Stereotype Threat and College Academic Performance: A Latent Variables Approach*

Jayanti Owens and **Douglas S. Massey**
Office of Population Research, Princeton University

Abstract

Stereotype threat theory has gained experimental and survey-based support in helping explain the academic underperformance of minority students at selective colleges and universities. Stereotype threat theory states that minority students underperform because of pressures created by negative stereotypes about their racial group. Past survey-based studies, however, are characterized by methodological inefficiencies and potential biases: key theoretical constructs have only been measured using summed indicators and predicted relationships modeled using ordinary least squares. Using the National Longitudinal Survey of Freshman, this study overcomes previous methodological shortcomings by developing a latent construct model of stereotype threat. Theoretical constructs and equations are estimated simultaneously from multiple indicators, yielding a more reliable, valid, and parsimonious test of key propositions. Findings additionally support the view that social stigma can indeed have strong negative effects on the academic performance of pejoratively stereotyped racial-minority group members, not only in laboratory settings, but also in the real world.

The academic underperformance of racial and ethnic minorities at selective colleges and universities is well-documented (Charles et al. 2009; Bowen and Bok 1998; Massey et al. 2003). Minority students, particularly black and Hispanic students, tend to perform less well on exams, graduate at lower rates, and earn lower grades than one would predict on the basis of objective characteristics such as family income, parental education, and SAT scores (Espenshade and Walton-Radford 2009). In particular, black students perform less well than equally-qualified whites, even after introducing exhaustive controls for background differences (Bowen and Bok 1998). A variety of explanations have been advanced to account for racial gaps in academic performance, including intergroup differences in resources (Massey et al. 2003), oppositional identity (Fordham and Ogbu 1986), and stereotype threat (Steele 1988a).

Evidence suggests that differences in access to financial, social, human, and cultural capital play an important role in explaining academic performance differentials; but they by no means eliminate intergroup gaps (Massey et al. 2003). Quantitative evidence in support of oppositional identity—the idea that minorities resist educational achievement as a racial betrayal and "acting white"—is remarkably weak (Downey 2008). More support has been found for stereotype threat—the view that minority students underperform because of

---

*Please direct communications to Jayanti Owens, 284 Wallace Hall, Office of Population Research, Princeton University, Princeton, NJ 08544 or jowens@princeton.edu.

pressures created by negative stereotypes about their group. Steele's original studies examine how black undergraduates at Stanford responded to race primes in carrying out everyday academic tasks (Steele 1988a; Steele 1988b). Since then, the stereotype threat model has been validated in numerous laboratory experiments (Major and O'Brien 2005; Spencer, Logel, and Davies forthcoming), intervention field experiments (Aronson, Fried, and Good 2002; Cohen et al. 2006, 2009; Steele 1997; Walton and Cohen 2007), and meta-analytic studies of laboratory and field experiments (Walton and Spencer 2009).

At the same time, other studies have called into question the validity of stereotype threat. For example, a recent field experiment primed high school students in a treatment group to think about their race and/or gender before taking higher-stakes Advanced Placement exams or Computerized Placement Tests and found no effects of race/gender priming on performance (Stricker and Ward 2004). Re-analysis of the Stricker and Ward (2004) study by Danaher and Crandall (2008), however, questioned Stricker and Ward's conclusions, instead arguing that their findings indeed support stereotype threat if a different interpretation of the results is applied. In general, then, experimental studies tend to support the theory of stereotype threat, and laboratory studies strongly support it.

In terms of non-experimental research, some observational studies point to a rather tenuous association between stereotypes and academic or occupational performance. For example, one recent study exploits patterns in military performance data and college grade performance—adjusting for scores on the Armed Services Vocational Aptitude Battery (ASVAB) and the Scholastic Aptitude Test (SAT), respectively—to see if they predict performance in military jobs and college GPA differentially for blacks and whites. It concluded that the relationships posited by stereotype threat do not hold in observational settings as well as they do in the laboratory (Cullen, Hardison, and Sackett 2004). Other studies use different domains, such as women in mathematics and geographic differences in the extent of stereotype threat, but come to similar conclusions, finding limited support for stereotype threat theory (Cullen, Waters, and Sackett 2006; Pope and Sydnor 2010). Although these studies underscore the fact that eliminating stereotypes cannot be expected to erase the black-white achievement gap in real-world settings entirely, they acknowledge that further testing of stereotype threat in non-experimental settings is needed (Cullen, Hardison, and Sackett 2004).

Other survey-based studies on stereotype threat have come to different conclusions, detecting relatively strong support for stereotype threat in real world settings in selective colleges and universities (Charles et al. 2009; Fischer and Massey 2007; Massey and Fischer 2005; Massey and Mooney 2007). Although these studies appear to confirm hypotheses derived from the stereotype threat model, they are nonetheless characterized by methodological shortcomings. For example, they measure key theoretical constructs using only summed indicators and predict relationships using ordinary least squares. These techniques lead not only to inefficiencies but also potential biases. Here we overcome these shortcomings by drawing on data from the National Longitudinal Survey of Freshmen (NLSF) to develop a latent construct model of stereotype threat in which theoretical constructs and equations are estimated simultaneously from multiple indicators, yielding a more reliable, valid, and parsimonious test of key propositions. By elaborating a model of the psychological processes by which stereotype threat undermines performance, our analysis lends additional support to the view that social stigma can indeed have strong negative effects on the academic performance of pejoratively stereotyped minority group members, not only in laboratory settings, but also in the real world.

STEROTYPE THREAT AND ACADEMIC PERFORMANCE

In their original analysis of NLSF data, Massey and Fischer (2005) synthesized findings from prior laboratory experiments to derive an Ordinary Least Squares model of stereotype threat that could be estimated using survey data. In their model specification, the existence of negative societal beliefs about a minority's intellectual ability was hypothesized to affect members' academic performance through two psychological pathways. Figure 1 shows the two pathways along with the direction of the effects hypothesized under the theory of stereotype threat. These pathways occur to the extent that individual minority group members internalize or externalize negative stereotypes about their group's intelligence. Externalization (EXT) occurs when minority students believe that members of the majority perceive them stereotypically as less intelligent and thus judge them invidiously in the performance of academic tasks. Internalization (INT) occurs when minority students themselves buy into the stereotype of intellectual inferiority at some level and come to fear that it may pertain to their own academic abilities.

The internalization and externalization of negative stereotypes yield two separate pathways to underachievement. Minority students follow the internalization pathway when they "disidentify" with academic performance as a determinant of self-esteem to relieve the psychological distress of potentially confirming the negative stereotype (Steele and Aronson 1995; Crocker, Major, and Steele 1998), typically achieved through a reduction of academic effort (EFF). The reduction of effort relieves the psychological pressure, because when minority students earn low grades they can attribute their failure to measure up to their lack of effort rather than any ostensible limitations of their intellectual abilities. They follow the externalization pathway when they become preoccupied with possibly confirming the negative stereotype rather than concentrating on the task at hand, yielding an academic performance burden (APB) that majority students do not face. We refer to the effects of internalization and externalization on academic performance as they operate through academic effort and performance burden as the "internalization mechanism" and "externalization mechanism," respectively. Although Massey and Fischer (2005) thought that externalization could affect disinvestment and internalization might affect the performance burden, this proved not to be the case empirically.

As noted above, most prior research on stereotype threat consists of laboratory experiments in which minority students are randomly assigned to a treatment group in which race is primed before administering an intellectual task, and a control group where race is not primed. These studies demonstrate the ease with which stereotype threat can be created and manipulated experimentally to push academic performance up or down (see Spencer et al. forthcoming for a review). Although laboratory experiments confirm the hypothesis of stereotype threat with a high degree of internal validity, such studies lack external validity in the sense that results cannot readily be generalized to real world settings such as a college campus or classroom (Campbell and Stanley 1963).

It is not much of a conceptual leap to speculate that minority students attending a range of selective, four-year, residential colleges might underperform academically owing to stereotype threat. Minority college students can expect to experience racial priming from a variety of sources in the course of daily life—from teachers, students, administrators, the media, and society in general. Moreover, because selective colleges and universities exhibit many characteristics of what Goffman (1962) calls total institutions, stereotypes are likely to carry particular weight on campus. With young students segregated in dorms, subject to a strong collective labeling, exposed to social sanctions from each other, and vulnerable to academic sanctions from faculty, even subtle evocations of minority status may trigger

heightened self-awareness to undermine academic performance (McGuire, McGuire, and Winton 1979).

In order to detect the potential influence of stereotype threat on campus, researchers began to add items to student surveys to measure constructs such as internalization, externalization, academic performance burden, and academic effort. Studies based on survey data generally provide empirical support for the view that the internalization and externalization of stereotypes independently influence academic underperformance (Charles et al. 2009; Fischer and Massey 2007; Massey and Fischer 2005; Massey and Mooney 2007). At the same time, unlike laboratory experiments, it is difficult to explicitly measure the extent of racial priming that may influence college performance.

As already noted, previous social survey-based studies of stereotype threat suffer from serious methodological and theoretical problems. First, they employ simple summed indices to measure internalization, externalization, and the academic performance burden, implicitly assuming that each item in the summed index carries equal weight in the underlying construct when, in fact, different items are likely have different measurement properties with different errors, especially across groups. A latent variables approach to model each stereotype threat concept, on the other hand, allows us to estimate factor loadings for the indicators of each latent variable—maximizing efficiency and accuracy when fitting measurement models for each latent construct and empirically examining each indicator's statistical significance and magnitude (i.e. the weights of each indicator) varies.

Second, previous studies using ordinary least square regressions rely on the theoretical assumption that the mechanisms of internalization and externalization are empirically distinct paths. Using multiple, distinct regressions to test paths between each construct not only prevents a formal, empirically-based test of whether the mechanisms of internalization and externalization are theoretically and statistically distinct, modeling constructs and equations separately also reduces statistical efficiency and introduces potential bias if in fact the internalization and externalization mechanisms are not distinct in practice. A method like structural equation modeling that relies on latent variables fit using measurement models, on the other hand, allows for the simultaneous modeling of relationships between all stereotype threat constructs, estimating error correlations between internalization and externalization and academic effort and performance burden to empirically test whether the mechanisms are distinct.

Third, previous studies rely on just a single variable—number of hours studied—to measure academic effort, assuming without justification that this single indicator fully captures the complex psychological reality of disidentification. The use of a single measure to capture academic effort is problematic for both substantive and theoretical reasons. Substantively, it is well-known that multiple measures yield more reliable scales than single indicators. Theoretically, we argue that disidentitifcation--the removal of academic achievement from the domain of self-evaluation and self-esteem—is likely to be expressed in both subjective and objective ways. Hence, in addition to hours studied, we add two subjective indicators of academic effort: self-assessed academic effort and the self-rated importance of learning course material.

Although we believe using multiple indicators to measure academic effort offers a more accurate strategy, we acknowledge the relative subjectivity of all indicators—not only of academic effort, but of each stereotype threat construct. We address this limitation by leveraging one of the advantages of latent variables analysis: the ability to test whether each indicator improves the fit of its respective latent construct. Throughout this study we make the fundamental assumption that students with higher levels of stereotype internalization

and/or externalization have been more "primed" to think about race in their daily lives than those who report lower levels. As a broader check to our model, we also fit our model for white students by way of comparison to show that stereotype threat from negative racial stereotypes is indeed unique to minority students.

DATA AND METHODS

Sample

The present study seeks to remedy the foregoing limitations to obtain more accurate estimates of effects associated with two the pathways specified in the Massey-Fischer (2005) model. As before, our data come from the National Longitudinal Survey of Freshmen (NLSF), a stratified random sample of 3,924 college students who entered 28 selective four-year colleges and universities throughout the United States in the fall of 1999 (see Massey et al. 2003; Charles et al. 2009). Students were interviewed in-person during the fall of their freshman year to collect retrospective information about social and educational experiences from childhood through high school. They were subsequently re-interviewed by phone every spring from 2000 through 2003 to learn about their social and academic experiences in college. Here we draw upon data gathered in the baseline and first two follow-up surveys, which had respective response rates of 86%, 96%, and 90%. We selected Hispanic and black students for study because, as stigmatized minorities, they are most susceptible to stereotype threat. For comparison purposes, however, we run the full structural model described below for white students, confirming that the stereotype threat model does not hold for these groups *not* historically stigmatized with negative stereotypes about their academic or intellectual abilities. Multiple imputation of five datasets was used to deal with item non-response (which ranged between .5% and 5%, as shown in Table 2), yielding a sample of 1,710 students, 918 black and 794 Hispanic. Comparisons of the stereotype threat model estimated for white students rely on a sample of 853 whites after multiple imputation. To make sure that results were not driven by imputation procedures, analyses were replicated on a sample of 1,579 students based on list-wise deletion of cases with missing data, for a final count of 839 black students and 740 Hispanic students. Results from analyses carried out using the imputed data or the list-wise deleted data yield the same substantive findings (results available upon request).

Variables

The dependent variable in our analyses is academic performance, which we measure as average GPA earned by students during the spring and fall of 2000-or the spring of their freshman year and the fall of their sophomore year. We derive estimates of the theoretical constructs shown in Figure 1—INT, EXT, EFF, and APB—using multiple indicators drawn from the survey, which are listed in Table 1. The items we drew upon were explicitly developed by the survey's authors in a conscious effort to capture different facets of stereotype threat. They are not a convenience sample of items pressed into the service of measurement ex post facto, but were intended from the start to get at the issue of stereotype threat.

We operationalize the internalization construct (INT) using three items that assess the degree to which respondents believe that members of their own group are lazy (v1 in Table 1), unintelligent (v2), and give up easily (v3). Externalization (EXT) is indexed by four items on whether respondents think whites (v4) and Asians (v5) discriminate and the degree to which they think other instructors (v6) and students (v7) base academic evaluations on group stereotypes.

Unlike Massey and Fischer, here we measure academic effort (EFF) using three items: the average number of hours studied per seven-day week during the academic year (v8), how

important the respondent believes it is to try hard to learn the course material in college classes (v9), and how much self-reported effort they estimate they have put into their studies over the past year (v10). Finally, academic performance burden (APB) is assessed by five items that ask respondents to report on the degree that instructors will think less of them if they have difficulty in class (v11), their individual performance reflects positively (v12) or negatively (v13) on their group, their own apprehensions about appearing foolish or stupid before others (v14), and the extent to which they believe not doing well academically will cause people to look down on others like them (v15).

In addition to measuring these theoretical constructs, all models include institution-level fixed effects. Our models also control for a variety of background variables that prior work using the NLSF has shown to be important in determining academic performance (see Charles et al. 2009; Massey et al. 2003), including gender, having a foreign-born parent, number of siblings in the household of origin, whether the student was raised in an intact, two-adult household, parental educational attainment, percent of college paid by family members, and the average number of hours worked for pay during the typical academic week in the first two years of college. We also control for differences in racial exposure and identity—the percentage of high school friends who were black or Hispanic, perceived social distance from whites, the strength of in-group identity, and skin tone. Differences in the degree of academic preparation are held constant by measuring the number of advanced placement courses taken in high school, cumulative high school GPA, and self-rated academic preparation.

We deploy these controls to help eliminate equally plausible alternative interpretations of the study's results. For example, it is possible that students clutch to negative stereotypes as a way to rationalize their own poor performance, but we try to eliminate this alternative by controlling for students' prior academic performance. Similarly, we try to control for the possibility that students attended poorer-performing schools and therefore arrive less-well prepared even if they received high grades by controlling for measures of social class—like neighborhood and school racial composition and percent of college paid for by family. Of course these are by no means perfect or exhaustive measures, but they allow us to help isolate the effects of internalization and externalization, via academic effort and academic performance burden, respectively, on college academic performance.

Methods

Figure 2 shows the path diagram used to model stereotype threat. Each circle represents a latent construct from the theoretical model shown in Figure 1, and arrows between these latent constructs represent regression paths to be estimated. Squares represent observed variables that are used to measure each construct and are identified by their corresponding variable number and explained above, as well as in Table 1.

We use multiple-group structural equation modeling (SEM) to investigate our research questions (see Bollen 1989). Broadly, SEM analysis allows us to take into account the measurement properties of the indicators underlying each of our key theoretical constructs— internalization, externalization, academic effort, and academic performance burden—and to assess whether they function similarly for black and Hispanic students. More specifically, multiple-group SEM simultaneously fits the model shown in Figure 2 (consisting of measurement models of internalization, externalization, etc., regression paths between each of these constructs, and error correlations between the internalization and externalization mechanisms) for each race group, leading to two different sets of estimates for the factor loadings on the indicators of each construct and for the regression paths between constructs. If the measurement properties of the indicators for each theoretical construct do not function similarly for blacks and Hispanics, and measurement differences are ignored, then apparent

differences between groups may not, in fact, reflect real differences. Instead, they may result from measurement differences between groups.

Prior work investigating stereotype threat using OLS regression modeling with summed multi-item indices assumes that the unique variance of each indicator is itself a robust measure of the construct for which it is a proxy, essentially meaning that OLS takes all measures of latent constructs to be equally valid. SEM allows us to relax that key assumption by measuring the latent constructs based on only the overlapping variation between indicators of each latent construct, as shown in Figure 3. Measuring each latent variable based only on the joint variation between all its indicators, each indicator's factor loading serves as a weight for the latent variable. As the unique variance of an indicator approaches zero, its standardized factor loading approaches one, which equals a perfect correlation between that indicator and the latent variable it measures. The result is that SEM controls for measurement error, leading to more accurate measures of the latent constructs if in fact factor loadings are all significant and vary in magnitude, indicating uneven contributions to their respective latent construct.

In addition, by allowing the factor loadings to vary between blacks and Hispanics, a latent variables approach provides additional flexibility by allowing the "contributions" of each indicator to the measurement of each latent construct to vary by race. For example, in the measurement of internalization, perceptions of one's own group's intelligence may receive a larger weight for blacks than for Hispanics, for whom perceptions of one's own group's hard work may contribute more heavily to the measurement of internalization.

Finally, fitting the full SEM shown in Figure 2 simultaneously for blacks and Hispanics (each with their own sets of measurement models, regression paths between constructs, and error correlations) and SEM's use of multiple dependent variables (i.e. internalization, externalization, etc.) allow us to examine error correlations between relevant latent variables (internalization with externalization and effort with academic performance burden) and to test whether the internalization and externalization mechanisms are truly distinct, which has been implicitly assumed in prior work. If the two mechanisms are not truly distinct, this paper and future work will need to investigate additional paths through which stereotype threat might operate.

Because structural relationships are conventional regression paths estimated using maximum likelihood methods, SEM makes many of the same assumptions as ordinary least squares regression. For example, a key assumption of SEM is that, taken together, indicators of a given latent variable and the latent variable itself have a joint multivariate normal distribution. However, violating this assumption does not seem to be consequential for the integrity of the structural model (see Bollen 1989). In addition, whereas OLS regression loses efficiency by making the assumption that error terms between variables are uncorrelated, SEM does not. Indeed, because SEM simultaneously models multiple structural relationships, it estimates correlations between the error terms of endogenous variables and thereby increases efficiency in the estimates themselves (see Bollen 1989).

To test whether a latent variables approach offers a more accurate and efficient modeling strategy, we carried out two preliminary steps. First, we tested whether each indicator improved the fit of each latent stereotype threat construct, which we accomplished by adding each indicator to the measurement model one at a time. We found that the fit improved with each additional indicator (results available upon request) and that, together, all indicators of a given construct more accurately capture the concept being modeled. Second, we also tested whether it is sufficient to assign equal weight to each indicator in a given latent construct by examining the magnitude and significance of individual factor

loadings. Here we found that magnitudes vary across indicators, highlighting the need for latent variables as opposed to summed indices (factor loadings shown in Appendix B).

We begin by using separate measurement models to estimate internalization, externalization, academic effort, and academic performance burden for each group. After validating our measurement models—i.e. establishing their reliability and validity for the aggregate group of blacks and Hispanics—we compare our results to measurement models corresponding to summed indices that have been used in the past. Once the measurement components of the analyses are complete, we estimate the full multiple-group SEM model shown in Figure 2 to investigate directly whether the model holds for both blacks and Hispanics, and identify whether there are intergroup differences in the structural effects of stereotype threat. The SEM is based on the following equations, modeled simultaneously:

$$\text{PERF}^*_i = \beta_{0i} + \beta_{1i}(\text{INT}^*) + \beta_{2i}(\text{EXT}^*) + \beta_{3i}(\text{EFF}^*) + \beta_{4i}(\text{APB}^*) + \beta_{5i}(\text{male}) + \ldots + \beta 17_i(\text{self−rated prep}) + e_1$$
$$\text{EFF}^*_i = \gamma_{0i} + \gamma_{1i}(\text{INT}^*) + \gamma_{2i}(\text{male}) + \ldots + \gamma 14_i(\text{self−rated prep}) + e_2$$
$$\text{APB}^*_i = \alpha_{0i} + \alpha_{1i}(\text{EXT}^*) + \alpha_{2i}(\text{male}) + \ldots + \alpha 14_i(\text{self−rated prep}) + e_3$$
$$\text{INT}^*_i = \delta_{0i} + \delta_{1i}(\text{male}) + \ldots + \delta_{13i}(\text{self−rated prep}) + e_4$$
$$\text{EXT}^*_i = \xi_{0i} + \xi_{1i}(\text{male}) + \ldots + \xi_{13i}(\text{self−rated prep}) + e_5$$

where i represents a separate estimation equation for blacks and Hispanics, respectively, and * represents latent variables/constructs.

After determining that intergroup differences do exist, we estimate final models for both groups. In the full SEM models, we control a number of background characteristics, which are shown in Table 2. Finally, we carry out a number of sensitivity analyses to examine the robustness of the stereotype threat model given the subjective measures we use to operationalize each latent construct. Although the variables we used were designed specifically to measure each stereotype threat concept, we first verify the reliability of each latent construct even if different indicators are used to measure each construct. Our goal is to ensure that each construct is not heavily reliant on a single indicator to capture the complex underlying concept. Fitting the full SEM fifteen times, each time excluding one of the fifteen indicators of stereotype threat described in the variables section, an examination of the magnitudes and significances of the paths indicate robustness against slight changes in latent variable specification (shown in Appendix C).

These results indicate that no single indicator is driving the relationship between each latent measure of stereotype threat (results available upon request). Second, we verify the value-added of each indicator by examining the magnitudes and significances of the factor loadings of each indicator on its respective latent construct, as shown in Appendix B. The significance of each indicator suggests that each additional item used adds important variation to the measurement of each latent variable, thus more fully capturing each underlying latent concept. The differential magnitude of each factor loading confirms that differential weights on indicators are indeed necessary in order to most accurately model each latent construct of stereotype threat.

MEASURING THEORETICAL CONTRUCTS

Table 2 shows means of indicators for each theoretical construct and the corresponding summed indices (assuming no measurement differences between groups) for blacks and Hispanics with whites shown for comparative purposes. Although there are some differences in the means of summed indices and their constituent items between blacks and Hispanics (most noticeably with Hispanics having lower mean levels of externalization and performance burden), the largest overall differences are between whites on the one hand and

blacks and Hispanics on the other. Whites have significantly lower levels of externalization and academic performance burden than either blacks or Hispanics. In terms of the summed indicator for internalization and academic effort, however, the mean values for Hispanics, blacks, and whites are nearly identical.

Even though the means of the summed scales for internalization and academic effort are similar for blacks, Hispanics, and whites, it is nonetheless possible that the constituent items have different weights in the latent constructs. Table 3 shows results comparing each of the four theoretical constructs estimated by group using two measurement models: a full model that estimates errors and factor loadings empirically (a confirmatory factor analysis) and a restricted model that constrains factor loadings to be 1 and assumes measurement error to be 0 (the constrained factor analysis, which is comparable to a summed indices approach). Table 3 reports three common goodness-of-fit measures for each of the models: the model chi-square, the Root Mean Square Error of Approximation (RMSEA), and the Comparative Fit Index, or CFI (Bollen 1989). Appendix B further demonstrates the improved accuracy of using latent variables versus summed indices by documenting the significance and varied magnitudes of each of the indicators of the given stereotype threat constructs.

In the full measurement model for internalization, the chi-square was 35.52 and was highly statistically significant (4 d.f., p<0.001) with RMSEA and CFI of 0.09 and 0.91, respectively, confirming the sufficient fit of this measurement model to the data. In contrast, the measurement model equivalent to using summed indexes (the constrained analysis) fits the data much less well, with a model chi-square of 46.32 (9 d.f., p<0.001), an RMSEA of 0.15, and a CFI of 0.90. A difference chi-square test comparing the two models yields a value of 14.96 (5 d.f., p<0.001), indicating a significantly poorer fit for the constrained than the full model, demonstrating the inferiority of the summed index approach used in prior survey-based studies.

Similar findings emerge with respect to the other theoretical constructs. For externalization, the full measurement model yields a chi square of 48.45 on 8 d.f. (p<0.001). While this chi square suggests a poorer fit than for the internalization construct, the RMSEA was nonetheless 0.08 and the CFI was 0.95, indicating that the fit, while not perfect was reasonably good. In contrast, the chi-square for the summed index model was 316.03 on 15 d.f., suggesting a very poor fit indeed, and the RMSEA and CFI were 0.15 and 0.63, respectively, confirming this diagnosis. The chi-square difference test was highly significant (chi square=247.87,7d.f., p<0.001).

Similarly, in measuring academic effort, the full measurement model yields a chi square of 14.70 on 7 d.f. (p>0.05) with RMSEA and CFI of 0.03 and 0.97, respectively, indicating another excellent fit, whereas the corresponding summed index for academic effort has a chi square of 238.59 (17 d.f., p<0.001), RMSEA of 0.12 and CFI of 0.20, indicating poor fit. The difference in chi square for the measurement model versus the summed index for academic effort is 214.20 on 10 d.f., indicating the full measurement model offers significantly better fit. Finally, in estimating the academic performance burden, the full measurement model yields a chi square of 100.32 on 16 d.f. (p<0.001). Although this fit is again somewhat poorer than for internalization or reduction of academic effort, the RMSEA of 0.08 and the CFI of 0.94 nonetheless indicate a reasonably good fit nonetheless. Once again, the summed index indicated a very poor fit on each goodness-of-fit statistic, with a model chi-square of 407.75 on 25 degrees of freedom, RMSEA of 0.13 and CFI of 0.73; and as before the chi square difference test indicated a significantly poorer fit compared with the full measurement model (difference chi square=305.84 with 9 d.f., p<0.001).

All in all, these results suggest that using a measurement model that allows for differences in measurement properties across groups is warranted when assessing the relationship between theoretical constructs pertaining to the stereotype threat model. Our findings imply that if measurement differences are ignored, estimates of the structural relations between constructs may be biased because of measurement differences. Thus we employ unconstrained latent variables in our SEM analysis of stereotype threat, yielding estimated effects that should be free of this bias.

PATHWAYS TO UNDERPERFORMANCE

Turning to the structural model summarized in Figure 2, our first step was to determine whether stereotype threat functions similarly for Hispanics and blacks. In Table 4 we show results for two different measurement strategies, one estimating the structural model simultaneously for blacks and Hispanics and the other estimating separate models. The model chi-square was 981.99 and statistically significant (576 d.f., p<0.001) when blacks and Hispanics were estimated separately, and the CFI and RMSEA were 0.91 and 0.03, respectively, indicating an acceptable fit. To test whether black and Hispanic students experience stereotype threat in the same way, we estimated a single SEM for the two groups collapsed together by constraining the measurement and structural models to be equivalent across groups. This model produced a chi square of 1104.67 (590 d.f., p<0.001), a CFI of 0.88 and an RMSEA of 0.04, indicating significant measurement and structural differences between blacks and Hispanics. The difference in chi square test was 122.68 (14 d.f.m p<0.001), indicating significant differences between black and Hispanic students in how stereotype threat affects academic performance.

Table 5 presents estimates the three different models estimating the effects of internalization and externalization on academic performance through academic effort and the academic performance burden, thus fully operationalizing the diagram shown in Figure 1. Given the nature of social survey studies, these results do not establish causal mechanisms, of course, but causality has already gained strong support in Steele's original experimental work (Steele 1988a, 1988b). Here we seek only to detect evidence of stereotype threat in the real world. Nonetheless, by deploying a variety of methods that come to similar conclusions our findings withstand various tests of robustness.

In order to examine whether the stereotype threat model holds for non-minority students, model 1 fits latent variables for stereotype threat for whites, blacks and Hispanics, but excludes the racial in-group exposure variables. Racial in-group exposure variables are only measured for blacks and Hispanics and are therefore excluded so that whites can be included in the model. Model 2 of Table 5, together with Figure 4, show the results of a final latent variables model for only blacks and Hispanics, including the racial in-group exposure variables, in which parameters were allowed to vary by racial group. Regression coefficients capturing the effects of each control variable on latent theoretical constructs of model 2 are shown in Appendix A. No systematic differences in the influence of controls were found between Hispanics and blacks. Factor loadings for each of the indicators of the latent variables are shown in Appendix B. The significance and varied magnitudes of the indicators confirm the value-added of each indicator in measuring the given latent variable. Finally, model 3 of Table 5 presents results from a path analysis model of stereotype threat, in which summed indices of respective indicators are used in place of latent variables when estimating relationships between the constructs postulated by stereotype threat. Though magnitudes and significances between paths are generally attenuated and weaker, results from the path analysis indicate similar patterns of relationships between constructs.

As noted above, the full structural model shown in model 2 of Table 5 and in Figure 4 fits the data quite well. The externalization of negative stereotypes is associated with a significant increase in academic performance burden for both black and Hispanic students. A black or Hispanic student whose level of externalization is one standard deviation above that of another student's of his/her same racial group is, on average, expected to experience a 0.08 and 0.03 standard deviation higher level of performance burden, respectively, net of controls—consistent with expectations derived from stereotype threat theory. Moreover, as hypothesized by the model, a black or Hispanic student whose level of performance burden is one standard deviation above that of another student of his/her same race is expected to perform 0.06 and 0.03 standard deviations lower, respectively, net of controls. The direct effect of externalization on academic performance is statistically zero for both groups, meaning that externalization affects academic performance through the posited mechanism of academic performance burden.

As expected, the effect of internalization on academic effort is negative for both groups. Among black students, one standard deviation more internalization on average yields a 0.18 standard deviation decline in academic effort. Among Hispanic students, the association is slightly larger—a Hispanic student who is one standard deviation higher in his/her level of internalization is expected to put forth 0.21 standard deviations less academic effort on average, net of controls, compared to a student whose level of internalization is one standard deviation lower. Consistent with theoretical predictions, the effect of academic effort on performance is positive and significant for both groups, so that the indirect effect of internalization through academic effort is negative. Internalizing negative stereotypes brings about disidentification and a reduction of academic effort which, in turn, yields lower grades. The effect, however, is greater for Hispanics than blacks. Whereas a black student who puts forth one standard deviation more effort receives a grade-point that is on average 0.17 standard deviations higher than another student with the same background characteristics, the same relationship is associated with a grade-point that is 0.36 standard deviations higher for a Hispanic student with the same background characteristics as another Hispanic student but for putting forth one standard deviation worth more academic effort.

In addition, the direct effect of internalization on academic performance is positive and significant for both black and Hispanic students, implying that the internalization mechanism accounts for a significant but nonetheless incomplete portion of the total effect of internalization on academic performance. A black student who experiences one standard deviation more internalization than another black student with the same background characteristics on average earns a grade-point that is 0.11 standard deviations higher. On the other hand, a Hispanic student who experiences one standard deviation more internalization than another Hispanic student with the same background characteristics earns a grade-point that is an even larger and more highly significant 0.17 standard deviations higher on average. The fact that internalization has a *positive* direct effect on performance runs counter to the theory of stereotype threat and will be explored more in the discussion below. Here we simply note that the magnitudes of the direct paths between internalization and performance are of notably smaller magnitude for both blacks and Hispanics than are the paths posited through the internalization mechanism via academic effort. Finally, error correlations between internalization and externalization on the one hand and academic effort and performance burden on the other are non-significant and of small magnitude, confirming that the dual-pronged model of stereotype threat via the internalization and externalization mechanisms operate as two distinct pathways to underperformance among minorities.

Unlike for blacks and Hispanics, when the stereotype threat model is fit for whites, results from model 1 of Table 5 show significant structural paths only for the internalization mechanism. Among whites, a student who perceives one standard deviation more negative

stereotypes about his/her group's intelligence, hard work, and persistence puts forth, on average and net of controls, 0.14 standard deviations less academic effort than an otherwise comparable white student who does not perceive these stereotypes. The magnitude of the effect, however, is about 75% as large as that for blacks and Hispanics, on average. As logically expected, however, a one standard deviation increase in academic effort is associated with an increase in performance of roughly similar magnitude and significance as for Hispanics, on average, net of controls. The lower magnitude of the effect of effort on performance for blacks, however, is salient—a one standard deviation increase in academic effort is associated with an increase in performance that is only about half as large as the average of that for whites and Hispanics (0.15 for blacks compared to roughly 0.30 for whites and Hispanics, on average).

As with blacks and Hispanics, the magnitude and significance of the direct association between internalization and performance is positive, counter to the theory of stereotype threat, again suggesting that the internalization of negative stereotypes influences grade-performance through channels other than a reduction in academic effort. Some of this positive, direct effect may be due to minorities wanting to show that they are "not like others" of the same, negatively stereotyped racial group. The externalization mechanism, however, is entirely non-significant for whites. Furthermore, the internalization and externalization mechanisms do not appear to be theoretically distinct concepts. As discussed in more detail in the discussion, model 1 results for whites indicate whites' general lack of fit within the stereotype threat framework, particularly with regard to the externalization of negative stereotypes.

CONCLUSION: STEREOTYPE THREAT IN REAL WORLD SETTINGS

Stereotype threat theory has historically been tested in laboratory experiments using domestic college students engaging mundane academic tasks within a short period of time. More recently studies have examined the degree to which stereotype threat functions in natural social settings using survey data spanning multiple years of college. Although some of these studies have generally confirmed the operation of stereotype threat on campuses of selective colleges and universities (Charles et al. 2009; Massey and Fischer 2005; Massey and Mooney 2007), they have relied on relatively weak modeling strategies that do not exploit the full variation in the measures of stereotype threat and thus limit the accuracy of the model and more complete testing of its validity. This study was the first to improve measurement by deploying a latent variable structural equation model with more robust and realistic measurement assumptions.

Because the causal effects of negative stereotypes have been generally supported through randomized experiments, this study aims to accurately model and test whether the causal mechanisms demonstrated in laboratory experiments can be detected in the real world. To this end, this analysis lends further credibility to survey-based research on stereotype threat and lends further credibility to the notion that the internalization and externalization of negative stereotypes undermines the academic performance of minority students in theoretically expected ways, even after measurement error is effectively modeled and controlled.

This growing body of survey-based research on stereotype threat should be viewed as a complement to existing experimental research. Although experimental research is of utmost importance in establishing the causal relationships between race primes, negative racial stereotypes, and academic performance, its reliance on relatively small numbers of students brings with it a loss of generalizability outside labs and across universities. Survey-based studies offer a valuable complement in their ability to gain generalizability across

institutions once the validity of experimental results is highly supported. Instead of using direct race primes, this study uses as its starting point estimates of how much negative stereotypes have been internalized and externalized, and then assesses the influence of these psychological states on academic outcomes after controlling for relevant background factors possibly associated both with grades and internalization and externalization (for example, factors like growing up in neighborhoods with high proportions of black residents or having the same race friends as oneself).

Specifically, we show that the externalization of negative stereotypes—expecting to be judged invidiously by majority group members on the basis of a stereotypical belief in minority intellectual inferiority—increases the performance burden experienced by individual minority group members and that this extra psychological burn, in turn, lowers grade performance. Moreover, the lack of a direct effect between externalization and academic performance suggests that the entire effect operates through the hypothesized indirect pathway. Although statistically significant using a one-tailed test, however, the externalization pathway is not particularly strong.

For blacks and Hispanics, the hypothesized pathway through internalization appears to be much stronger in certain respects. To the extent that minority group members internalize negative stereotypes—believing at some level that the canard of intellectual inferiority might actually apply to them—they reduce their academic efforts in keeping with a psychological process of disidentification, which involves disengaging from grade achievement as a domain of self-evaluation. However, the strong, direct relationship between internalization and academic performance suggests other mechanisms are at work beyond those specified by stereotype threat theory.

That internalization is directly associated with higher grades calls into question the validity of oppositional culture theory (Fordham and Ogbu 1986). In other words, academic disengagement does not seem to be the only response to the internalization of negative stereotypes among minority students. In fact, a recent study suggests that if the reframing of threats as opportunities to overcome challenges is taking place, internalization may actually not effect, or even improve, performance for threatened groups (Alter et al. 2010). What this strong, direct relationship between internalization and academic performance really means should be the focus of additional work.

In addition, our findings with regard to blacks and Hispanics suggest that aggregation across minority groups misses significant variation along the lines of ethnicity and national origin (Portes and Rumbaut 2001), in this case between African Americans and Hispanics. Unfortunately, owing to sample size limitations, we were unable to disaggregate the black and Hispanic students by ethnic origin. There is reason to believe, however, that ethnic origin may emerge as a primary line along which students experience differences in the effects of stereotype threat. African Americans of immigrant origin students may experience stereotype threat differently than those of native origin, for example, and Mexicans may be more or less susceptible to stereotype threat than Cubans. Future research should endeavor to consider differences by national origin as well as broader racial-ethnic categories.

For whites, the coherence of the internalization mechanism is perhaps best understood in the context of research related to the effects of stereotype threat on athletes and women in mathematics and science—the majority of whom are white (Harrison et al. 2009; Huguet and Regner 2009; Yopyk and Prentice 2005). To the extent whites who identify as athletes or "women in the sciences and mathematics" fields (a la former Harvard President Larry Summers' comments in 2005), the broad items used to measure internalization in this study would allow white respondents to identify "his/her group" as that of athletes or women in

particular male-dominated fields. If these sub-groups of white students experience depressed performance on the basis of negative stereotypes about athletes, women in mathematics, sciences, engineering or other male-dominated fields, or otherwise, negative stereotypes about these groups' academic intelligence may account in large part for the apparent relationships between internalization, academic effort, and performance we find.

The lack of the influence of the externalization mechanism, on the other hand, may be explained if whites in the particular sub-groups susceptible to threat do not perceive negative stereotypes to be widely accepted by members outside the university community in general, or that they are not readily identifiable as members of the subgroups in question. In either case, it would follow logically that whites in these sub-groups would not experience a strong enough sense of externalization to drive an overall effect of stereotype threat.

Overall, the primary findings of the current research pertain to minority students attending selective colleges and universities. Future survey-based work might seek to expand the analysis to minority students at non-selective schools, including both two-year and four-year institutions, who comprise the bulk of minority students in higher education. Doing so will reveal whether stereotype threat is exacerbated or diminished by institutional selectivity and enable generalization to a population of students that more closely represent minority students attending college and university in the United States today. It may also help shed further light on the potentially contradictory findings between this study and Stricker and Ward's (2004) field experiment. Even though both measure the effects of stereotype threat —such as through grades or AP scores—in ways that reflect students' sustained academic achievement over time, they come to conclusions that are not immediately reconcilable.

Acknowledgments

References

Alter, Adam L.; Aronson, Joshua; Darley, John M.; Rodriguez, Cordardo; Ruble, Diane N. Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. Journal of Experimental Social Psychology. 2010; 46(1):166–171.

Aronson, Joshua; Fried, Carrie B.; Good, Catherine. Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence. Journal of Experimental Social Psychology. 2002; 38:113–125.

Bollen, Kenneth A. Applied probability and statistics section, wiley series in probability and mathematical statistics. 1. New York: Wiley; 1989. Structural equations with latent variables.

Bowen, William G.; Bok, Derek C. The shape of the river: Long-term consequences of considering race in college and university admissions. Princeton, N.J: Princeton University Press; 1998.

Campbell, DT.; Stanley, JC. Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally; 1963.

Charles, Camille Z.; Fischer, Mary J.; Mooney, Margarita; Massey, Douglas S. Taming the river: Negotiating the academic, financial, and social currents in selective colleges and universities. 1. Princeton, NJ: Princeton University Press; 2009.

Cohen, Geoffrey; Garcia, Julio; Apfel, Nancy; Master, Allison. Reducing the Racial Achievement Gap: A Social-Psychological Intervention. Science. 2006; 313(5791):1307–1310. [PubMed: 16946074]

Cohen, Geoffrey; Garcia, Julio; Purdie-Vaughns, Valerie; Apfel, Nancy; Brzustoski, Patricia. Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap. Science. 2009; 324:400–403. [PubMed: 19372432]

Crocker, Jennifer; Major, Brenda; Steele, Claude M. Social stigma. In: Gilbert, Daniel T.; Fiske, Susan T.; Gardner, Lindzey, editors. Handbook of social psychology. 4. Vol. 2. Boston, MA: McGraw-Hill; 1998. p. 504-553.

Cullen, Michael J.; Hardison, Chaitra M.; Sackett, Paul R. Using SAT-Grade and Ability-Job Performance Relationships to Test Predictions Derived from Stereotype Threat Theory. Journal of Applied Psychology. 2004; 89(2):220–30. [PubMed: 15065971]

Cullen, Michael J.; Waters, Shonna D.; Sackett, Paul R. Testing Stereotype Threat Theory Predictions for Math-Identified and Non-Math-Identified Students by Gender. Human Performance. 2006; 19(4):421–40.

Danaher, Kelly; Crandall, Christian S. Stereotype Threat in Applied Settings Re-Examined. Journal of Applied Social Psychology. 2008; 38(6):1639–1655.

Downey, Douglas B. Black/White differences in school performance: The oppositional culture explanation. Annual Review of Sociology. 2008; 34(1):107–26.

Espenshade, Thomas J.; Walton-Radford, Alexandria. No longer separate, not yet equal: Race and class in elite college admission and campus life. Princeton, NJ: Princeton University Press; 2009.

Fischer, Mary J.; Massey, Douglas S. The effects of affirmative action in higher education. Social Science Research. 2007; 36(2):531–49.

Fordham, Signithia; Ogbu, John U. Black students' school success: Coping with the burden of 'acting white'. The Urban Review. 1986; 18:176–195.

Goffman, Erving. Asylums: Essays on the social situation of mental patients and other inmates. Chicago, IL: Aldine; 1962.

Harrison, C Keith; Stone, Jeff; Shapiro, Jenessa; Yee, Sharon; Boyd, Jean A.; Rullan, Vashti. The Role of Gender Identities and Stereotype Salience With the Academic Performance of Male and Female College Athletes. Journal of Sport and Social Issue. 2009; 33(1):78–96.

Huguet, Pascal; Regner, Isabelle. Counter-Stereotypic Beliefs in Math do Not Protect School Girls from Stereotype Threat. Journal of experimental social psychology. 2009; 45(4):1024–1027.

Major, Brenda; O'Brien, Laure T. Social psychology of stigma. Annual Review of Psychology. 2005; 56(1):393–421.

Massey, Douglas S.; Charles, Camille Z.; Lundy, Garvey F.; Fischer, Mary J. The source of the river: The social origins of freshmen at America's selective colleges and universities. Princeton, NJ: Princeton University Press; 2003.

Massey, Douglas S.; Fischer, Mary J. Stereotype threat and academic performance: New findings from a racially diverse sample of college freshmen. Du Bois Review. 2005; 2(1):45–68.

Massey, Douglas S.; Mooney, Margarita. The effects of America's three affirmative action programs on academic performance. Social Problems. 2007; 54(1):99–117.

McGuire WJ, McGuire CV, Winton W. Effects of household sex composition on the salience of one's own gender in the spontaneous self-concept. Journal of Experimental Psychology. 1979; 15:77–90.

Pope, Devin G.; Sydnor, Justin R. Geographic Variation in the Gender Differences in Test Scores. Journal of Economic Perspectives. 2010; 24(2):95–108.

Portes, Alejandro; Rumbaut, Rubén G. Legacies: The story of the immigrant second generation. Berkeley, CA: University of California Press; 2001.

Spencer, Steven; Logel, Christine; Davies, Paul G. Stereotype threat. Annual Review of Psychology. 61(1) Forthcoming.

Steele, Claude M. The psychology of self-affirmation: Sustaining the integrity of the self. Advances in Experimental Social Psychology. 1988a; 21:261–302.

Steele, Claude M. A threat in the air: How stereotypes shape intellectual identity and performance. In: Eberhardt, Jennifer L.; Fiske, Susan T., editors. Confronting racism: The problem and the response. Thousand Oaks, CA: Sage Publications; 1988b. p. 234

Steele, Claude M. A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance. American Psychologist. 1997; 52:613–629. [PubMed: 9174398]

Steele, Claude M.; Aronson, Joshua. Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology. 1995; 69(5):797–811. [PubMed: 7473032]

Stricker, Lawrence J.; Ward, William C. Stereotype Threat, Inquiring about Test Takers' Ethnicity and Gender, and Standardized Test Performance. Journal of Applied Social Psychology. 2004; 34(4): 665–93.

Walton, Gregory M.; Cohen, Geoffrey L. A Question of Belonging: Race, Social Fit, and Achievement. Journal of Personality and Social Psychology. 2007; 92:82–96. [PubMed: 17201544]

Walton, Gregory M.; Spencer, Steven J. Latent Ability: Grades and Test Scores Systematically Underestimate the Intellectual Ability of Negatively Stereotyped Students. Psychological Science. 2009; 20(9):1132–1139. [PubMed: 19656335]

Yopyk, Darren JA.; Prentice, Deborah A. Am I an Athlete or a Student? Identity Salience and Stereotype Threat in Student-Athletes. Basic and Applied Social Psychology. 2005; 27(4):329–336.

Appendix A. Regression Coefficients for Effects of Controls on Structural Equation Model of Relationships Between Internalization, Externalization, Disidentification, Academic Performance Burden and Academic Performance

		Standardized Coefficients									
	Race (B=black, H=Hispanic)	Internalization		Externalization		Effort		Academic Perf. Burden		Performance	
Variables[1]		B	S.E.	B	S.E.	B	S.E.	B	S.E.	B	S.E.
Male	B	0.12**	0.04	−0.09*	0.04	−0.10*	0.05	0.09*	0.04	−0.09**	0.03
	H	0.12**	0.04	−0.16***	0.04	−0.10*	0.05	−0.07	0.04	−0.10**	0.04
# of Dependents	B	−0.02	0.04	0.02	0.04	0.06	0.05	−0.02	0.04	−0.04	0.03
	H	0.08	0.04	0.04	0.04	0.04	0.05	0.02	0.039	−0.03	0.04
Intact Family	B	0.08	0.05	0.01	0.04	0.16**	0.05	0.07	0.04	0.03	0.03
	H	−0.06	0.04	0.01	0.05	0.00	0.05	−0.02	0.04	0.02	0.04
First Generation Immigrant	B	−0.01	0.04	−0.07	0.04	0.06	0.05	0.08	0.04	−0.02	0.03
	H	−0.04	0.05	0.14**	0.05	0.08	0.06	0.03	0.047	−0.01	0.04
Second Generation Immigrant	B	−0.01	0.04	0.11*	0.04	0.08	0.05	0.07	0.04	0.00	0.03
	H	−0.04	0.05	0.06	0.05	−0.01	0.05	0.01	0.045	−0.02	0.04
1 parent B.A.(Ref=No parents B.A.)	B	−0.03	0.05	−0.10*	0.04	0.04	0.05	0.02	0.04	−0.01	0.04
	H	0.03	0.05	0.11*	0.05	0.02	0.05	−0.03	0.045	0.06	0.04
2 parents B.A.	B	−0.08	0.05	−0.06	0.04	−0.01	0.05	−0.04	0.04	0.06	0.03
	H	0.01	0.05	0.07	0.05	−0.02	0.05	−0.04	0.045	0.10*	0.04
1 parent Advanced Degree	B	−0.01	0.05	0.03	0.05	−0.09	0.06	0.03	0.05	0.04	0.04
	H	−0.03	0.05	0.00	0.06	−0.02	0.06	−0.02	0.05	0.12**	0.04
2 parents Advanced Degree	B	−0.01	0.05	−0.04	0.05	−0.03	0.06	0.02	0.05	0.05	0.04
	H	−0.03	0.05	−0.05	0.06	−0.03	0.06	−0.07	0.049	0.17***	0.04
% of college paid by family (%/10)	B	0.06	0.05	−0.02	0.04	−0.01	0.05	0.05	0.04	0.01	0.03
	H	0.12*	0.05	0.10	0.05	−0.03	0.06	−0.12*	0.047	−0.024	0.04
Hours Work for Pay (/10)	B	−0.03	0.04	−0.01	0.04	0.02	0.05	0.02	0.04	0.09**	0.03
	H	−0.14**	0.04	−0.07	0.05	−0.13**	0.05	−0.01	0.04	0.02	0.04
Strength of ingroup racial identity (0–10)	B	−0.13**	0.04	0.28***	0.04	−0.03	0.05	0.14**	0.04	0.05	0.04

Standardized Coefficients										
Variables[1]	Race (B=black, H=Hispanic)	Internalization		Externalization		Effort		Academic Perf. Burden		Performance	
		B	S.E.	B	S.E.	B	S.E.	B	S.E.	B	S.E.
	H	0.02	0.04	0.15[**]	0.05	0.03[**]	0.05	0.12[**]	0.04	0.00	0.04
Social distance from whites (0–10)	B	0.05	0.04	0.17[***]	0.04	−0.10[*]	0.05	0.07	0.04	0.05	0.03
	H	0.20[***]	0.05	0.17[***]	0.05	−0.02	0.05	0.07	0.042	0.03	0.04
% black or Hispanic neighborhood	B	0.02	0.05	0.02	0.05	0.04	0.06	0.04	0.05	−0.11[**]	0.04
	H	−0.02	0.06	−0.07	0.07	0.23	0.07	0.08	0.059	−0.22[***]	0.05
% same-race friends growing up (%/10)	B	0.01	0.05	0.10	0.05	0.11	0.06	0.03	0.05	0.02	0.04
	H	−0.08	0.06	0.01	0.06	−0.04	0.07	0.08	0.056	0.10[*]	0.05
# AP Courses	B	0.06	0.05	−0.01	0.04	0.01	0.05	−0.00	0.04	0.14[***]	0.03
	H	0.11[*]	0.05	0.10[*]	0.05	0.06	0.05	0.02	0.044	0.07	0.04
Self-rated preparation (0–10)	B	0.02	0.04	0.09[*]	0.04	−0.09	0.05	−0.04	0.05	0.03	0.03
	H	0.00	0.05	−0.02	0.05	−0.19[***]	0.05	0.02	0.042	0.03	0.04
High School GPA (0–4)	B	0.08[*]	0.03	0.01	0.03	0.25[***]	0.05	0.03	0.04	0.22[***]	0.04
	H	0.08	0.05	0.01	0.05	0.20[***]	0.05	−0.05	0.044	0.14[**]	0.04

[*] Significant at .05,

[**] Significant at .01,

[***] Significant at .001. One-tailed test.

NOTE: Reporting Standardized coefficients. Model includes university-level fixed effects. Effects of regression paths between constructs in full SEM shown in Table 5, model 2.

[1] See Table 2 for fuller description of control variables.

Appendix B. Factor Loadings for Each Indicator of the Constructs of Stereotype Threat, by Race

	Standardized Coefficients			
	Black		Hispanic	
Constructs of Stereotype Threat	B[1]	SE	B[1]	SE
Internalization (INT)				
Own group's intelligence (0–6)	0.49	0.04	0.65	0.03
Own group is hard working (0–6)	0.70	0.04	0.64	0.03
Own group perseveres (0–6)	0.51	0.03	0.64	0.03
Externalization (EXT)				
Whites treat other races equally or discriminate (0–10)	0.75	0.04	0.71	0.05
Asians treat other races equally or discriminate (0–10)	0.58	0.03	0.63	0.05
Instructors' stereotypes do not affect evaluations of members of stereotyped groups (0–10)[2]	−0.13	0.03	0.06	0.04
Students' stereotypes do not affect evaluations of members of stereotyped groups (0–10)[2]	−0.15	0.03	0.00	0.04
Academic Effort (EFF)				

	Standardized Coefficients			
	Black		Hispanic	
Average Number of Hours Studied in a 7 day Week/10 (0–12)	0.40	0.04	0.48	0.04
Importance of learning course material (0–10)	0.40	0.03	0.46	0.03
Self-reported [academic] effort during past year of college (0–10)	0.65	0.04	0.66	0.04
Academic Performance Burden (APB)				
Instructors think less of me for having difficulty in class (0–10)	0.38	0.04	0.24	0.04
Excelling academically reflects positively on my racial/ethnic group (0–10)	0.42	0.05	0.72	0.02
Doing poorly academically reflects negatively on my racial/ethnic group (0–10)	0.51	0.05	0.84	0.02
I don't want to look foolish or stupid in class (0–10)	0.49	0.04	0.35	0.04
If I don't do well, people will look down on others like me (0–10)	0.75	0.04	0.58	0.03
Academic Performance (GPA)				
Average Second and Third Semester GPA	1.00	0.00	1.00	0.00

NOTE: Results from the full structural equation model are shown in Table 5, Model 3 and are based on multiple imputation of 5 datasets.

[1] Each indicator is statistically significant at .001, indicating a significant contribution to the variance of its respective latent construct.

[2] This item was reverse-coded to accurately reflect its negativelyl valenced properties in comparison to the other items in the externalization construct.

Appendix C. Fifteen Structural Equation Model (SEM) Results for Black and Hispanic Students, Each Model Omitting One Indicator of Stereotype Threat to Test the Robustness of Structural Paths to Changes in Model Specification (N=1,710)

*
Significant at .05,

Significant at .01,

***
Significant at .001; one tailed test.

NOTE: Reporting Standardized coefficients from sample resulting from multiple imputation of five datasets. Standard errors in parentheses. Models include institution-level fixed effects.

[1] Structural residual error correlations are estimated between internalization and externalization and between academic performance burden and disidentification. Measurement error correlations are estimated between indicators v6 and v7 and v10 and v11 (See Table 1 for list of variables by number).

[2] See Table 1 for a list of variables by number.
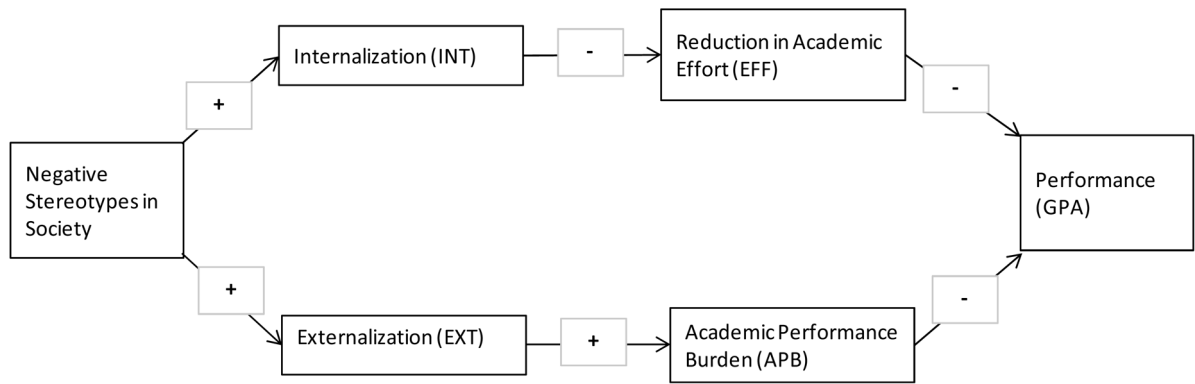
Figure 1. Conceptual Model of Stereotype Threat with Expected Direction of Relationships between Concepts
NOTE: The boxes overlaying arrows contain '+' or'−' signs that indicate the expected direction of the effect based on the theory of stereotype threat.'+' represents an expected positive relationship between the two concepts, whereas'−' represents an expected negative relationship.
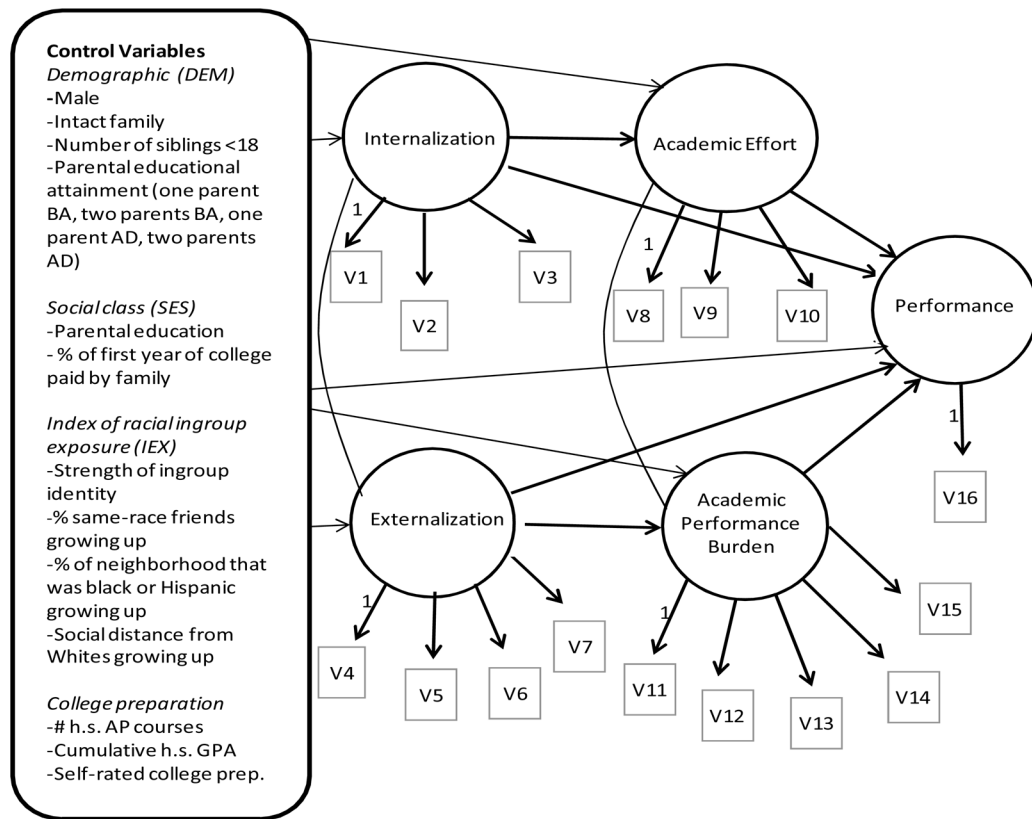
Figure 2.
Path Diagram for Stereotype Threat[1]
[1]See Table 1 for specification of measures v1–v16. Arrows signify regression paths. Lines without arrows signify estimated error correlations.
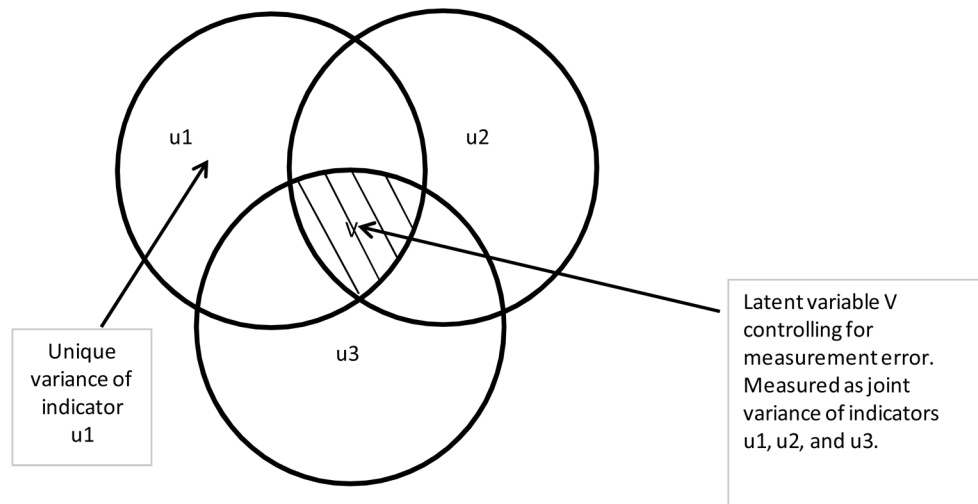
Figure 3.
Measurement of a Latent Variable as the Joint Variance of Three Indicators (u1, u2, and u3)
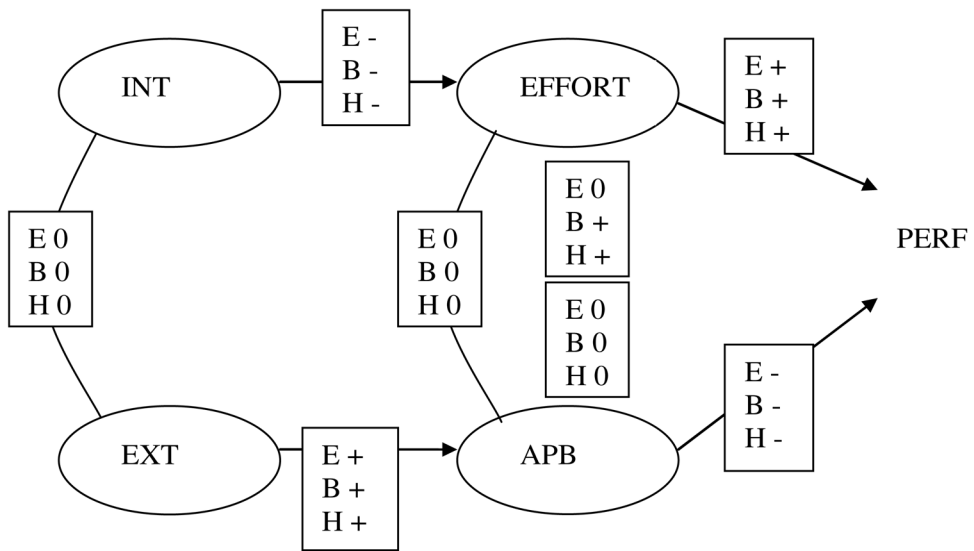
Figure 4. Direction of Significant Paths in Full Structural Equation Model, by Race, Compared to Expected Direction of Effects Based on Stereotype Threat Theory

Note: E=expected direction of effect based on stereotype threat theory, B=black students, H=Hispanic students. Only significant paths are shown with a + or − sign; all 0's indicate statistically zero coefficients.

Table 1

Indicators and Dimensions of Stereotype Threat Used in the Analyses

Internalization (INT)
1. On a scale of 0 (lazy) to 6 (hardworking), do members of your own racial group tend to be lazy or hardworking?
2. On a scale of 0 (unintelligent) to 6 (intelligent), do you think people in your own racial group tend to be unintelligent or intelligent?
3. On a scale of 0 (give up easily) to 6 (stick with it), in general, do you think people of your own racial group tend to give up easily or stick with a task until the end?
Externalization (EXT)
4. On a scale of 0 (treat equally) to 10 (discriminate against others), do you think Whites tend to treat members of other racial groups equally, or do they tend to discriminate against people who are not in their group?
5. On a scale of 0 (treat equally) to 10 (discriminate against others), do you think Asians tend to treat members of other racial groups equally, or do they tend to discriminate against people who are not in their group?
6. On a scale of 0 (total agreement) to 10 (total disagreement), to what extent do you agree that: If instructors hold negative stereotypes about certain groups, it will not affect their evaluations of individual students from that group.
7. On a scale of 0 (total disagreement) to 10 (total agreement), to what extent do you agree that: If other students hold negative stereotypes about certain groups, it will not affect their evaluations of individual students from that group.
Academic Effort (EFF)
8. How many hours (between 0–120) do you spend studying in the average seven-day week during the academic year?
9. In thinking about how hard to try in your college studies on a scale from 0 (no importance) to 10 (utmost importance), how important for you is it to learn the course material?
10. On a scale of 0 (no effort) to 10 (maximum possible effort), how hard would you say you have been trying [academically] during this past year of college?
Academic Performance Burden (APB)
11. On a scale of 0 (total disagreement) to 10 (total agreement), if I let my instructors know that I am having difficulty in class, they will think less of me.
12. On a scale of 0 (total disagreement) to 10 (total agreement), if I excel academically, it reflects positively on my racial or ethnic group.
13. On a scale of 0 (total disagreement) to 10 (total agreement), if I do poorly academically, it reflects negatively on my racial or ethnic group.
14. On a scale of 0 (total disagreement) to 10 (total agreement), I don't want to look foolish or stupid in class.
15. On a scale of 0 (total disagreement) to 10 (total agreement), if I don't do well, people will look down on others like me.
Academic Performance (PERF)
16. Students' second and third semesters average grade-point-average (GPA)

Table 2

Means and Standard Deviations of Stereotype Threat Constructs and Control Variables, by Race, with Stereotype Threat Constructs Represented as Summed Indexes of Their Respective Indicators and Control Variables[1]

	Black		Hispanic		White		% Missing (black/ Hispanic respondents)	% Missing (white respondents)
	Mean	Stnd. Dev.	Mean	Stnd. Dev.	Mean	Stnd. Dev.		
Constructs of Stereotype Threat								
Internalization (INT)	**7.57**	**2.37**	**7.61**	**2.45**	**7.97**	**2.01**		
Own group's intelligence (0–6)	2.40	1.00	2.64	1.03	2.57	0.86	0.03	0.01
Own group is hard working (0–6)	2.56	1.12	2.43	1.15	2.72	0.82	0.02	0.01
Own group perseveres (0–6)	2.61	1.14	2.54	1.07	2.68	0.95	0.04	0.02
Externalization (EXT)	**26.57**	**6.70**	**23.46**	**6.24**	**22.81**	**5.91**		
Whites treat other races equally or discriminate (0–10)	6.90	2.17	5.98	2.17	5.94	1.90	0.02	0.01
Asians treat other races equally or discriminate (0–10)	5.77	2.38	5.07	2.26	4.93	1.97	0.03	0.01
Instructors' stereotypes do not affect evaluations of members of stereotyped groups (0–10)[2]	6.87	2.84	6.04	2.84	5.74	2.83	0.00	0.00
Students' stereotypes do not affect evaluations of members of stereotyped groups (0–10)[2]	7.03	2.69	6.37	2.56	6.21	2.50	0.00	0.00
Academic Effort (EFF)	**18.06**	**3.68**	**18.13**	**3.75**	**17.64**	**3.52**		
Average Number of Hours Studied in a 7 day Week/10 (0–12)	2.72	1.54	2.86	1.56	2.65	1.47	0.00	0.00
Importance of learning course material (0–10)	8.35	1.90	8.39	1.71	8.08	1.57	0.00	0.00
Self-reported [academic] effort during past year of college (0–10)	7.00	1.83	6.87	1.86	6.91	1.79	0.00	0.00
Academic Performance Burden (APB)	**24.91**	**9.46**	**21.40**	**9.89**	**17.17**	**8.22**		
Instructors think less of me for having difficulty in class (0–10)	2.29	2.54	2.46	2.44	2.38	2.14	0.00	0.00
Excelling academically reflects positively on my racial/ethnic group (0–10)	6.77	3.03	5.62	3.22	3.77	2.96	0.00	0.00
Doing poorly academically reflects negatively on my racial/ethnic group (0–10)	5.84	3.17	4.60	3.18	2.95	2.63	0.00	0.00
I don't want to look foolish or stupid in class (0–10)	5.60	3.14	5.21	2.90	5.36	2.57	0.00	0.00
If I don't do well, people will look down on others like me (0–10)	4.40	2.97	3.51	2.71	2.71	2.29	0.00	0.00
Academic Performance (GPA)								
Average Second and Third Semester GPA	2.98	0.47	3.10	0.49	3.34	0.42	0.00	0.00
Control Variables								

NIH-PA Author Manuscript   NIH-PA Author Manuscript   NIH-PA Author Manuscript

	Black		Hispanic		White		% Missing (black/Hispanic respondents)	% Missing (white respondents)
Demographic (DEM)								
Male	0.36	0.48	0.42	0.49	0.50	0.49	0.00	0.00
Intact Family	0.52	0.50	0.66	0.47	0.39	0.81	0.00	0.00
Avg. # Dependents (0–18yrs) at ages 6, 13, 18	2.18	0.89	2.18	0.89	0.78	2.14	0.00	0.00
First Generation Immigrant	0.08	0.26	0.20	0.40	0.05	0.21	0.00	0.00
Second Generation Immigrant	0.18	0.38	0.48	0.50	0.09	0.29	0.00	0.00
Socio-Economic Status (SES)								
One parent has B.A. (or equivalent)	0.16	0.37	0.16	0.37	0.08	0.28	0.00	0.00
Two parents have B.A. (or equivalent)	0.12	0.32	0.11	0.32	0.16	0.36	0.00	0.00
One parent has Advanced Degree	0.23	0.42	0.25	0.43	0.35	0.48	0.00	0.00
Two parents have Advanced Degree	0.15	0.36	0.16	0.36	0.30	0.46	0.00	0.00
% of college paid for by family (%/10)	0.39	0.36	0.48	0.38	0.69	0.33	0.04	0.02
Average Hours of Work for Pay During First Two Years of College (/10)	0.69	0.97	0.62	0.86	0.36	0.71	0.00	0.00
Index of Racial Ingroup Exposure (IEX)	**15.17**	**6.08**	**10.10**	**6.16**	**-**	**-**		
Strength of ingroup racial identity (0–10)	2.25	0.91	1.70	0.90	-	-	0.02	-
Percent of same-race friends growing up (%/10)	5.71	3.43	2.72	3.06	-	-	0.00	-
Percent black or Hispanic neighborhood composition growing up (%/10)	4.91	2.83	3.84	3.02	-	-	0.05	-
Social distance from whites growing up (0–10, reverse-coded)	2.29	0.82	1.84	0.86	-	-	0.00	-
Academic Preparation (DAP)								
Number of AP Courses Taken	2.59	1.92	3.19	2.13	3.55	2.09	0.00	0.00
High School GPA (0–4)	3.57	0.36	3.71	0.32	3.79	0.26	0.00	0.00
Self-rated preparation level (0–10)	5.57	1.65	5.51	1.67	5.50	1.56	0.05	0.02
N	833		736		814			

NOTE: Standard deviations in parentheses. Descriptive statistics based on sample resulting after list-wise deletion.

[1] Means of each construct of stereotype threat and the index of racial ingroup exposure (IEX) are shown as indexes based on sums of their respective indicators.

[2] This item was reverse-coded to accurately reflect its negativelyl valenced properties in comparison to the other items in the externalization construct.

Table 3

Results of Multiple Group Analysis Comparing Full Measurement Models to Summed Indexes for Black and Hispanic Students[1]

Latent Construct[3]	Confirmatory Factor Analysis			Constrained Factor Analysis			Difference χ^2 (d.f.)
	χ^2 (d.f.)	RMSEA	CFI	χ^2 (d.f.)[2]	RMSEA	CFI	
Internalization (INT)	35.52(4) ***	0.09	0.91	46.32(9) ***	0.15	0.90	14.96(5) *
Externalization (EXT)	48.45(8) ***	0.08	0.95	316.03(15) ***	0.15	0.63	247.87(7) ***
Academic Effort (EFF)	14.70(7) *	0.03	0.97	238.59(17) ***	0.12	0.20	214.20(10) ***
Academic Performance Burden (APB)	100.32(16) ***	0.08	0.94	407.75(25) ***	0.13	0.73	305.84(9) ***

*
Significant at .05,

**
Significant at .01,

***
Significant at .001. One-tailed test. Factor analyses include institution-level fixed effects.

[1]
Blacks and Hispanics are estimated separately for purposes of estimating the fit of both the confirmatory factor analyses and the summed indexes. The Difference in chi-square statistic displays a goodness-of-fit test between the confirmatory and constrained factor analyses for each latent construct using the Satorra-Bentler Scaled Chi-squared Difference Test.

[2]
Factor loadings for all indicator variables are set to 1 (within and across race and immigrant generation groups). Residual variances are equal across groups for each indicator variable.

[3]
Academic performance is not shown here because it has a single indicator and therefore a CFA-summed index comparison does not apply.

Table 4

Changes in Structural Equation Model Fits by Estimation Strategies for Black and Hispanic Students (N=1380)[1]

Model	Chi-square (d.f.)	Diff in Chi-sq (d.f.)[2]	Goodness of Fit Measures
Blacks and Hispanics estimated separately	981.99(576)***	N/A	CFI: .91
			RMSEA: .03
Blacks and Hispanics estimated together	1104.67(590)***	122.68(14)***	CFI: .88
			RMSEA: .04

*
Significant at .05,

**
Significant at .01,

***
Significant at .001. One-tailed test.

[1]
List-wise deletion leads to a sample of 735 black students and 645 Hispanic students. Models include institution-level fixed effects.

[2]
Difference in Chi-square statistic indicates difference from the model where blacks and Hispanics are esimated separately.

Table 5

Comparison of Structural Equation Model (SEM) and Path Analysis Results Showing Relationships Between Internalization, Externalization, Disidentification, Academic Performance Burden and Academic Performance with and without Racial Exposure Control Measures, by Race

Standardized Coefficients

	(1) SEM: Latent Variables with Multiple Indicators for Each Stereotype Threat Construct			(2) Full SEM: Latent Variables with Multiple Indicators for Each Stereotype Threat Construct[2]		(3) Path Analysis: Summed Indices for Stereotype Threat Constructs	
	White	Black	Hispanic	Black	Hispanic	Black	Hispanic
Structural Parameters for Paths Between Constructs[1]							
Internalization → Academic Effort (EFF)	−0.14**	−0.17**	−0.19**	−0.18**	−0.21***	−0.09**	−0.15***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.03)	(0.03)
Academic Effort (EFF) → Academic Performance (GPA)	0.27***	0.15**	0.32***	0.17***	0.36***	0.09**	0.20***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.03)	(0.04)
Internalization → Academic Performance (GPA)	0.12**	0.10*	0.14**	0.11*	0.17**	0.07*	0.09**
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.03)	(0.04)
Externalization → Academic Performance Burden (APB)	0.04	0.13*	0.05*	0.08*	0.03*	0.03	0.06
	(0.05)	(0.06)	(0.02)	(0.04)	(0.01)	(0.03)	(0.04)
APB → Academic Performance (GPA)	0.02	−0.05*	−0.03*	−0.06*	−0.03*	−0.03	−0.02
	(0.04)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)
Externalization → Academic Performance (GPA)	−0.04	−0.01	−0.03	−0.03	−0.07	−0.08*	−0.05
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)
Internalization with Externalization	0.31***	0.23***	0.39***	0.23	0.37	0.10**	0.15***
	(0.06)	(0.06)	(0.06)	(0.13)	(0.20)	(0.03)	(0.04)
Academic Effort (EFF) with Academic Performance Burden (APB)	0.08**	0.13*	0.04	0.16	0.04	0.10**	0.05
	(0.05)	(0.06)	(0.05)	(0.10)	(0.06)	(0.03)	(0.04)
Measures of Ingroup Exposure[3]	-	-	-	X	X	X	X
N (by race)	853	918	792	918	792	918	792
N (total sample, by model)	2563			1710		1710	

Standardized Coefficients

	(1) SEM: Latent Variables with Multiple Indicators for Each Stereotype Threat Construct			(2) Full SEM: Latent Variables with Multiple Indicators for Each Stereotype Threat Construct[2]		(3) Path Analysis: Summed Indices for Stereotype Threat Constructs	
	White	Black	Hispanic	Black	Hispanic	Black	Hispanic
Chi-squared (d.f.)	4563.05(2400)***			2421.06(1289)***		29.64(8)***	
RMSEA	0.03			0.03		0.05	
CFI	0.85			0.91		0.81	

*
Significant at .05,

**
Significant at .01,

***
Significant at .001; one tailed test.

NOTE: Reporting Standardized coefficients from sample resulting from multiple imputation of five datasets. Standard errors in parentheses. Models include institution-level fixed effects.

[1]
Structural residual error correlations are estimated between internalization and externalization and between academic performance burden and disidentification. Measurement error correlations are estimated between indicators v6 and v7 and v10 and v11 (See Table 1 for list of variables by number).

[2]
Factor loadings for the indicators of each construct of stereotoype threat in this full model are shown in Appendix B.

[3]
Measures of Ingroup Exposure (IEX) include: Strength of racial ingroup identity (0–10), social distance from whites growing up (0–10), percent of friends of same race as respondent while growing up (scaled 0–10), and percent black or Hispanic neighborhood composition growing up (scaled 0–10). IEX measures reported in Appendix A.