CrossMark

Does a Self-Affirmation Intervention Reduce Stereotype Threat in Black and Hispanic High Schools?

Jenifer L. Bratter[1] · Kristie J. Rowley[2] · Irina Chukhray[3]

Abstract The risk of confirming negative stereotypes about one's social group, known as *stereotype threat*, depresses academic achievement among students of color and contributes to racial gaps in achievement. Some work finds that stereotype threat may be alleviated through self-affirmation exercises, translating into improved performance among students vulnerable to threat. However, this work has been conducted primarily in settings where students of color represent a relatively small segment of the student population. The current study explores whether this intervention is efficacious in schools where students of color are the majority. Through a randomized controlled trial of 886 students in three high schools (one predominantly black, one predominantly Hispanic, and one mixed race school), we administered self-affirmation exercises over the course of an academic year. We find no clear evidence that self-affirmation promoted higher standardized test scores or higher grades within the sample. The null findings highlight the complex nature of academic challenges in segregated contexts and raise important questions about the nature of stereotype thereat in such contexts. Importantly, this suggests that solely enhancing self-integrity may not be sufficient to close academic race-based gaps.

✉ Jenifer L. Bratter
jbratter@rice.edu

1 Department of Sociology, MS-28, Rice University, 6100 Main Street, Houston, TX 77005, USA

2 Department of Sociology, Brigham Young University, Provo, UT, USA

3 Houston Education Research Consortium, Kinder Institute for Urban Research, MS-258, Rice University, 6100 Main Street, Houston, TX 77005, USA

Introduction

Achievement gaps among whites, blacks, and Hispanics in standardized test scores and grade point averages (GPAs) are an on-going unfortunate feature of educational performance in many of today's public schools (Berends and Penaloza 2010). Black and Hispanic students score as much as a standard deviation or a grade level behind their peers on a variety of standardized assessments, indicating academic struggles that may impact minorities' futures as these groups are projected to occupy significant shares of the working age population (Lichter 2013; Passel 2011). This "puzzle of underachievement" (see Massey et al. 2011) is even more vexing given that a student's socioeconomic disadvantage, while a core driver of academic ability, does not fully explain race/ethnic gaps in academic performance (Coley 2002; Downey 2008; Fryer and Levitt 2010; Massey et al. 2011; Phillips 2000). Even middle class minority students attending elite schools underperform when compared to their white peers (Massey et al. 2011). Understanding and ameliorating these gaps remains a persistent challenge to policy makers and researchers.

A small but growing body of work has explored correcting achievement gaps by addressing social psychological "threats" to student performance. *Stereotype threat*, the anxiety of confirming group-specific stereotypes of poor intellectual ability, has well-established links to the poor performance of racial minority students in college, high school, and even in earlier grades. Ameliorating experiences of threat is the focus of interventions aimed at enhancing student self-concept and self-integrity by

exposing students to writing exercises that are "self-affirming" or allowing students to write freely on what they value. Cohen et al. (2006) have generated one such intervention that effectively closed the race/ethnic gap in grade point average (GPA) for students exposed to the treatment condition.

While impressive, the question remains whether such gains can be replicated across contexts, particularly in schools and school districts where students of color represent a large share of the district and achievement gaps are likely to be large. The current study reports the results of an intervention implementing the self-affirmation exercise developed by Cohen et al. (2006) within a large urban school district in the state of Texas. Adding to a growing body of work that has fielded this intervention, we add to this literature by fielding these exercises within a unique race/ethnic context. This urban school district is majority–minority where 8% of the students are white (non-Hispanic), 25% are black and about 60% are Hispanic.[1] Additionally, several schools are composed entirely of black or Hispanic students, while white students represent one-third or more of students in other schools. We implement this exercise in three distinctive high school contexts: a predominantly black school, a predominantly Hispanic school, and a school whose composition is divided roughly into thirds: black, Hispanic, and white. Few if any studies have fielded an intervention targeting stereotype threat such contexts. This is a critical gap in light of the fact that schools attended by black and Hispanic students are decreasingly likely to include white students and these schools are most likely to be poor performing (Condron et al. 2013). Therefore, exploring the efficacy of this intervention in such a context has the potential to speak to race/ethnic gaps that plague a growing number of American schools.

Our findings draw on a sample of 886 ninth graders who received this intervention during their English class at multiple points during the 2012–2013 academic year. We explored its effectiveness by comparing academic achievement on two standardized examinations and students' Spring semester grade in their English class (the targeted class) of those receiving the "affirmation" treatment exercise and those receiving the "neutral" control exercise, where the exercise prompt presents students with a neutral or non-emotion eliciting topic to write about. We first gauge whether there are stronger treatment effects

among black and Hispanic students than white students in test scores and grades; we then gauge whether treatment effectiveness varies by race/ethnicity of student or by school setting. In addition, we examine whether the effect of the treatment for black and Hispanic students varies by their school's racial/ethnic composition.

Stereotype Threat and Affirmation Exercises: Can the Impact of Anxiety be Reduced?

The role of social psychological issues, specifically experiencing anxiety, has emerged as a potent contributor to the academic abilities of marginalized students. Stereotype threat, as identified by Steele and Aronson, refers to anxiety over "being at risk of confirming…a negative stereotype about one's group" and this experience inhibits performance on academic tasks (Steele and Aronson 1995: 797). Therefore, while anxiety affects everyone, this form is linked specifically with being reflective about membership in a minority or socially subordinate group (Downey 2008). Stereotype threat occurs in the presence of some trigger, either the explicit linking of a task to intellectual ability, or a reference to marginalized status (e.g., marking race on a form or mentioning race) or even being in a context where a student's minority status is made salient through being one of the only members of this group in one's class. Members of groups who are stereotyped as having low intellectual ability experience anxiety when facing these triggers and perform worse on intellectual tasks than students who are not members of minority groups. Therefore, it is that anxiety, not poor ability, that drives down performance (Osborne 2001).

A growing body of scholarship finds evidence that stereotype threat negatively affects the academic performance of black and Hispanic students across a variety of age groups and school settings (Brown and Day 2006; Cohen et al. 2006; McKay et al. 2002; Nguyen and Ryan 2008; Schmader and Johns 2003; Steele and Aronson 1995; Walton and Spencer 2009; Wasserberg 2014). Stereotype threat impacts performance on cognitive tasks conducted by black and Hispanic students as young as elementary school age (McKown and Weinstein 2003; Wasserberg 2014). Most studies, however, focus on minority college students, while only a select few focus on minority middle school students, and even fewer studies focus on stereotype threat impacting minority high school students (see Marchand and Taasoobshirazi 2013). This is a critical gap as Osborne's (2001) analysis reveals that threat explains a meaningful portion of the racial differences in academic performance among high school students. For example, Kellow and Jones (2008) find that black high school students score significantly lower on a high-stakes test than whites when they are told that their performance on the test

---

[1] The school district that generated the information on composition employs race/ethnic categories used by the U.S. government that reflect the standards set forth by the Office of Management and Budget. Therefore Hispanic, while technically an ethnic group, is treated as a racial one, amidst other (non-Hispanic) categories of white, black, Asian, American Indian, "Two or more Races", and some other race.
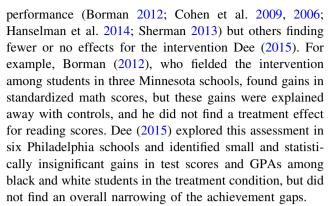
will predict their performance on an important standardized test.

A critical question emerging from this work is whether the impact of *threat* can be reduced and academic performance improved. There is evidence that academic ability can improve when threat is relieved (Cohen et al. 2006). Enhancing students' self-integrity, that is their feelings about their self-worth, can be a way to limit or even remove the influence of threat-driven anxiety. While several similar interventions exist (see Walton and Spencer 2009), Cohen et al. (2006) reported the most dramatic results. This intervention is, according to Cohen et al. (2006), designed to specifically target self-confidence in an attempt to find methods that lead to higher test scores. This self-affirmation exercise is aimed to function as a "catalyst" that boosts students' self-confidence and self-integrity while allowing their abilities to be unencumbered, translating, theoretically, into better performance (Purdie-Vaughns et al. 2009), specifically during an anxiety-provoking moment when a student is about to be evaluated.

The intervention, which is a writing exercise completed four times during the school year, was fielded by middle school students to affirm their self-integrity and raise their grades (Cohen et al. 2006: 1307). The intervention takes an average of 15 min to complete and merely asks students to write about what is important to them at several points during the year. Among seventh graders, 70% of black student participants at a middle school benefited from the writing intervention designed to reduce stereotype threat; the overall racial gap decreased by 40%, with treatment students experiencing significantly higher GPAs (Cohen et al. 2006). Similarly promising results are reported in a two-year follow-up study indicating that the lowest-achieving black students who took the intervention improved the most when compared to the control group (i.e., students who completed a similar exercise that was designed to have a neutral impact on their self-integrity) (Cohen et al. 2009). Likewise, an enhanced replication with a different body of students and teachers, in which teachers read students' writing assignments in an effort to alter their presumed stereotypical views of students, researchers found that the writing exercise significantly improved black students' overall GPA and narrowed the race gap in GPA to nearly nothing, while gaps persisted for those taking the control or neutral exercise (Bowen et al. 2013).

Placing Self-Affirmation in Context: How does Racial Composition in Schools Matter?

These results are impressive, but an important remaining question is whether similar treatment effects can be found across a variety of school settings. The results thus far are mixed, with some studies finding gains in academic performance (Borman 2012; Cohen et al. 2009, 2006; Hanselman et al. 2014; Sherman 2013) but others finding fewer or no effects for the intervention Dee (2015). For example, Borman (2012), who fielded the intervention among students in three Minnesota schools, found gains in standardized math scores, but these gains were explained away with controls, and he did not find a treatment effect for reading scores. Dee (2015) explored this assessment in six Philadelphia schools and identified small and statistically insignificant gains in test scores and GPAs among black and white students in the treatment condition, but did not find an overall narrowing of the achievement gaps.

While prior work establishes that gains are possible using this intervention, these studies provide little insight into the role of the school itself. In an innovative departure, Hanselman et al. (2014) explored the efficacy of the assessment in contexts that may be deemed high threat compared to those that are low threat. Contexts are considered *high threat* when black or Hispanic students comprise a relatively small share of the school's racial composition. and interracial contact is a regular facet of a student's experience in the classroom. When students are in the racial minority, they are more aware of their minority status and this awareness cues their anxiety of fulfilling stereotypes of the group (Cohen and Garcia 2005; Martiny et al. 2012). Notably, this work identifies stronger treatment effects for students in high-threat contexts.

Children of color increasingly attend schools that are composed entirely of other black or Hispanic students. These are contexts that have largely been untapped for assessment of stereotype threat or for interventions aimed at reducing its effects. Minority students attending schools with a large concentration of black and Hispanic students fall even more behind their white counterparts in both reading and math, compared to minority students in racially integrated schools or schools dominated by whites (Condron et al. 2013). Since the number of black and Hispanic students attending schools with large shares of whites has declined dramatically (Orfield et al. 2014), understanding the experience of students attending majority black or Hispanic schools is critical to addressing ongoing achievement gaps.

Fielding the Intervention in a Majority-Minority School District and All Black/Hispanic Schools

How might the effect of a self-affirmation exercise vary across these different types of schools? On the one hand, we may find few effects of self-affirmation exercises across these schools because these are all low-threat contexts according to previous research. For example, in the study by Hanselman et al. (2014), high-threat contexts were schools where blacks and Hispanics represented 13 and 14% of the

student body, while low-threat contexts were schools with 25 and 27%, respectively. By these standards, all the schools in the current study, and nearly all in this district, would be characterized as low threat. Presumably, whether a student of color is in a mixed environment or attending a school where their group is a large majority, they would likely not stand out as racially distinctive in a way that would correspond to a threat (i.e., be the only one in their classroom) and perhaps are less likely to confront the burdens of race-based stereotyping. In addition, even if white students are present, these students may be more cognizant of racial difference (i.e., negative stereotypes about one's racial group) in these environments and may develop a strong affinity to communities of color (Morris 2006; Perry 2002). While students of color face academic challenges in these environments, it is not clear if they are due to stereotype threat.

On the other hand, it is possible that this intervention is effective in raising the academic performance of black and Hispanic students in these contexts, even if they could be understood as low threat, because students of color may encounter a high-threat context in their classrooms. Moreover, issues of race/ethnic disparities are not necessarily absent in schools where multiple groups are equally represented, since race shapes students' self-perception and understanding of their own chances to succeed, as well as how students understand and view other students (Bettie 2003; Lewis 2003; Morris 2006; Tyson et al. 2005).

Additionally, it may be possible that self-affirmation may be effective even if students of color are the majority. Schools composed entirely of black or Hispanic students are more likely to be lower performing, least desired by parents, and most likely to face instability in terms of teacher and administrator turnover and even possible closure (Condron et al. 2013; Lee and Klugman 2013; Orfield et al. 2014). As a result, it is possible that even in the absence of interactions with racial out-group members, students' self-confidence and their sense of self-worth are compromised in these settings, producing lower academic performance, independent of ability. In particular, high-stakes testing moments are classically associated with anxieties and boosting self-integrity might be particularly effective in these environments.

Research Questions

We advance the following three research questions. First, we test whether receiving the self-affirmation treatment exercise led to better academic performance, independent of the race/ethnicity of the student, the race/ethnic composition of the school, or other relevant background characteristics. To capture academic performance, we employ two central metrics

(a) teacher assigned grades in English (the class where the intervention is fielded) and (b) performance on standardized tests. Prior assessments have uncovered improvement in grades (see Cohen et al. 2006), but fewer have identified improvement in standardized test performance. We compare the grades and standardized test scores of those receiving the self-affirmation *treatment* to those who received an exercise that did not include an affirmation component (i.e., the *control* exercise). If the treatment has the intended effect, students who received and completed the treatment exercises should have better academic performance at the close of the academic year than the students in the control condition.

Second, we ask whether the intervention is more or less effective for students by race/ethnicity. As this intervention is designed to alleviate anxiety that taxes minority students' performance in particular, we expect the strength of the association to vary by race, with greater benefits for black and Hispanic students, than for white students. A contribution of our work is addressing whether Hispanic and black students experience the same treatment effects. The earliest and most persistent evidence has been found among black students experiencing stronger treatment effects than white students, but we have less information on Hispanics (Steele and Aronson 1995).

Third, we advance a similar question regarding race/ethnic composition of the school as prior research has been fielded in settings where white students are the majority and the probability of interracial contact is high (see Hanselman et al. 2014). Specifically, we investigate whether the treatment effects for black and Hispanic students are stronger in a racially mixed school, where they encounter race or ethnically different students than the corresponding school where they are members of the majority and thus only interacting with students of the same racial group. As stereotype threat is cued in contexts where inter-group contact is frequent, we anticipate the impact of the treatment on academic performance will be highest in the racially mixed school and relatively weaker in the predominantly black and Hispanic school.

Data and Methods

The Schools

With growing evidence that USA schools are becoming more racially segregated (Frankenberg and Lee 2002; Orfield et al. 2014), a growing number of schools are almost entirely composed of one race/ethnic minority group, either black or Hispanic. We argue above that this may shift the experience of stereotype threat and present a compelling context to field an intervention aimed at reducing its impact on academic performance. We drew on

data collected by the school district. The district requested that the Houston Education Research Consortium (HERC) analyze the data. HERC conducted a randomized experimental design to field self-affirming writing exercises to students in a large metropolitan school district in Texas. The district administered the exercises, collected the data, and shared the data with HERC.

The specific schools were identified through a stratified school search based on these racial/ethnic criteria.[2] Our sampling approach prioritized identifying different schools with distinctive race/ethnic compositions, highlighting the different types of schools that black and Hispanic students find themselves. We fielded the intervention in a school district where only 8% of the students are white (non-Hispanic), 25% are black, and more than 60% are Hispanic. Moreover, there is a great degree of variation across school campuses, with several schools composed almost entirely (in excess of 90%) of black students or Hispanic students, while other schools are more than 50% white or Asian, despite representing less than 15% of the district (Texas Education Agency 2012). Notably, there are no schools in this district where white students are a large majority.

We identified three types of schools that represent three distinct but common configurations within the district. We identified a *mixed* school with roughly equal shares of black, white, and Hispanic students; a *predominantly black* school where black students made up almost 100% of the school; and a *predominantly Hispanic* school where Hispanic students made up more than 95% of the school.

Ideally, these schools would be similar in many respects to truly tap the effect of race/ethnic composition; however, they differ on several dimensions. The predominantly black and Hispanic schools are lower performing than the mixed race campus. As of 2011, one year prior to the administration, students from the predominantly black and Hispanic schools earned scores well below the state average on state standardized tests.[3] A salient question is whether this

intervention can be effective in low-performing contexts. We will return in our discussion to this issue. Schools also differed in terms of class size. Across the total 54 classrooms where the intervention was fielded, the school we defined as "mixed" (i.e., roughly one-third black, white, and Hispanic students), the average class size was 13 students; meanwhile, the average class size was 19 students in the "predominantly Hispanic" school and 22 students in the "predominately black" school. Despite these differences across schools, a comparison across these contexts signals an important contribution as these educational settings are largely absent in the work on stereotype threat.

Procedure

The administration aimed to follow Cohen et al. (2006) affirmation writing exercises. Ninth-grade students (roughly aged 14) were given a series of four short writing exercises (lasting between 15 and 20 min) in their English classrooms during the 2012–2013 academic year. With the approval of the district and the principals and deans of instruction at the individual schools, this exercise was integrated into their daily assignments in class and we were not obligated to gain informed consent from the students or their parents. To identify moments of high academic stress, when anxiety related to stereotype threat is theoretically operating, these occurred no more than three days prior to four high-stakes tests: (1) a national pre-college exam administered in October, (2) the semester final exam in December, (3) a school-based examination in February, and (4) an end-of-the-year standardized test that is required for graduation and is used to determine promotion to the next grade in April.

We aimed to administer the intervention in a way that mimicked, as much as possible, its appearance as part of the regular curriculum. We therefore employed a double-blind approach. We began by soliciting school administrators to allow their teachers, as opposed to members of the research team, to give the exercises to their students during class. Once we received approval, we trained teachers on the administration. They were informed that our goal was to examine the potential benefits of a special type of writing assignment; one where students were allowed to write freely therefore students should be instructed not to worry about spelling or grammar. We also explained that we were fielding various formats of the writing exercise but did not indicate that these signaled a "treatment" or "control" condition; therefore, students should be instructed to not compare their assignments to their peers. During the training, the research team intentionally did not mention stereotype threat, race, or self-affirmation. We, then, provided the exercises directly to teachers a day before the intended administration date in a

---

[2] We compiled a list of schools in the district by racial/ethnic composition identifying schools composed of 80% or greater minority students ($N = 17$). Similarly, we compiled a list of schools with at least 20% white students. Very few have equal shares across the three major race/ethnic groups; as a result there were fewer "mixed" schools to select for this study. Once targeted, we gained consent from school officials as well as the district to field the intervention in ninth grade English classrooms.

[3] This information comes from an analysis of the now phased-out TAKS (Texas Assessment of Knowledge and Skills). The Texas Education Agency rates the accountability of schools in terms of meeting state standards in standardized assessments as well as longitudinal completion rates and annual dropout rates with the highest possible rating of "exemplary". As of 2011, the accountability rating of the mixed school was "recognized," the rating of the majority Hispanic school was "academically acceptable," and the rating of the majority black school was "academically unacceptable" (Texas Education Agency 2011).

sealed envelope that only had the specific student's name at the top.

The exercise itself included a list of 12 items that reflected values students might hold (e.g., spending time with friends or family, enjoying sports, enjoying music). Students assigned to the treatment circled those values that were important to them and then elaborated on why these values mattered (e.g., why the student values sports or spending time with friends and family). The proprietary nature of the self-affirmation exercise limits our ability to provide more description, but we invite readers to see Cohen et al. (2006) for more specifics. The control condition provided a similar list of values but instead of a prompt to elaborate; the control provided students with a brief writing prompt on a "neutral" scenario (e.g., whether the USA should continue using the penny given increasing production costs). Later in the cycle, students were given modified versions of these exercises to reduce repetition. Students in the control conditions were provided a new neutral scenario for each exercise.[4]

These writing exercises were adapted from the self-affirming intervention developed by Cohen et al. (2006), which has been fielded successfully in several other studies (Borman 2012; Dee 2015; Hanselman et al. 2014; Sherman 2013). In adapting these exercises, we employed several changes. First, our writing prompt for the control exercise differs due to pilot testing that suggested the original control writing prompt was confusing for students. The purpose of the control is to elicit non-emotional student responses and our revised control prompt succeeded in gathering written responses that did not elicit student emotional reactions. Second, we have English teachers administer the writing exercises, as opposed to researchers, to signal that this exercise is part of the classroom curriculum. Finally, in order to maintain the appearance of a typical English curriculum assignment, we did not employ the racial climate survey Cohen and colleagues administered that gauges the centrality of race for students daily experiences. We acknowledge that these changes limit the degree to which this is a true replication of the Cohen and colleagues study, but we have maintained all other features of the self-affirmation exercises.

Sample

Identifying our study sample from the population of students engaged the following steps. Once three schools were selected, we obtained the roster of roughly 1600 ninth-grade students (with an average age of 14–15) across the schools. Administrative data on students' race/ethnic background, gender, prior achievement on standardized examinations were merged on to these files. Students were stratified by race/ethnicity across all three schools to ensure representation of each race/ethnic group (black, white, Hispanic) in the treatment and control groups. We then conducted a student level randomization within each school, randomly assigning students to either treatment or control conditions, a designation they maintain throughout the school year. We provided the intervention materials to teachers in individually marked envelopes for each classroom. Teachers passed out the materials with student names already printed on each intervention so that students received their assigned treatment or control exercise. Follow-up tests indicated no statistically significant differences between treatment and control in gender or race/ethnicity across the full sample and within schools. Notably, the balance on race/ethnicity held even within schools identified as consisting of predominantly black and Hispanic students. We excluded students in English as a Second Language (ESL) or other Language English Proficient (LEP) programs or special education. We also excluded students who entered school after the school year began, those who shifted classrooms, or who exited the school before the year was over. This limited our initial sample to 1245 students.

Our goal was to limit our analysis to students who are "fully exposed" to the administration—that is they were present in the classroom and received as well as returned the appropriate exercise. We ultimately dropped students who received or returned the incorrect exercise, were absent during the administration, or changed classrooms. This resulted in the removal of roughly 28% ($N = 359$) of the initial sample of 1245 students. While, ideally, these exclusions should be random, ancillary analyses reveal students included in our sample are disproportionately white, female, and not disadvantaged. Also, students excluded from our analyses are more likely to attend the predominately black school in our sample. The significant association between inclusion in our sample and these characteristics raise important questions for the analysis that we return to in the discussion. Our final study sample includes 886 students, and we specify the number of students in each school in the forthcoming tables.

Measures

Dependent Variables

We employ three different indicators of student achievement to capture different facets of academic performance: students' standardized test results in Reading and Algebra

---

[4] Administration-related difficulties delayed the administration of the first exercise in one school following the first examination date. While we regard this as a lost outcome, it does not have any bearing on our analysis presented in this paper as we focus on the patterns that occurred at the close of the academic year.

and the English grade teachers assigned students at the close of the Spring semester (or academic year). Cohen's initial test that revealed positive results employed GPA and others have found varying treatment effects of GPA versus standardized test scores (Dee 2015; Hanselman et al. 2014). We assess the academic performance of students with a continuous measure of the grade received, ranging from 0 to 100, at the close of the Spring semester in English class (the course where they received the exercises). We refer to this outcome as "Spring Semester English grade". We also draw on their standardized test score in English Reading I and the standardized test score in Algebra I on the STAAR-EOC (State of Texas Assessments of Academic Readiness-End of Course Results). We hereafter refer to these assessments as *Reading* and *Algebra*. The district administered this assessment for the first time during the year in which we conducted this intervention. The EOC assesses students' knowledge and skills necessary for success in future academic courses.

We examine STAAR performance as a continuous outcome and observe whether taking the treatment relative to the control translates into increases in average score performance, which aligns with the emphasis of previous tests of this intervention. The STAAR scores are horizontally scaled, which allows for comparison across test forms from year to year for a specific subject assessment. Horizontal scale scores cannot be compared to the scale scores of other students in other grades in a different subject area; scale scores can be compared if in the same subject area.

Independent Variables

Our key independent variables are presence in the treatment or control conditions (1 = treatment, 0 = control), race/ethnicity, and school. Our analysis only focuses on the three largest racial/ethnic groups in the district, under the Office of Management and Budget designation that is adopted in the categories provided by the PEIMS (Public Education Information Management System) data: whites (non-Hispanic), blacks (non-Hispanic), and Hispanics. Students in all other categories, including Asians, Pacific Islanders, multiracial students, and students classified as "Some other race" are placed in an "Other race" category in our analyses.

Controls

We adjust for a variety of characteristics, coded dichotomously, that may have implications for academic performance. These are gender (male/female), economic disadvantage, and English proficiency. Gender is dichotomous (male = reference). Economic disadvantage is a three category variable where students are not disadvantaged (reference), Free/Reduced Lunch (FRL), or living in poverty (have household income at or below the poverty line). Status as FRL is a measure of student socioeconomic status gauging students who are eligible to partake in the Free Lunch Program under the National School Lunch Act and is determined by household size and income or based on categorical eligibility, which is a proxy for income data, roughly less than 18% of the federal poverty line. In our data, we further distinguish students whose household incomes are below the federal poverty line. Status as LEP, also known as English Language Learners (ELLs), refers to students who are eligible for language assistance programs (e.g., English as a Second Language or bilingual education).

Analysis Plan

Our goal is to compare treatment and control students in terms of academic performance. We therefore conduct a series of OLS regression models predicting our key outcomes, all of which are continuous, assessed at the close of the academic year. To test for the effect of the intervention (research question 1), we assess the average academic performance of treatment relative to control students in a baseline (Model 1) and a fully adjusted model (Model 2) that includes covariates for race/ethnicity and school and other relevant characteristics. To test whether the intervention's effect intersects with race/ethnicity (research question 2) and school (research question 3), we introduce a series of interactions between receiving the treatment and student's race/ethnic background and school (Models 3 and 4, respectively).

Although ideally, both sets of interactions would appear in the same model, our sample cannot sustain such a modeling approach because there is limited racial variation across schools. Specifically, white students are only present in one school ("mixed") and black and Hispanic students are only present in the "mixed" school and the school where they are the large majority. However, we do explore how differing school composition may shape the effect of the intervention for specific groups of students (research question 3) in models where we stratify our sample by race/ethnicity and explore the effect of the intervention and its interaction with school on academic performance among black students (Table 5) and then Hispanic students (Table 6). In these models, we explore if the effect of the treatment for black or Hispanic students differs by the schools attended (i.e., either the mixed-setting or the corresponding predominate setting). Similar comparisons for white students are not possible as white students are only present in one school.

Results

Descriptive Statistics

We begin with a description of our respondents and the distribution of their characteristics in Table 1. In all, we have 886 students who received and returned either the treatment or control conditions four times over the course of the academic year. The number of students in either condition, listed in the bottom row, is basically equivalent (number of treatment students = 430, control students = 456). We show percentage distributions of the entire sample in the "All" column and the sample stratified by treatment and condition in their respective columns. We also show the corresponding frequency within each category in the "Frequency" column.

We present information on students' race/ethnicity, gender, FRL and LEP status, and school type in Table 1. In our sample, we find generally equivalent percentages in both treatment and control conditions. Regarding race, our sample is broadly reflective of the district. The majority of the students in the sample are Hispanic (more than 60%), roughly a quarter are black, and 11% are white, with slightly more white students in the control condition. The remaining 1% includes all other race/ethnic groups, including Asians, Pacific Islanders, Native American,

Table 1 Distribution of independent variables, by treatment condition

	Total	Treatment %	Control %	Frequency
Race				
Black	26	26	25	227
Hispanic	62	63	61	547
White	11	9	13	99
Other	1	2	1	13
Gender				
Female	50	50	50	443
Economic status				
Not disadvantaged	27	25	29	242
Free/reduced lunch	38	41	35	338
Live in poverty	35	33	36	306
Current LEP status	7	9	6	66
School type				
Mixed	35	35	35	309
Majority Hispanic	50	50	50	444
Majority black	15	15	15	133
Total	886	430	456	886

LEP Limited English proficient

Asterisks indicate statistical significance of difference between treatment and control ($^+$ p<0.10, * $p < .05$, ** $p < .01$, *** $p < .001$). None of the differences are significant

"Some Other race" students, and multiracial students. Half of all students in the sample are female, roughly one quarter are "not disadvantaged", while almost 40% qualify as free/reduced lunch and the other one-third are "in poverty." About 7% of students in our sample are classified as limited English proficiency (LEP). Half of our students come from the school composed of majority Hispanic student population, 35% are from the mixed school, and 15% are from the majority black school.

Fortunately, we find no statistically significant differences in these characteristics across the treatment and control, indicating that our efforts to randomize the administration yielded balance among those included in the analytical sample. The racial and gender distributions are nearly identical across both conditions. We see slightly more free and reduced lunch students among the treatment and slightly fewer students in the majority Hispanic school who received the control. Our forthcoming analyses will control for these characteristics as they have meaningful associations with academic performance; however, we do not anticipate that they will explain away the role of the treatment or control on grades or standardized scores.

Multivariate models

We now turn to our multivariate models. In Tables 2, 3, and 4, we explore our research questions in a series of models that assess the association of exposure to the treatment and Spring semester grade in English (Table 2), standardized scores on Reading (Table 3) and standardized test scores in Algebra (Table 4). In each table, we show the coefficients and associated standard errors and signal statistical significance with asterisks. We begin with the association between exposure to the treatment and grade received in English at the end of the semester. The coefficient in Model 1 reveals small differences in Spring semester grades between the students taking the treatment and their peers taking the control, and notably, our adjusted R^2 registers no variation explained. In Model 2, we continue to observe no effect, net of demographic characteristics, including race/ethnicity, gender, socioeconomic status, LEP status, and specific school. However, this model reveals very clear racial gaps in end of semester grades in English. Model 3 tests whether the efficacy of the treatment varies by race/ethnicity of the student. Not surprisingly, none of the effects are statistically significant and they are nominally small in size. In our final interactive model (Model 4) testing the treatment's potential impact across schools, we continue to find little evidence that the effect of the treatment on semester grades in English varies across school context. In sum, we find little to no evidence that the exposure to the treatment corresponded to higher grades in English and no evidence that the treatment's

Table 2 Regression models predicting Spring Semester English grade

	Model 1	Model 2	Model 3	Model 4
Treatment	−0.073 (0.85)	0.12 (0.77)	0.09 (2.24)	0.04 (1.27)
Student race (ref = white)				
Black		−6.09*** (1.57)	−6.81** (1.99)	−6.06*** (1.58)
Hispanic		−6.09*** (1.51)	−5.87** (1.83)	−6.07*** (1.51)
Other		−4.44 (3.33)	−3.21 (4.67)	−4.46 (3.33)
Student gender (ref = male)				
Female		3.62*** (0.76)	3.62*** (0.76)	3.63*** (0.76)
Economic status (ref = not disadvantaged)				
Eligible for free or reduced lunch		0.49 (1.08)	0.52 (1.08)	0.46 (1.08)
Live in poverty		−3.21** (1.13)	−3.12** (1.14)	−3.20** (1.13)
Current LEP status		−0.07 (1.54)	−0.01 (1.54)	−0.03 (1.54)
School type (ref = mixed)				
Majority Hispanic		−6.30*** (1.09)	−6.33*** (1.09)	−6.05*** (1.38)
Majority black		−8.44*** (1.45)	−8.43*** (1.46)	−9.64*** (1.89)
Treatment × student race				
Treatment × black			1.38 (2.72)	
Treatment × Hispanic			−0.44 (2.44)	
Treatment × other			−2.45 (6.65)	
Treatment × school type				
Treatment × majority Hispanic				−0.50 (1.67)
Treatment × majority black				2.49 (2.44)
Adjusted R^2	0.00	0.20	0.20	0.20
Total	807	807	807	807

LEP Limited English proficient

Standard errors are included in the parenthesis

Asterisks indicate significance according to a two-tailed test ($^+$ $p<0.10$, * $p<.05$, ** $p<.01$, *** $p<.001$)

effect varied by race/ethnicity or depending on specific school.

Tables 3 and 4 show the results for the STAAR Reading and Algebra standardized tests, respectively. Using the same modeling approach, we again find little evidence that students who received the treatment did significantly better on standardized tests in Reading or Algebra, for the most part. Models 1, 2, and 3 for both outcomes assess the effects of the treatment and reveal no statistically significant difference in the academic outcomes between treatment and control students. We do find sustained racial gaps. On average, blacks and Hispanics score between 50 and nearly 100 points lower on standardized reading tests, relative to white students (see Table 3) and score 150 points lower on standardized tests in Algebra. Achievement gaps are also apparent across schools as those in the "mixed" school scored considerably higher than students in the other two schools. Turning to Model 3, we find no evidence that the influence of the treatment on either Reading or Algebra scores by race/ethnicity of the student.

Despite these patterns of null effects, we do find some evidence of the treatment's impact when appraising the interactions across schools in Reading (Table 3, Model 4), but no such effect is identified in Algebra scores. Specifically, we uncover a positive and significant effect of the interaction between treatment and school context for the majority Hispanic school, indicating potentially greater gains in standardized reading scores for these students than was achieved in the mixed school.

To further our examination of the roles of race and context, we turn to Tables 5 and 6 that explore the impact of the treatment among black (Table 5) and Hispanic students (Table 6). Unlike white students who are solely present in the "mixed" context, black and Hispanic students encounter schools where they are either in the minority (i.e., in the "mixed" school) or the large majority of students in their school. Theoretically, these conditions may shape the experience of stereotype threat and thus impact the treatment's effect on their academic performance. We present the results for each outcome across three panels using a similar modeling approach as found in

Table 3 Regression models predicting STAAR Reading scores

	Model 1	Model 2	Model 3	Model 4
Treatment	−22.52 (16.83)	−16.33 (13.81)	−18.20 (40.41)	−53.08* (22.84)
Student race (ref = white)				
Black		−86.58** (28.24)	−79.61* (35.96)	−83.20** (28.25)
Hispanic		−97.81*** (26.99)	−101.15** (33.08)	−95.31*** (26.98)
Other		−50.78 (59.89)	−88.16 (84.22)	−43.18 (59.92)
Student gender (ref = male)				
Female		39.45** (13.76)	39.50** (13.78)	38.94** (13.74)
Economic status (ref = not disadvantaged)				
Eligible for free or reduced lunch		−25.50 (19.78)	−25.72 (19.84)	−24.45 (19.76)
Live in poverty		−86.88*** (20.57)	−87.96*** (20.69)	−86.50*** (20.54)
Current LEP status		−150.38*** (28.21)	−151.13*** (28.28)	−153.20*** (28.21)
School type (ref = mixed)				
Majority Hispanic		−145.94*** (19.63)	−145.24*** (19.68)	−176.95*** (24.71)
Majority black		−270.05*** (25.93)	−270.16*** (25.97)	292.15*** (33.19)
Treatment × student race				
Treatment × black			−13.36 (48.83)	
Treatment × Hispanic			6.77 (44.05)	
Treatment × other			74.83 (119.81)	
Treatment × school type				
Treatment × majority Hispanic				61.86* (30.03)
Treatment × majority black				42.98 (43.38)
Adjusted R^2	0.00	0.34	0.33	0.34
Total	806	806	806	806

LEP Limited English proficient

Standard errors are included in the parenthesis

Asterisks indicate significance according to a two-tailed test ([+] $p<0.10$, * $p<.05$, ** $p<.01$, *** $p<.001$)

the previous tables. Among black students, we find no significant effects of the treatment on their final grade in English. However, we do find a significant association (at the $p < .10$ level) in Model 2 between exposure to the treatment and Reading scores (see Panel B) and Algebra scores (see Panel C). However, students receiving the treatment earned *lower* scores in their Reading and Algebra than those in control. This is also apparent in Model 3 for Reading scores, but not Algebra, where we gauge the potential interaction of treatment and school attended. Patterns for Hispanic students, shown in Table 6, echo the findings in Tables 2, 3 and 4, as the treatment seems to have little bearing on the academic performance, regardless of outcome.

Discussion

This study reports the findings of a fielded intervention aimed at relieving stereotype threat and ultimately, translating into better academic performance among minority students. Stereotype threat is a theory suggesting that poorer performance by minorities (racial or otherwise) in academic achievement can be traced back, in part, to performance anxiety that reflects fears of confirming stereotypes of minority groups. Although the intervention itself has been shown to be effective in some contexts, it has never been fielded in districts or schools with a heavy concentration of blacks and Hispanics as is found in our study site. Districts and schools that are composed entirely of black or Hispanic students are increasingly common and unfortunately distinctive as academic achievement within such schools tends to be poorer than schools that are more racially integrated.

Our analysis reveals several notable findings. Overall, we find very little evidence that the self-affirmation exercise enhances students' grades or their standardized test scores within these school contexts. Scores on standardized tests in Reading and Algebra as well as English Spring semester grades did not, for the most part, differ significantly between students who in the treatment condition (i.e., received the self-affirmation exercise) and those in the control (i.e., received a modified version with no affirming content). However, there were some exceptions to this

Table 4 Regression models predicting STAAR Algebra scores

	Model 1	Model 2	Model 3	Model 4
Treatment	−14.03 (35.17)	−25.69 (33.48)	−10.99 (123.55)	−51.94 (63.54)
Student race (ref = white)				
Black		−195.38* (78.43)	−169.98⁺ (102.68)	−193.74* (78.44)
Hispanic		−174.74* (74.87)	−173.89⁺ (96.75)	−175.72* (74.89)
Other		6.78 (147.86)	−74.99 (199.82)	12.22 (147.99)
Student gender (ref = male)				
Female		−12.32 (33.47)	−10.83 (33.59)	−13.02 (33.52)
Economic status (ref = not disadvantaged)				
Eligible for free or reduced lunch		61.55 (48.64)	61.70 (48.75)	64.86 (48.94)
Live in poverty		−36.80 (48.72)	−38.88 (48.92)	−35.83 (48.78)
Current LEP status		−68.35 (61.39)	−69.78 (61.54)	−70.27 (61.44)
School type (ref = mixed)				
Majority Hispanic		−82.84⁺ (48.90)	−80.52 (49.07)	−114.83⁺ (63.65)
Majority black		−313.72*** (63.45)	−313.34*** (63.58)	−289.10*** (80.53)
Treatment × student race				
Treatment × black			−53.40 (139.52)	
Treatment × Hispanic			−5.05 (130.39)	
Treatment × other			183.05 (297.94)	
Treatment × school type				
Treatment × majority Hispanic				65.17 (77.92)
Treatment × majority black				−64.17 (105.19)
Adjusted R^2	0.00	0.10	0.10	0.10
Total	583	583	583	583

LEP Limited English proficient

Standard errors are included in the parenthesis

Asterisks indicate significance according to a two-tailed test (⁺ $p<.10$, * $p<.05$, ** $p<.01$, *** $p<.001$)

pattern of null findings. Specifically, we uncovered a positive interaction between receiving the treatment and presence in a majority Hispanic school when assessing reading scores (see Table 3, Model 4). These findings are somewhat tenuous, as we uncover no differences among Hispanic students across different school types (mixed vs. majority Hispanic). Additionally, while the interaction tests do not reveal that the treatment's effect varied by race, we do find some evidence that exposure to the treatment significantly impacted the performance of black students on standardized tests (see Table 5). Notably, this occurred in the opposite direction than expected as treatment students earned lower scores than their peers in the control. Ancillary analyses (not shown, available from authors by request) reveal that the pattern is driven by those in the mixed school, but is not apparent for those attending the "majority black" school. Taken together, the treatment's effectiveness is still unclear within these school contexts.

There are many important implications of these patterns. The first is reconciling the relatively dramatic patterns identifed by Cohen et al. (2006) with the largely null associations that we uncover. While Cohen et al. (2006)

identified dramatic improvement in GPA, we note little changes in grades or test scores following the intervention. Our study is not the first to uncover a lack of association between this intervention and student achievement (see Hanselman et al. 2014). Our work runs parallel to other studies revealing this intervention was considerably less successful in schools where students had fewer opportunities for contact with racially different peers, and specifically, where minority students did not encounter white students as often as in other schools (Hanselman et al. 2014). These patterns potentially point to differences in the operation and experience of stereotype threat or other identity threats within segregated contexts. When contexts are both racially homogenous and dominated by racial minority members, they are less likely (or wholly unlikely) to provide the type of inter-group contact that triggers the sense that a minority student is being judged by outsider stereotypes.

Notably, we also find that patterns of black and Hispanic students diverge as black student achievement was *negatively impacted* by the treatment and provided fewer clear results for Hispanic students. This challenges the initial

Table 5 Regression models predicting Spring Semester English grades, STAAR Reading, and STAAR Algebra scores for black students

	Model 1	Model 2	Model 3
Panel A. Semester English grade			
Treatment	2.48 (1.61)	1.46 (1.40)	−0.38 (2.04)
School type (ref= mixed)			
Majority black		−9 79*** (1.53)	−11 46*** (2.03)
Student gender (ref = male)			
Female		3.03* (1.41)	3.10* (1.41)
Economic status (ref = not disadvantaged)			
Eligible for free or reduced lunch		1.75 (1.92)	1.60 (1.92)
Live in poverty		(−0.82) (1.85)	−0.96 1.85)
Treatment × school type			
Treatment × majority black			3.46 (2.78)
Adjusted R^2	0.01	0.27	0.27
Total	179	179	179
Panel B. STAAR Reading			
Treatment	−22.34 (37.27)	−54.62[+] (28.71)	−97.80* (41.98)
School type (ref= mixed)			
Majority black		−297.92*** (31.55)	−336.03*** (41.52)
Student gender (ref = male)			
Female		53.70[+] (28.72)	55.61[+] (28.67)
Economic status (ref = not disadvantaged)			
Eligible for free or reduced lunch		21.89 (39.68)	18.83 (39.63)
Live in poverty		−37.97 (38.18)	−41.15 (38.14)
Treatment × school type			
Treatment × majority black			80.09 (56.93)
Adjusted R^2	0.00	0.42	0.42
Total	182	182	182
Panel C. STAAR Algebra			
Treatment	−84.70 (74.17)	−122.71[+] (67.16)	−171.83 (####)
School type (ref= mixed)			
Majority black		−393.89*** (76.79)	−431.74*** (####)
Student gender (ref = male)			
Female		80.71 (67.26)	79.49 (67.47)
Economic status (ref = not disadvantaged)			
Eligible for free or reduced lunch		178.26[+] (96.49)	182.61 (97.05)
Live in poverty		111.39 86.51)	110.57 (86.75)
Treatment × school type			
Treatment × majority black			78.12 (####)
Adjusted R^2	0.00	0.20	0.20
Total	136	136	136

LEP Limited English proficient

Standard errors are included in the parenthesis

Asterisks indicate significance according to a two-tailed test ([+] $p<0.10$, * $p<.05$, ** $p<.01$, *** $p<.001$)

theorizing that suggests that all *minority* groups encounter and are similarly impacted by stereotype threat or similarly benefited by self-affirmation. Recent debates about the changing nature of racial stratification assert that difference from whites should be unpacked further as racialized disadvantages are often more acute for African Americans compared to other groups (Lee and Bean 2010) and that Latinos' experiences are strongly shaped by skin color, generation, and nativity (Faught and Hunter 2012). More research is needed to identify interventions that are

Table 6 Regression models predicting Spring Semester English grades, STAAR Reading, and STAAR Algebra scores for Hispanic students

	Model 1	Model 2	Model 3
Panel A. Semester English grade			
Treatment	−0.28 (1.15)	−0.38 (1.09)	0.15 (2.34)
School type (ref= mixed)			
Majority Hispanic		−6.80*** (1.37)	−6.44** (1.97)
Student gender (ref = male)			
Female		3.79** (1.10)	3.81** (1.10)
Economic status (ref = not disadvantaged)			
Eligible for free or reduced lunch		1.82 (1.56)	1.79 (1.56)
Live in poverty		−3.01[+] (1.66)	−3.04[+] (1.66)
Treatment × school type			
Treatment × majority Hispanic			−0.67 (2.65)
Adjusted R^2	0.00	0.09	0.09
Total	484	484	484
Panel B. STAAR Reading			
Treatment	−21.51 (18.92)	−19.10 (17.58)	−54.25 (37.30)
School type (ref= mixed)			
Majority Hispanic		−160.14*** (22.14)	−183.70*** (31.25)
Student gender (ref = male)			
Female		18.25 (17.58)	17.01 (17.62)
Economic status (ref = not disadvantaged)			
Eligible for free or reduced lunch		−30.08 (25.68)	−28.36 (25.73)
Live in poverty		−87.13** (27.16)	−85.56** (27.20)
Treatment × school type			
Treatment × majority Hispanic			45.10 (42.21)
Adjusted R^2	0.00	0.14	0.14
Total	479	479	479
Panel C. STAAR Algebra			
Treatment	−9.15 (42.53)	−11.01 (42.32)	−70.44 (101.49)
School type (ref= mixed)			
Majority Hispanic		−97.70[+] (57.30)	−137.83 (84.66)
Student gender (ref = male)			
Female		−33.39 (42.47)	−35.00 (42.57)
Economic status (ref = not disadvantaged)			
Eligible for free or reduced lunch		88.84 (64.76)	95.50 (65.64)
Live in poverty		−15.20 (66.19)	−9.91 (66.75)
Treatment × school type			
Treatment × majority Hispanic			72.16 (111.98)
Adjusted R^2	0.00	0.01	0.01
Total	367	367	367

LEP Limited English proficient

Standard errors are included in the parenthesis

Asterisks indicate significance according to a two-tailed test ([+] $p<0.10$, * $p<.05$, ** $p<.01$, *** $p<.001$)

effective for black and Hispanic students. We note here, however, that these schools differ on more dimensions that merely their race/ethnic composition and in more ways than may be captured in our models. Future work needs to examine more carefully the presumption that these two groups (or other race/ethnic groups) may experience threat in different ways, potentially impacting the means through it can be relieved.

The current analysis has a few limitations. First, many concepts go unmeasured that have bearing on this relationship. We do not, for example, directly test for the existence of "threat". Stereotype threat is often identified after its been cued up, and we present no such trigger for these students. However, we operate under the notion that there is a "threat in the air" (Steele and Aronson 1995) suggesting that all persons classified as members of minority groups are susceptible to feeling as if they are being judged in any moment when they are evaluated. We accept this as an operating premise, however, it is unclear how threat may emerge, or how it might inhibit achievement, in contexts where children of color are the large majority. Enhancing self-integrity through the exercise may not be able to counter threat if none is present. Prior analyses gauge the level of race consciousness through fielding a survey on racial climate, a feature that is absent in our intervention. Our aim is to introduce a scenario that can be replicable at a large scale and fielding such a survey repeatedly may not be feasible. However, the result to exclude the racial climate survey may have removed the sensitivity to race that may have underlined the experience of stereotype threat in other assessments. Future research might explore the impact of the presence or absence of a climate survey for these interventions.

Additionally, our analyses do not adjust for academic ability of students. Theoretically, these exercises should be efficacious regardless of student academic ability (Cohen et al. 2006), however, prior academic achievement may drive or suppress the effectiveness of the affirmation. Ancillary analyses presented in Appendix Table 8 reveal that adjusting for performance on a standardized examination taken prior to entering ninth grade called the Stanford 10, does not alter the effect of the treatment, though this score is strongly related to academic performance a year later (at the close of ninth grade).

We make a few suggestions for future research. Future work should aim to expand the assessment of self-affirmation by exploring some of the qualitative dimensions. When writing about their personal experiences, students' reflections, either positive or negative, may be critical to the power the affirmation exercise may have. We explored this preliminarily by coding each response to the open-ended portion of the exercise, regardless of the condition, for their use of positive or life-affirming content, as opposed to neutral or negative content (i.e., conveying unhappiness or some degree of emotional upset). The analyses shown in Appendix Table 7 compared students in the treatment who reported affirming responses to those in the control who used neutral content in their exercises. In all, we still find few differences in academic performance between these two groups; however, analyses of students reporting negative content (not shown) suggest this could

be detrimental to their academic performance.[5] These patterns warrant further exploration as they suggest that the "power" of self-affirmation may be conditional on whether students actually find it to be affirming and, more importantly, that there may be costs for students for whom reflecting on their values is an upsetting experience.

We submit that the null findings strongly suggest that academic challenges faced by many of these students are complex, shaping the potential of enhancing self-integrity for ameliorating gaps in school performance. Notably, these schools differ in more ways that just race/ethnic composition and in ways that outstrip our adjustment in our models. Our attempt to adjust for socioeconomic status between students, for example, draws on Free-Reduced Lunch (FRL) status. While researchers frequently use FRL (Hoffman 2012), this measure has noted flaws (Harwell and LeBeau 2010). Moreover, our results reveal improvement in standardized test scores, but with less clear impact on grades, suggesting that these aspects of academic performance may respond to the influence self-affirmation differently. The academic challenges of students in segregated schools (often from high poverty neighborhoods) are tremendous, and research is still unraveling its complex nature. Many have argued that segregation in today's schools is reaching pre-1960s levels (Frankenberg and Lee 2002; Orfield et al. 2014, 2012). However, what this means for identity threats remains unclear, suggesting a variety of tools are required to address these experiences.

---

[5] We coded each exercise, regardless of the condition, for their use of positive or life-affirming content, as opposed to neutral or negative content. Our analyses incorporate the frequency of these codes across all four administrations, distinguishing the following categories: those students in the treatment condition that provided affirming responses across all four exercises (hereafter affirming treatment), students in the control condition who provided neutral responses across all four exercises, and all the remaining students who provided both positive and negative responses. Volunteer undergraduate student readers were trained with examples from respondents' actual responses to code the exercises as affirmative, neutral, or negative. Each exercise was coded by two readers with high inter-rater reliability.

Appendix

See Tables 7 and 8.

Table 7 Regression models predicting Spring Semester English grade, STAAR Reading, and STAAR Math scores for students providing affirming responses on treatment or neutral responses on control

	Semester English grade	STAAR Reading	STAAR Math
Affirming treatment (ref = neutral control)	−0.75 (0.97)	−5.79 (20.07)	32.54 (47.27)
Student race (ref = white)			
Black	−5.05** (1.92)	−10.95 (39.90)	−109.16 (106.7)
Hispanic	−3.05+ (1.79)	−40.80 (37.29)	−133.48 (102.6)
Other	−2.45 (4.03)	−78.40 (83.57)	−149.37 (205.4)
Student gender (ref = male)			
Female	3.323** (0.96)	18.21 (19.98)	−33.09 (46.81)
Economic status (ref = not disadvantaged)			
Eligible for free or reduced lunch	−3.18* (1.40)	−62.69* (29.04)	2.14 (70.98)
Live in poverty	−5.13*** (1.44)	−109.07*** (29.85)	−62.63 (69.74)
Current LEP status	−1.57 (1.92)	−174.28*** (39.84)	−106.86 (80.92)
School type (ref = mixed)			
Majority Hispanic	−5.50*** (1.31)	−162.49*** (27.07)	−99.46 (66.75)
Majority black	−6.46** (1.86)	−282.46*** (38.79)	263.34** (90.88)
Adjusted R^2	0.24	0.35	0.06
Total	379	378	261

Responses refer to content conveyed in the open−ended section of the treatment or control exercise. Affirming refers to positive content related to the values reported. Neutral refers to content reflecting neither positive nor negative sentiment in the open−ended component

LEP Limited English proficient

Standard errors are included in the parenthesis

Asterisks indicate significance according to a two-tailed test (+ $p<0.10$, * $p<.05$, ** $p<.01$, *** $p<.001$)

Table 8 Regression models predicting Spring Semester English grade, STAAR Reading, and STAAR Math scores, controlling for performance on Stanford 10

	Semester English grade	STAAR Reading	STAAR Math
Treatment	0.29 (0.73)	−14.50 (10.52)	−16.37 (28.31)
Student race (ref = white)			
Black	2.73+ (1.53)	14.34 (21.97)	−25.53 (66.67)
Hispanic	−2.75+ (1.48)	−4.18 (21.15)	−70.31 (63.34)
Other	−4.15 (3.26)	3.39 (47.19)	23.84 (124.91)
Student gender (ref = male)			
Female	2.69*** (0.73)	24.39* (10.48)	19.55 (28.43)
Economic status (ref = not disadvantaged)			
Eligible for free or reduced lunch	0.30 (1.06)	3.83 (15.36)	28.84 (42.29)
Live in poverty	−2.77* (1.11)	−39.47* (15.99)	−50.35 (42.23)
Current LEP status	3.85* (1.57)	−38.34+ (22.44)	−5.90 (53.49)
School type (ref = mixed):			
Majority Hispanic	−3.44** (1.05)	−57.50*** (15.14)	17.12 (40.69)
Majority black	−4.47** (1.45)	−138.70*** (20.46)	−140.37* (55.33)
Stanford 10	0.136*** (0.01)	4.99*** (0.20)	8.54*** (0.48)

Table 8 continued

	Semester English grade	STAAR Reading	STAAR Math
Adjusted R^2	0.30	0.65	
Total	708	722	

The Stanford 1 exam is a standardized test offered in eighth grade. This covariate aims to adjust for prior academic performance

LEP Limited English proficient

Standard errors are included in the parenthesis

Asterisks indicate significance according to a two-tailed test ($^+$ $p<0.10$, * $p<.05$, ** $p<.01$, *** $p<.001$)

References

Berends, M., & Penaloza, R. (2010). Increasing racial isolation and test score gaps in mathematics: A 30-year perspective. *The Teachers College Record, 112*(4), 978–1007.

Bettie, J. (2003). *Women without class: Girls, race, and identity*. Berkeley, CA: University of California Press.

Borman, G. (2012). Examination of a self-affirmation intervention in St. Paul Public Schools. In *The senior urban education research fellowship series* (Vol. 9, pp. 1–44). The Council of the Great City Schools.

Bowen, N. K., Wegmann, K. M., & Webber, K. C. (2013). Enhancing a brief writing intervention to combat stereotype threat among middle-school students. *Journal of Educational Psychology, 105*(2), 427–435. doi:10.1037/a0031177.

Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on raven's advanced progressive matrices. *Journal of Applied Psychology, 91*(4), 979.

Cohen, G. L., & Garcia, J. (2005). "I am us": Negative stereotypes as collective threats. *Journal of Personality and Social Psychology, 89*(4), 566–582. doi:10.1037/0022-3514.89.4.566.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*(5791), 1307–1310.

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science, 324*(5925), 400–403. doi:10.1126/science.1170769.

Coley, R. J. (2002). *An uneven start: Indicators of inequality in school readiness*. Princeton, NJ: Educational Testing Service, Policy Information Center.

Condron, D. J., Tope, D., Steidl, C. R., & Freeman, K. J. (2013). Racial segregation and the black/white achievement gap, 1992 to 2009. *The Sociological Quarterly, 54*(1), 130–157.

Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness, 8*(2), 149–168.

Downey, D. B. (2008). Black/white differences in school performance: The oppositional culture explanation. *Annual Review of Sociology, 34*, 107–126. doi:10.1146/annurev.soc.34.040507.134635.

Faught, J., & Hunter, M. (2012). Latinos and he skin color paradox: Skin color, national origin, and political attitudes. *The Sociological Quarterly*, *53*, 676–701.

Frankenberg, E., & Lee, C. (2002). *Race in American public schools: Rapidly resegregating school districts*. Cambridge, MA: The Civil Rights Project, Harvard University.

Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal-Applied Economics, 2*(2), 210–240. doi:10.1257/app.2.2.210.

Hanselman, P., Bruch, S. K., Gamoran, A., & Borman, G. D. (2014). Threat in context school moderation of the impact of social identity threat on racial/ethnic achievement gaps. *Sociology of Education, 87*(2), 106–124.

Harwell, M., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. *Educational researcher, 39*(2), 120–131.

Hoffman, L. (2012). Free and Reduced-Price Lunch Eligibility Data in Ed" Facts": A White Paper on Current Status and Potential Changes. *Office of Planning, Evaluation and Policy Development, US Department of Education*.

Kellow, J. Thomas, & Jones, B. D. (2008). The effects of stereotypes on the achievement gap: Reexamining the academic performance of African American high school students. *Journal of Black Psychology, 34*(1), 94–120. doi:10.1177/0095798407310537.

Lee, J., & Bean, F. (2010). *The diversity paradox: Immigration and the color line in the 21st century*. New York City: Russell Sage Foundation.

Lee, J. C., & Klugman, J. (2013). Latino school concentration and academic performance among latino children. *Social Science Quarterly, 94*(4), 977–1015.

Lewis, A. (2003). *Race in the schoolyard: Negotiating the color line in classrooms and communities*. New Brunswick, NJ: Rutgers University Press.

Lichter, D. T. (2013). Integration or fragmentation? Racial diversity and the American future. *Demography, 50*(2), 359–391.

Marchand, G. C., & Taasoobshirazi, G. (2013). Stereotype threat and women's performance in physics. *International Journal of Science Education, 35*(18), 3050–3061.

Martiny, S. E., Roth, J., Jelenec, P., Steffens, M. C., & Croizet, J. C. (2012). When a new group identity does harm on the spot: Stereotype threat in newly created groups. *European Journal of Social Psychology, 42*(1), 65–71. doi:10.1002/ejsp.840.

Massey, D. S., Charles, C. Z., Lundy, G., & Fischer, M. J. (2011). *The source of the river: The social origins of freshmen at America's selective colleges and universities*. Princeton: Princeton University Press.

McKay, P. F., Doverspike, D., Bowen-Hilton, D., & Martin, Q. D. (2002). Stereotype threat effects on the raven advanced progressive matrices scores of African Americans. *Journal of Applied Social Psychology, 32*(4), 767–787.

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development, 74*(2), 498–515. doi:10.1111/1467-8624.7402012.

Morris, E. W. (2006). *An unexpected minority: White kids in an urban school*. New Brunswick: Rutgers University Press.

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*(6), 1314–1334. doi:10.1037/a0012702.

Orfield, G., Frankenberg, E., Ee, J., & Kuscera, J. (2014). Great progress, a long retreat and an uncertain future. Retrieved October 5, 2014, from http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/brown-at-60-great-pro

gress-a-long-retreat-and-an-uncertain-future/Brown-at-60-0518 14.pdf.

Orfield, G., Kucsera, J., & Siegel-Hawley, G. (2012). E pluribus… separation: Deepening double segregation for more students. Civil Rights Project at UCLA.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology, 26*(3), 291–310. doi:10.1006/ceps. 2000.1052.

Passel, J. S. (2011). Demography of immigrant youth: Past, present, and future. *The Future of Children, 21*(1), 19–41.

Perry, P. (2002). *Shades of white: White kids and racial identities in high school*. Durham: Duke University Press.

Phillips, M. (2000). Understanding ethnic differences in academic achievement: Empirical lessons from national data. In D. W. Grissmer & J. M. Ross (Eds.), *Analytic issues in the assessment of student achievement* (pp. 103–132). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Purdie-Vaughns, V., Cohen, G.L., Garcia, J., Sumner, R., Cook, J. C., & Apfel, N. (2009). Improving minority academic performance: How a values-affirmation intervention works. *Teachers College Record*.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440.

Sherman, D. K. (2013). Self-affirmation: understanding the effects. *Social and Personality Psychology Compass, 7*(11), 834–845.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797.

Texas Education Agency. (2012). Reports and data. Retrieved February 1, 2015, from http://tea.texas.gov/.

Tyson, K., Darity, W., & Castellino, D. R. (2005). It's not "a black thing": Understanding the burden of acting white and other dilemmas of high achievement. *American Sociological Review, 70*(4), 582–605.

Walton, G. M., & Spencer, S. J. (2009). Latent ability grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20*(9), 1132–1139.

Wasserberg, M. J. (2014). Stereotype threat effects on african american children in an urban elementary school. *Journal of Experimental Education, 82*(4), 502–517. doi:10.1080/ 00220973.2013.876224.