper, however, Douglas Massey and Fischer (2003) note that college preparation is determined by pre-existing demographic and background characteristics. Indeed, they write:

> "Given the intergroup variation we have documented with respect to demographic characteristics, child-rearing practices, school conditions, peer environments, and social attitudes, it would be surprising if such background differences were *not* related in some way to differences in preparation between groups. We thus begin our assessment of how preparation affects achievement by attempting to account for intergroup differences in preparation using the various background measures we have collected."

In this analysis, Douglas Massey and Fischer (2003) find that intergroup differences in college preparation are mitigated when background differences are controlled for statistically, and, on some occasions, variation is eliminated. They observe that different background factors are more or less salient in the determining different kinds of preparation. While academic preparation in terms of AP credits is "determined basically by parental education, wealth, and child-rearing patterns, [...] the most important determinants of [high-school] grades earned were the academic and peer environment the respondent experienced in high school."

Douglas Massey and Fischer (2003) come to these conclusions by regressing each indicator of preparation on thirty-nine background characteristics, but do not include such demographic variation in their specifications for academic outcomes. Their ultimate regression equation for first-semester GPA includes but nineteen control variables: the four races included in the NLSF, their ten indicators of level of preparation, and five indices of vulnerability:

$$GPA = \beta_0 + \beta_1 Black + \beta_2 Latin + \beta_3 Asian + \beta_4 APCourses + \beta_5 HighSchoolGPA + \beta_6 Preparation+$$

$$\beta_7 FamilyContribution + \beta_8 PeerInfluence + \beta_9 MinorityDistance + \beta_{10} WhiteDistance+$$

$$\beta_{11} SelfEsteem + \beta_{12} SelfEfficacy + \beta_{13} SelfEfficacy + \beta_{14} SelfEfficacy + \beta_{15} Overconfidence+$$

$$\beta_{14} Stereotype1 + \beta_{14} Stereotype2 + \beta_{15} Racelessness1 + \beta_{16} Racelessness2 + \mu$$

We investigate the robustness of their results to changes in the specification, and find that they are fragile. The effect of stereotype vulnerability, as measured by the first index, becomes statistically insignificant once controls for background characteristics are introduced.

Douglas Massey and Fischer (2003)'s findings that controlling for pre-existing demographic differences mitigates, but does not eliminate, intergroup variation in first semester grade point average is upheld. As we see in Table 4, average first semester GPA has a standard deviation of 0.59 for Latino students, and 0.57 for black students. We evaluate the magnitude of intergroup variation in fall GPA in light of this.

Without controls for preparation, vulnerabilities, or other background characteristics, black students earn, on average, grades 0.36 grade points *less* than their white peers. At just under two-thirds of a standard deviation, this effect is sizeable. Controlling for indicators of preparation reduces the mangitude of this effect, but it remains negative and significant. With controls for preparation, black students earn, on average, grades a quarter of a grade point lower, and with additional controls for other background characteristics, this effect is reduced only slightly, to 0.23. In both cases, the coefficient for black students is approximately two-fifths of a standard deviation.

Latino students, without controlling for preparation, vulnerability, or background, earn grades just over a quarter less than their white peers. This effect is almost half a standard deviation. With controls for

Table 3: Effect of Indicators of Preparation on First Semester Grade Point Average

	Dependent variable:		
	Group Effects	MCLF Model	Full Model
	(1)	(2)	(3)
Black	−0.363*** (0.025)	−0.253*** (0.034)	−0.231*** (0.044)
Latin	−0.269*** (0.026)	−0.162*** (0.029)	−0.098** (0.040)
Asian	−0.029 (0.026)	−0.009 (0.027)	0.003 (0.042)
White			
AP Courses Taken		0.016*** (0.005)	0.011 (0.007)
High School GPA		0.424*** (0.033)	0.368*** (0.045)
Self-Rated Preparation		0.047*** (0.003)	0.042*** (0.005)
Share of Cost Paid by Family		0.034 (0.028)	−0.072* (0.043)
Susceptibility to Peer Influence		−0.013*** (0.002)	−0.009*** (0.003)
Social Distance from Minorities		0.005** (0.002)	0.004 (0.003)
Social Distance from Whites		0.0003 (0.002)	0.001 (0.003)
Self-Esteem		0.003 (0.002)	0.005* (0.003)
Self-Efficacy		−0.004 (0.004)	−0.004 (0.005)
Self-Confidence		0.004* (0.002)	0.003 (0.003)
Overconfidence		0.034 (0.040)	−0.031 (0.054)
Stereotype Threat Index 1		−0.175*** (0.063)	−0.138 (0.084)
Stereotype Threat Index 2		0.077 (0.061)	−0.007 (0.080)
Racelessness Index 1		0.133 (0.101)	0.226* (0.124)
Racelessness Index 2		0.107** (0.041)	0.096 (0.059)
Male			−0.033 (0.026)
Religiosity			−0.005 (0.005)
Number of Siblings			−0.003 (0.012)
Two Parents All Ages			0.064* (0.037)
Two Parents Some Ages			0.470** (0.182)
Single Parents All Ages			0.419** (0.185)
Both Parents Natives			−0.009 (0.108)
Foreign Born Parent			0.041 (0.106)
Foreign Born Respondent			−0.015 (0.038)
Less Than 14,999			−0.132 (0.092)
15,000 — 24,999			−0.048 (0.072)
25,000 — 49,000			−0.096** (0.039)
50,000 — 74,999			−0.051 (0.034)
75,000 or More			
Neither Parent College Grade			−0.060 (0.042)
One Parent College Grad			
Two Parents College Grade			0.015 (0.034)
One Parent Advanced Degree			0.047 (0.032)
Both Parents Advanced Degree			0.048 (0.033)
Value in 100,000			−0.001 (0.006)
Ever on Welfare			0.018 (0.048)
Cultivation of Human Capital			−0.003* (0.002)
Cultivation of Social Capital			−0.002 (0.004)
Cultivation of Cultural Capital			0.001 (0.002)
Cultivation of Intellectual Independence			0.001 (0.002)
Strictness of Discipline			−0.008*** (0.002)
Use of Shame of Guilt			0.001 (0.003)
Private			−0.033 (0.030)
Less Than 20 Percent Minority			0.112** (0.054)
20 — 39 Percent Minority			0.065 (0.055)
40 — 59 Percent Minority			0.041 (0.056)
60 — 79 Percent Minority			−0.003 (0.068)
Greater Than 80 Percent Minority			
Infrastructure			0.001 (0.004)
Academic Support			−0.006 (0.006)
Disorder/Violence Index 00			0.0002 (0.0003)
Support for Academic Effort			0.001 (0.004)
Support for Intellectual Independence			0.005 (0.004)
Support for Delinquency			−0.009** (0.004)
Constant	3.337*** (0.018)	1.307*** (0.155)	1.163*** (0.317)
Observations	3,640	2,595	1,638
R²	0.076	0.241	0.291
Adjusted R²	0.075	0.236	0.267
Residual Std. Error	0.546 (df = 3636)	0.475 (df = 2576)	0.468 (df = 1583)
F Statistic	99.565*** (df = 3; 3636)	45.409*** (df = 18; 2576)	12.056*** (df = 54; 1583)

Note: *p<0.1; **p<0.05; ***p<0.01

This regression uses data from the National Longitudinal Survey of Freshmen (NLSF), and specifies a set of dichotomous variables for race, indicators of preparation, psychological vulnerabilities, and other background characteristics. Background characteristics were selected from the regression specifications found in Chapter 8 of The Source of the River. First semester college GPA was calculated on a 4.00 scale from respondents' retrospective self-reports of courses taken and grades earned. Qualitative grades were not considered in these calculations. Two indicators of vulnerability to stereotype threat were developed. The first identifies black and Latino students of the first camp that simultaneously reported being "extremely self-conscious about their teachers' perceptions of them, and second, did not consider themselves very good students" (i.e they said it was "not true" or only "somewhat true" that they were good students). The second identifies black and Latino students that reported believing that being American ought to be more important to minorities than being black, Latin, or Asian, and also gave their own racial group an unintelligence rating of five or more on a ten-point scale. Each index was represented by a dichotomous variable, which was coded 1 if the respondent displayed the vulnerability, and 0 otherwise.

Table 4: First Semester GPA: Standard Deviation and Deficit

Relation to GPA	Asians	Latinos	Blacks
Standard Deviation	0.52	0.59	0.57
Deficit: Race Only	-0.03	-0.27	-0.36
Deficit: Massey (2003) Controls	-0.01	-0.16	-0.25
Deficit: All Background Controls	0.00	-0.10	-0.23

[a] This table shows the standard deviation of first semester grade point average for Asians, Latinos, and Blacks, along with the magnitude of the deficit associated with each group as regression specifications changed.

indicators of preparation and vulnerabilities, Latino students earn grades just 0.16 grade points less than their peers, and with additional background controls, they earn grades, on average, just one-tenth lower than their white peers. In this case, we see that the intergroup variation in fall GPA is markedly reduced. At just one sixth of a standard deviation, this -0.10 effect is quite small, although significant in statistical terms. These changes in the magnitude of intergroup variation with the addition of controls are represented visually in Figure 1.

Although intergroup variation persists, the effect of stereotype threat, per our first index, goes from negative and significant at the 0.01 level, to statistically indistinguishable from 0. We also see that high school GPA and self-rated preparation continue to have positive, and statistically significant effects on first-semester grade point average, after the introduction of other background variables. The effect of number of AP courses taken in high school becomes statistically indistinguishable from 0. We also find that among the variables introduced in our final model, each variable for family structure (two parents all ages, two parents some ages, and single parent all ages) had statistically significant effects on GPA. The coefficients for all three of these variables were positive.

Figure 1: Deficit in GPA Earned by Minorities Compared with Whites
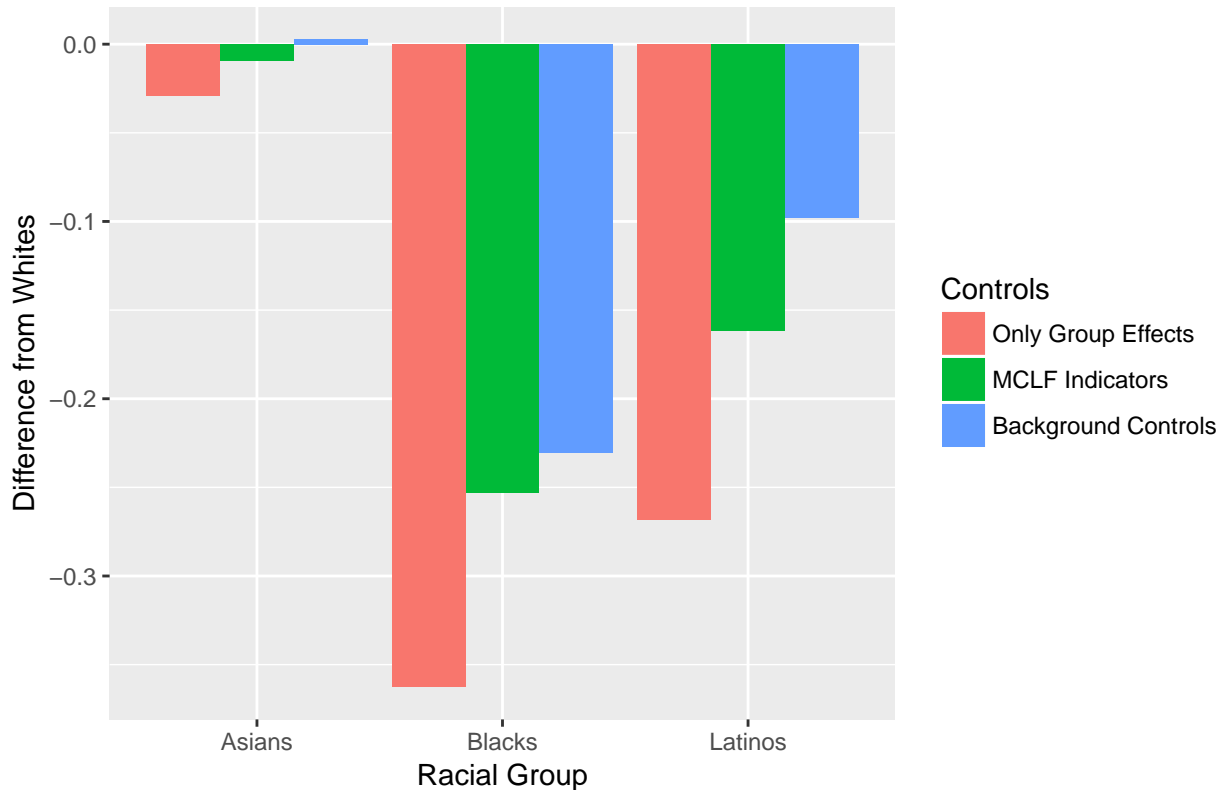
5 Conclusion and Discussion

Critics of affirmative action have levied a variety of charges against race-conscious admissions policies. We focus on what Douglas Massey and Fischer (2003) call the stereotype threat hypothesis, which posits that affirmative action stigmatizes all minority students as intellectually inferior, creating negative psychological pressure and undermining minority performance. To evaluate this proposition, Douglas Massey and Fischer (2003) used data from the National Longitudinal Survey of Freshmen, and created indexes of vulnerability to stereotype threat. Their first index identified "blacks [and] Latinos who expressed reservations about their academic abilities and were unusually self-conscious of teachers" and their second index pinpointed black and Latino students "who expressed doubts about their [own] group's ability while adhering to a mainstream "American" identity." The authors hypothesized that students meeting either of these criterion would be particularly vulnerable to stereotype threat. Douglas Massey and Fischer (2003) used these indicators to estimate first semester college grade point average, and found evidence of stereotype threat—measured by their first index—undermining minority achievement.

We analyzed the robustness of this finding in the face of changes to the regression specification, and found that the authors' results were fragile. Controlling for background differences, the effect of stereotype threat became statistically indistinguishable from zero. Our analysis yielded no support whatsoever for the stereotype threat hypothesis. Perhaps, it is unsurprising that we've reached this result. Douglas Massey and Fischer (2003) themselves find that there are sizeable and statistically significant relationships between various background characteristics and the indicators of preparation for which they control. And, as ("PAD 705 Handout: Omitted Variable Bias" 2003) explain, omitted variable bias occurs "when a variable that is correlated with both the dependent *and* one or more *included* independent variables is omitted from a regression equation." Future analyses on this topic could explore the extent to which the indicators of stereotype threat developed by Douglas Massey and Fischer (2003) were correlated with various background characteristics introduced.

We did, however, find support for the broader findings of Douglas Massey and Fischer (2003). Even with

controls for indicators of preparation, psychological vulnerabilities, *and* other background characteristics, we observed intergroup variation in first-semester grade point average. As others have found, our estimates suggest that academic preparation is the principal determinant of first semester GPA. Indeed, M. J. Fischer and Massey (2006) find, as we did and as Douglas Massey and Fischer (2003) did, that high school GPA has a highly significant and positive effect on first semester GPA for black and Latino students. Further, M. J. Fischer and Massey (2006) include individual SAT score in their regression specification, and find that it, too, has a significant and positive effect on GPA. Future analyses could include more indicators of academic preparation.

An array of theories exist that could explain the intergroup variation that persists in the face of controls. Perhaps, as White and Lowenthal (2011) explain, "minority students openly resist the adoption of the very discursive skills they need to thrive [in] college," and its this unconscious decision making that undermines their success. Or, as Feagin (1992) postulates, perhaps the "sustained obstacle of cumulative discrimination" characterizes the college experiences of minority students, and prevents them from achieving. It's also quite possible that it's some mix of these and other hypotheses. Our specification does not enable us to empirically evaluate those theories, but, we can conclude that it is not stereotype threat that is responsible for the variation that we continue to observe.

References

Claude M. Steele, Joshua Aronsonl. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans" 69. Journal of personality; social psychology: 797–811.

David Chambers, William Kidder, Timothy Clydesdale, and Richard Lempert. 2005. "The Real Impact of Eliminating Affirmative Action in American Law Schools: An Empirical Critique of Richard Sander's Study." *Stanford Law Review* 57. Stanford Law Review: 1855–98.

Delpit, Lisa D. 1988. "The Silenced Dialogue: Power and Pedagogy in Educating Other People's Children." *Harvard Educational Review* 58. President; Fellows of Harvard College: 132–59.

Douglas Massey, Garvey Lundy, Camille Charles, and Mary Fischer. 2003. "The Source of the River: The Social Origins of Freshmen at America's Selective Colleges and Universities." Princeton University Press, 1–607.

Feagin, Joe. 1992. "The Continuing Significance of Racism: Discrimination Against Black Students in White Colleges" 22. Journal of Black Studies: 546–78.

Fischer, Mary J., and Douglass S. Massey. 2006. "The Effects of Affirmative Action in Higher Education" 36. Social Science Research: 531–49.

Leonard, H. Jeffrey. 1974. "The 'Reverse Discrimination' Backlash." The Harvard Crimson.

Massey, Douglas, and Camille Charles. "About: National Longitudinal Survey of Freshmen." Princeton University.

Ogbu, John. 1986. "Black Students' School Success: Coping with the Burden of 'Acting White'." *The Urban Review* 18: 176–206.

"PAD 705 Handout: Omitted Variable Bias." 2003. Rockefeller College of Public Affairs; Policy.

Powell, Lewis F. 1978. "Judgement of the Court." Cornell Law School.

Rosenberg, Morris, and Roberta Simmons. 1971. "Summated Rating Scale Construction : An Introduction." *American Sociological Association.* University of Michigan, Scholarly Publishing Office, 132–59.

Sander, Richard. 2004. "A Systematic Analysis of Affirmative Action in American Law Schools." *Stanford Law Review* 57. Stanford Law Review: 367–83.

West, Martha S. 1998. "The Historical Roots of Affirmative Action." *Berkeley La Raza Law Journal* 10: 607–30.

White, John, and Patrick Lowenthal. 2011. "Minority College Students and Tacit 'Codes of Power': Developing Academic Discourses and Identities" 34. The Review of Higher Education: 283–318.