Accepted Manuscript

Affirmation effects on math scores: The importance of high school track

Amanda Bancroft, Jenifer Bratter, Kristie Rowley

PII: S0049-089X(16)30627-5

DOI: 10.1016/j.ssresearch.2016.10.001

Reference: YSSRE 2002

To appear in: *Social Science Research*

Received Date: 24 June 2015

Revised Date: 30 June 2016

Accepted Date: 5 October 2016

Please cite this article as: Bancroft, A., Bratter, J., Rowley, K., Affirmation effects on math scores: The importance of high school track, *Social Science Research* (2016), doi: 10.1016/j.ssresearch.2016.10.001.

Affirmation Effects on Math Scores: The Importance of High School Track

Amanda Bancroft[a1,*]

Jenifer Bratter[a2]

Kristie Rowley[b]

[a]Rice University
Department of Sociology, MS-28
P.O. Box 1892
Houston, TX 77251

[b]Brigham Young University
2008 JFSB
Provo, UT 84602

*Corresponding author: Email: abancroft@rice.edu

1. INTRODUCTION

Social scientists have observed stereotype threat – the phenomenon of as “being at risk of confirming, as self-characteristic, a negative stereotype about one's group,” – in academic settings among racial minorities and among women and girls, who are threatened by stereotypes about academic intelligence and quantitative skills, respectively (Steele and Aronson 1995). Among interventions aimed at stereotype threat reduction are a series of short writing exercises, or “affirmations,” designed to encourage students to write important personal values, thereby affirming self-worth and academic self-concept. Studies suggest that improved self-concept arising from participation in these exercises can improve academic performance for students experiencing the negative effects of stereotypes (Cohen Garcia, Apfel, and Master 2006; Cohen, Garcia, Purdie-Vaughns, Apfel, and Brzustoski 2009; Miyake, Kost-Smith, Finkelstein, Pollock, Cohen, and Ito 2010; Wisconsin Center for Education Research 2013; Hanselman, Bruch, Gamoran, and Borman 2014). Giving such interventions to students throughout the school year is thought to lead to help them “gain relatively ‘quick wins.’ As these accumulate in a recursive process, like a chain reaction, the gains are carried forward” (Wisconsin Center for Education Research 2014).

During the 2012-2013 school year, the Houston Education Research Consortium (HERC[1]) implemented affirmation interventions in three local public high schools to reduce the effects of stereotype threat among ninth grade students and narrow existing racial achievement gaps between whites and African Americans, and whites and Hispanics, which have been the subject of much investigation in Houston and elsewhere (Steele and Aronson 1995; Cohen and Garcia 2005; Kelly 2009; Wisconsin Center for Education Research 2013; HERC 2013;

[1] The Houston Education Research Consortium (HERC) is a partnership between Rice University and the Houston Independent School District (HISD) which aims to produce rigorous research for the purpose of closing the socioeconomic gaps in educational achievement and attainment in Houston.

Hanselman, et al. 2014).

The present study takes HERC's aims one step further, focusing on outcomes for minority[2] girls, who experience the burden of both race and gender – a "double bind" that has been linked to educational and professional underrepresentation of women in STEM, or science, technology, engineering, and math (Malcom, Hall, and Brown 1976; Collins and Matyas 1985; Brown 1994; Brown 1995; Gonzales, Blanton, and Williams 2002; Tate and Linn 2005; Charles and Bradley 2009; Mikyake et al. 2010; Malcom and Malcom 2011; Ong, Wright, Espinosa, and Orfield 2011; Espinosa 2011; Reyes 2011; National Science Foundation 2011). The project looks at stereotype threat at this intersection of race and gender, investingating the effectiveness of interventions in alleviating threat among minority girls and improving standardized math test scores, as math is considered a substantial barrier to entry into STEM education and employment (Sells 1973; Charles and Bradley 2009). While several environmental, institutional, and other features of the STEM pipeline have been found to affect STEM persistence over time, a focus on in-class intervention is one way that school districts can positively impact STEM success while using very few resources to do so.

2. LITERATURE REVIEW

Achievement gaps persist between racial groups in overall academic achievement, and between males and females in math achievement (Steele and Aronson 1995; Good, Aronson, and Inzlicht 2003; Kelly 2009; College Board 2012; Department of Research and Accountability, HISD 2012, 2013; Texas Education Agency 2012), although in some cases, the latter gap appears to be narrowing and even reversing (Department of Research and Accountability, HISD 2013). Some research has attributed these gaps to stereotype threat by demonstrating the ways in

[2] The term "minority" will be used hereafter to identify African American and Hispanic students.

which threat interferes with learning and academic performance and produces patterns of lowered expectations and low self-esteem, reduced high-level course enrollment, lowered involvement in or association with academic activity, disassociation with one's racial group, and disassociation with and withdrawal from academic pursuits (Steele and Aronson 1995; Major, Spencer, Schmader, Wolfe, and Crocker 1998; Cohen and Garcia 2005). One conceptual model of the "social identity" threat pathway, according to Hanselman et al. (2014) is that social contexts, individual characteristics, and a given task domain, create the conditions for the academic salience of a particular identity, which leads to threat (2014:108). Threat, in turn, leads to various responses (e.g., stress, decreased working memory) which deplete students' pool of resources that may be allocated to academic task, which decreases performance on those tasks and contributes to the existing negative self-concept (Hanselman et al. 2014:107). The particular benefit of stereotype threat interventions (examples of which will be discussed below) is that they are delivered within the academic task domain (e.g., standardized testing), in an attempt to moderate the effect of threat and its related response(s) in a student.

In the absence of such interventions, negative self-concepts and patterns of low performance separate low-achieving from high-achieving students (Wilson and Linville 1985; Major et al. 1998; Cohen et al. 2006; Cohen et al. 2009), and students' attribution of poor performance to inherent and inflexible group or personal characteristics contributes to this bifurcation (Cech and Blair-Loy 2010; Espinosa 2011). Even at the elementary school level, achievement gaps affect student self-concepts and outcomes – effects that reverberate into high school, university, and career (Alexander, Entwisle, and Horsey 1997; Huguet and Regner 2007; Burrelli 2009; Miyake et al. 2010; Cech and Blair-Loy 2010; Ong et al. 2011). For girls and women specifically, these effects can limit entrance into higher-level math education, a critical pathway to STEM (Sells 1973; Ayalon 1995; Shapka and Keating 2003; Charles and Bradley 2009).

Exact mechanisms are the cause of some debate, and different branches of stereotype threat research note that social, psychological, cognitive, and physiological factors can play a role when stereotyped identities are made salient, or when ability (based on identity) is evaluated. Increased anxiety and stress system arousal, reductions in or sacrifice of working memory, lowered self-regulation, and negative thinking and low self-esteem have each been shown to play a role in the experience of threat (Steele and Aronson 1995; Blascovich, Spencer, Quinn, and Steele 2001; Schmader and Johns 2003; Cohen and Garcia 2005; Beilock, Jellison, Rydell, McConnell, and Carr 2006; Inzlicht et al. 2006; Osborne 2007; Beilock, Rydell, and McConnell 2007; Krendl et al., 2008). Some argue that stereotype threat is a complex phenomenon, while others contend that multiple types exist (Shapiro and Neuberg 2007).

While the effects of stereotype threat have often been analyzed for race or gender in isolation (Steele Aronson 1995; Osborne 2001; Schmader and Johns 2003; Good et al. 2003; Cohen et al. 2006; Keifer and Sekaquaptewa 2007; Huguet and Regner 2007; Good, Aronson, & Harder 2008; Krendl, Richeson, Kelley, and Heatherton 2008; Miyake et al. 2010), there is little work on stereotype threat at the intersection of gender and race (Cheryan and Bodenhausen 2000; Gonzales et al. 2002) – surprising, given the ubiquity and high profile of the research and government policy work surrounding minority and female involvement in STEM (Malcom et al. 1976; Brown 1994, 1995; Ong 2002; Gonzales et al. 2002; Tate and Linn 2005; Carolene and Johnson 2007; Obama 2009; Ong et al. 2011; Espinosa 2011; Jones Taylor and Walton 2011; National Science Foundation 2011; Malcom and Malcom 2011). This project aims to occupy that space, supported by evidence that stereotype threat's role in achievement gaps may be manipulated through intervention techniques (Good et al. 2003; Cohen et al. 2006; Cohen et al. 2009; Miyake et al. 2010).

2.1. Race, Gender, Race x Gender: Patterns of Threat

Many have theorized that minority students' fears of furthering racial stereotypes directly interfere with cognitive functioning, and that a stereotype's salience before or during an intellectual exercise is enough to arouse those fears. Similar triggers activate stereotype threat in minorities, as well as in girls and women. In experimental settings, students for whom stereotypes are "cued" in certain ways – i.e., who are told that tests or activities are diagnostic of ability, are told that tests are biased, or are asked to record their race or gender before assessment – perform lower, relative to students for whom these stereotypes are not made salient (Shih, Pittinsky, and Ambady 1999; Shih et al. 1999; Spencer, Steele, and Quinn 1999; Gonzales et al. 2002; Schmader and Johns 2003; Gresky, Ten Eyck, Lord, and McIntyre 2005; Cohen and Garcia 2005; Kellow and Jones 2005; Inzlicht, McKay, and Aronson 2006; Good, Aronson, and Harder 2008; Krendl et al. 2008). Gonzales et al. (2002), one of few studies to evaluate stereotype threat among minority women, tested for ethnicity- and gender-based effects of threatand found a "double-minority effect" for Latinas, who performed lower than white women, white men, and Latino men when the double-effect was cued (Gonzales et al. 2002:668).

Race and gender stereotypes are both sustained and reproduced by such scenarios, as beliefs about stereotypes, and associated behaviors, constantly reinforce one another. Prior work notes that repeated, long-term exposure to threat creates a cyclical, compounded effect and results in outcomes like underachievement (Wilson and Linville 1985; Major et al. 1998). Even when women progress to further levels of math and science achievement study, or when minority students excel academically, stereotypes seem to work powerfully on the psychological and cognitive processes of the individuals associated with them (Good, Aronson, and Harder 2008; Cech and Blair-Loy 2010; Miyake et al. 2010).

2.2. Past Interventions

Several successful intervention techniques have been implemented to help combat manifestations of negative academic self-concept. Some studies have utilized attributional techniques, which help students explain their own educational difficulties by attributing failure and success to factors outside of intrinsic ability (e.g., to techniques such as effort, strategy, or practice), or by teaching students about how academic aptitude is amenable to improvement (Aronson, Fried, and Good 2002; Good et al. 2003; Dweck 1999). Other studies have shown that creating identity maps that highlight positive personal attributes, or identifying achievements of successful people who share certain stereotyped identities can positively affect performance (Gresky et al. 2005; McIntyre et al. 2003; Cohen and Garcia 2005). Some work has also shown that guaranteeing participants that a test is unbiased, or that it works in their favor, can also be effective (McIntyre, Paulson, and Lord 2003).

Self-affirmation techniques have been highly successful in alleviating stereotype threat specifically (Schimel, Arndt, Banko, and Cook 2004; Sherman and Cohen 2006; Martens, Johns, Greenberg, and Schimel 2006). The projects after which the HERC (2013) experiment is modeled attempted to bolster students' positive feelings about themselves and buffer the potential negative impact of stereotypes. Studies did this through the use of written exercises, prompting middle school students (Cohen et al. 2006; Cohen et al. 2009; Wisconsin Center for Education Research 2013; Hanselman et al. 2014), or women in college STEM majors (Miyake et al. 2010), to reflect on personal values in order to cue up more positive (less stereotype-centered) aspects of self-identity.

In previous research, writing exercises were designed to allow students the opportunity to affirm self-worth and self-integrity, which research suggests are deteriorated under repeated exposure to stressful situations like stereotype threat (Steele and Aronson 1995; Major et al. 1998; Sherman and Cohen 2006). Typically, students in the treatment or “affirmation” condition write about the values most important to them, prompting positive self-understanding and combating stereotypes. Those in the control condition write on an impersonal argumentative topic, giving control students an opportunity to gain writing practice while prompting an overall neutral effect (Cohen et al. 2006; Cohen et al. 2009; Miyake et al. 2010). The exercises are distributed at multiple points throughout the school year as a casual, ungraded free-writing activity, and are not framed by teachers as indicative of ability, or as related to gender or racial differences – conditions under which stereotype threat has been found to thrive.

[Insert Figures 1 and 2 about here]

Cohen et al.’s (2006, 2009) findings indicate that interventions interrupted potential downward trends, protecting African American students against early declines in academic performance and maintaining that protective effect over time. Specifically, interventions improved GPAs among African American students in the affirmation condition and led to a 40% reduction in the black-white racial achievement gap that academic year, while African Americans in the control group saw a downward trend in grades (Cohen et al. 2006). In a two-year follow up to their 2006 study, Cohen et al. (2009) found further indication of positive intervention effects. Namely, GPAs of African American students in the treatment group rose an average of 0.24 points, while low-achieving African Americans saw an even more striking mean increase of 0.41 points. The authors also discovered a positive increase in students’ involvement in advanced placement (AP) math courses, and noted upward-trending overall academic achievement for African American students in the affirmation group throughout the entirety of their middle school experience (2009).

In their study targeting women in STEM majors, Miyake et al. (2010) found that their own intervention narrowed gender gaps on the Force and Motion Conceptual Evaluation (a standardized physics exam), and that women in the affirmation group saw an average increase from C to B grades. Effects were particularly beneficial for those who had previously expressed support for gender stereotypes about science ability. The Wisconsin Center for Education Research study found that “Self-affirmation writing exercises improved the achievement, especially in mathematics, of students who may suffer from stereotype threat” (Wisconsin Center for Education Research 2014).

It appears that even small-scale interventions can have significant effects on stereotype threat, and thus on student achievement, by disrupting downward trends – especially for students most at risk. The HERC (2013) attempted to replicate this strategy by administering similar interventions at four points throughout the school year. The goal of a repetitive intervention cycle is to engage students in a positive self-affirming activity early in their high school careers before potential downward academic trends materialize, allowing them to revisit the activity throughout the year and reaffirm positive self-understandings without the threat of performing for assessment, or to attain a class grade. Borman and colleagues argue that “writing helps students gain relatively ‘quick wins.’ As these accumulate in a recursive process, like a chain reaction, the gains are carried forward” (Wisconsin Center for Education Research 2014). The HERC project implemented such interventions with the hope that their cost-effectiveness and easy replicability could have the potential to narrow racial and socioeconomic achievement gaps, in Houston and elsewhere.

3. HYPOTHESES, DATA AND METHODS

3.1. Hypotheses

The HERC (2013) study, like the others listed above, aims to narrow achievement gaps through “affirmation” interventions. The HERC project is novel in that it reached high school students – a group that has not yet been studied within this specific affirmation/intervention framework. Additionally, the current extension of the HERC project focuses on the understudied intersection of race and gender.

Net of economic disadvantage, academic track, and specific school environment, writing exercises will have a more positive effect on the standardized math test scores of:

- H1: all students in the treatment group, compared to their peers in the control group;
- H2: girls in the treatment group, compared to girls in the control group;
- H3: Hispanics in the treatment, compared to Hispanics in the control group; and
- H4: minority girls in the treatment, compared to minority girls in the control group.

This final comparison (H4) allows us to better understand the efficacy of the writing exercises for narrowing the raced and gendered achievement gaps in math that doubly disadvantage minority girls and constrain their entry into further math education and future STEM career participation.

3.2. Data and Methods

This study is a randomized controlled trial of an affirmation intervention on student standardized math performance. The data and the analysis of the proposed effects are cross-sectional at two discreet, but cumulative, points in time (at which successive interventions were implemented). Students at three high schools were randomized then placed in treatment and control groups, and received intervention (or a value-free writing exercise) based on their group.

Selection of the schools and the timeline and other details of implementation are discussed in full below.

3.2.a. The Schools

The HERC (2013) study was conducted among ninth grade students enrolled in grade-level English courses in three public high schools in the Houston Independent School District (HISD) in Houston, Texas. Each school is quite different demographically – one school is mostly Hispanic (hereafter HS1), one is mostly African American (HS2), and one has a distribution of roughly 1/3 Hispanic, 1/3 African American, and 1/3 white students (HS3). As the largest achievement gaps in HISD are between white and African American students, and white and Hispanic students, these groups are of highest interest to the project, and schools were chosen to include a representative number of students from each of these groups. In the STAAR Algebra I analyses (wave 4, the only wave in which all three racial groups could be compared), the percentages of students in each group approximate actual district percentages (see Tables 1.2 and 2.2).

[Insert Tables 1.2 and 2.2 about here]

Due to their small, unrepresentative population within the three schools in the sample, students who racially identify outside of the three main categories are not included in this project, although they participated in the broader study. HS3, the only school with a significant proportion of white students, is situated in a wealthy Houston neighborhood, while HS1 and HS2 are located in relatively less affluent areas, but the school still includes a large population of low-income and minority students, as do the other two schools. HS3 was thus included in the sample to capture some white (reference) students in the sample, as very few high schools in the district have a large enough population of white students to be analytically useful. The final sample – or, more precisely, the census of 9th grade students in these three schools – thus includes a racially and somewhat socioeconomically diverse group of students, whose achievement outcomes are assessed in conjunction with student-level data on race and SES received from the school district.

3.2.b. Randomization Procedures

The significance of race in a student's vulnerability to and experience of stereotype threat is of primary concern to the HERC (2013) study. Due to the stark racial demographic differences between HS1, HS2, and HS3, ninth grade students were first stratified by race across all three schools to ensure that each of the three main groups of interest – Hispanic, African American, and white – were represented in the treatment and control groups. Students were then individually randomized within each school, for placement into either the treatment or control group. At HS1, both the control group (CG) and treatment group (TG) are majority Hispanic (CG N = 200 Hispanic/ 13 African American; TG N = 210 Hispanic/ 10 African American). At HS2, both groups are majority African American (CG N = 88 African American/ 17 Hispanic; TG N = 95 African American/ 13 Hispanic). At HS3, there is roughly equal distribution of Hispanic, African American, and white students in both groups (CG N = 109 Hispanic/ 84 African American/ 97 white; TG N = 129 Hispanic/ 84 African American/ 90 white). Student racial totals across all three schools are as follows: Hispanic N = 678; African American N = 374; white N = 190. The total number of students in the population with a racial identification is 1276, plus 69 whose races are unaccounted for in available school records. There is a roughly equal distribution of girls and boys in both the treatment and control groups.[3] Basic descriptive statistics on the dependent variables, for all comparison groups studied in the present report, are listed in the Appendix in Tables 2.3-2.6.

[3] See Tables 2.1 and 2.2 in Appendix for Wave 1 and Wave 4 samples.

3.2.c. Writing Exercises[4]

Writing affirmations in the HERC (2013) study are modeled after the Cohen et al. (2006, 2009). These instruments were used with permission from Dr. Geoffrey Cohen. For the first wave of the HERC study, students in the treatment group received a writing affirmation exercise prompting them to circle two or three items from a list of thirteen values[5] (sequence was randomized to minimize order effects). Values were loosely based on those in the Cohen et al. (2006) study, and were further developed to appeal to young people with a variety of interests. After choosing two or three values, each student was prompted to write about the values and what they mean to him/her, then list the top two reasons the values are important. Exercises noted that students should not worry about spelling and grammar, and that there were no correct or incorrect answers. Later waves allowed students to choose new values, or revisit prior selections.

Students in the control group were first prompted to circle two or three values *not* important to them, in order to separate the positive, or "affirmation" experience of the treatment group from the neutral experience of the control group. The subsequent writing portion included a brief prompt about a particular debate (e.g., pros and cons of the U.S. penny, Pluto's status as a planet). Students were then asked to reduce their argument to its two main elements. Exercises noted that students should not worry about spelling and grammar, and that there were no correct or incorrect answers. For the final wave, control students had an opportunity to revisit their argument from the previous wave.

[4] A small pilot study conducted in September 2012 among a class of middle school students in a Houston public school confirmed that both the treatment and control exercises were well understood by participants, directions were easy to follow, and student responses were thorough.
[5] Enjoying Sports, Being Creative, Enjoying Art, Being Independent, Enjoying Music or Dance, Belonging to a Group, Following Current Events, Being Smart or Getting Good Grades, Playing Video Games, Being With Friends or Family, Being Religious, Having a Sense of Humor, Social Networking or Other Online Activities

Implementation of the four waves occurred over the course of the 2012-2013 academic year (two exercises per semester), each preceding a standardized test by a short period of time (from a few days to about a week). The tests were: (1) the Preliminary SAT/National Merit Scholarship Qualifying Test (hereafter PSAT); (2) school finals; (3) common assessments; (4) and the State of Texas Assessments of Academic Readiness (STAAR). The PSAT is taken every year of high school from 9th-11th grade, is used as practice for the SAT, and offers opportunities for merit scholarships (College Board 2012). The 2012-2013 school year marks the second administration of the STAAR tests, which act as end-of-course assessments. School finals and common assessments are created within the district by teachers and administrators, and test students' knowledge of core curriculum (Math, Science, Social Studies, and English). To provide some insight into HISD's achievement gaps, PSAT math scores for all comparison groups are listed in the Appendix.[6]

3.2.d. Teacher Involvement

HERC team members worked with a chosen liaison at each school to oversee implementation of the exercises. HERC also provided standardized teacher training on implementation, to ensure minimal introduction of bias. Teachers were not apprised of the study's significance for stereotype threat or achievement gaps specifically, but understood that interventions that have shown potential for improving student achievement. For each wave, teachers were instructed to introduce the exercise as a chance for students to express themselves and their thoughts, and asked not to frame exercises as indicative of ability, or as an activity for which students would be evaluated or graded. For each class period, teachers were asked to fill out an "observation sheet" that requested: (1) students absent from each class on the day of

[6] See Table 1.

implementation, to account for nonresponse; and (2) teachers' opinions on logistical success/failure, level of student effort/enjoyment, and suggestions for future implementation.

3.2.e. Measures and Analysis

Intervention effects on standardized math scores – of students in general and minorities, girls, and minority girls in particular – are of primary interest to this research, as success in math is a significant factor in students' continued participation in higher-level math and science (Catsambis 1994; Shapka and Keating 2003; Charles and Bradley 2009). I first loaded the wave 4 (or STAAR Algebra I) data into Stata 13.1, then ran standard regression models to test the variables of interest for that wave[7]. Dependent variables of interest are performance on the PSAT math exam and performance on the STAAR Algebra I exam. Only STAAR results are reported in the results section, due to study limitations (see 3.2.f. below).

[Insert Table 3.2 about here]

I then used postestimation commands to test whether race, gender, and treatment added anything of value to the non-interactive, standard regression model with those same variables in isolation. I found that the interaction term "race x gender x treatment" was not significant for either wave.[8] However, I also found that student academic track is highly significant for the model for each wave, at $p < 0.001$, which suggests (unsurprisingly) that student class placement is incredibly important for test performance, and also that those in higher tracks benefited more from the intervention.

To determine treatment effect, I compared treatment group performance to control group performance (for wave 1/PSAT math and wave 4/STAAR Algebra I).[9] Scores are based on

[7] See Tables 3.1 and 3.2 for variables used in regression analysis.
[8] The "testparm" postestimation command in Stata 13.1 was used to determine the significance of the interaction term. For further coding and commands, please contact author directly.
[9] The "lincom" postestimation command in Stata 13.1 was used to determine intervention effects for each group.

scales generated by the PSAT and STAAR test-makers (College Board and the Texas Education Agency, respectively). Net of economic disadvantage, academic track, and specific school environment, I tested whether (H1)[10] the effect of the intervention was significantly different than that of the control (for all students) suggesting the general utility of the exercises. I also compared the treatment effect for girls (H2), and for minorities (H3). Finally, I compared the performance of minority girls in the treatment to that of minority girls in the control (H4), to better understand the efficacy of the writing exercises for narrowing the raced and gendered achievement gaps in math that doubly disadvantage minority girls and lower their entry into further math education and future STEM career participation.

[Insert Graphs 4a and 4b about here]

3.2.f. Limitations

The experimental process – particularly at a school site – can be challenging and fraught with potential roadblocks throughout the process. Our experience with the HERC experiment was not without its problems, which are listed below.[11] Despite these challenges, it is our belief that analyses are still meaningful, and that further experiments of this type are warranted.

First, students in the re-stratified/-randomized control group at HS3 have statistically higher previous test scores than students in the treatment group. This is less problematic than if previous scores had been higher among students in the treatment group (potentially making any treatment effects more difficult to detect), and in fact points to the possibility that the results may actually be more conservative than if the groups had *more* similar previous scores. In other words, any significant results may underestimate the actual effects of the intervention on treatment vs. control students.

[10] Full hypotheses can be found in the “Hypotheses” section above.
[11] See complete overview of study limitations in Appendix.

Second, staff and faculty at HS2 did not administer the exercises before administration of the first standardized test, the PSAT. Nevertheless, the HERC team decided to administer the exercises at HS2 early in the week following the PSAT, as previous research shows a cumulative effect arising from multiple intervention waves. The same (incorrectly printed) control that HS1 and HS3 received was also used at HS2, in order to limit differences in implementation between schools. While analysis of the intervention effect on HS2's PSAT is thus unfeasible, the cumulative effect of the intervention over the entire school year may still be successfully analyzed using wave 4 results.[12] This is because administering such interventions to students throughout the school year is thought to lead to help them gain immediate increases in performance. When these gains "accumulate in a recursive process, like a chain reaction, the gains are carried forward" (Wisconsin Center for Education Research 2014). Please note that while the PSAT results are not discussed here, they can be found in the Appendix.

4. RESULTS

4.1. STAAR, Algebra I[13]

STAAR Algebra I analyses[14] include both African American and Hispanic students, as all three schools administered the fourth intervention wave prior to the STAAR exam. Using white students at the "diverse" high school (HS3), African Americans at the "diverse" high school (HS3), and African American students in the mostly African American high school (HS2), the treatment group experienced an insignificant effect (β = -109.114) compared to the control group on the STAAR (H1). Girls in the treatment group experienced an insignificant effect (β = 25.813) compared to girls in the control group (H2). As for race, African Americans in the

[12] PSAT results (which are only available for Hispanic and white students at 2 high schools) can be found in the Appendix.
[13] Effect sizes for wave 4/STAAR can be found in Graphs 4a and 4b.
[14] Please note that STAAR Algebra I scale scores range from 1367 to 4942.

treatment group experienced an insignificant effect (β = -37.609) compared to African Americans in the control group (H3). Finally, African American girls in the treatment group experienced an insignificant effect (β = -221.551) compared to African American girls in the control (H4). Due to the statistical insignificance of all findings, there is insufficient evidence to reject any of the null hypotheses for wave 4 (for African American and white students).

Using white students at the “diverse” high school (HS3), Hispanic students at the “diverse” high school (HS3), and Hispanic students in the mostly Hispanic high school (HS1), the treatment group experienced an insignificant effect (β = -101.395) compared to the control group on the STAAR (H1). As in the wave 4 comparisons listed above, girls in the treatment experienced an insignificant effect (β = 39.943) compared to girls in the control group (H2). Additionally, as in wave 1, Hispanics in the treatment group experienced an insignificant effect (β = 49.191) compared to Hispanics in the control group (H3). Finally, Hispanic girls in the treatment group experienced an insignificant effect (β = -381.445) compared to Hispanic girls in the control group. Due to the statistical insignificance of all findings, there is insufficient evidence to reject any of the null hypotheses for wave 4 (for Hispanic and white students).

[Insert Tables 4.2 and 4.3 about here]

Perhaps the most interesting finding from the entire set of analyses begins with the significance of student track in the regression models at the $p \leq 0.001$ level.[15] This significant finding is also displayed in secondary interaction models which interacted track with (1) race, (2) gender, and (3) treatment.[16] In Tables 4.2 and 4.3, results for these interactions show that *AP/IB and pre-AP/IB students in the treatment group performed significantly higher than AP/IB and*

[15] See Tables 4.1-4.3 for regression analyses.
[16] Using Stata 13.1 command “lincom”

pre-AP/IB students in the control group ($p \leq 0.01$), suggesting a beneficial interaction between treatment and track for students in wave 4.

5. CONCLUSION

Previous work using similar affirmation interventions suggests that giving such assignments, throughout the year, leads to a long-term change in students' self-esteem and academic self-concept, which in turn effect students' engagement in academics (in this case, math), which in turn effects academic achievement (in this case, two standardized math exams). Given this vast body of literature, the null findings in the current study were surprising. However, we posit a few considerations for future research of this type.

First, academic self-concept could be addressed in even more directly, perhaps by adding elements to intervention exercises which suggest that aptitude is not necessarily stable and may be improved or changed over time (Aronson, Fried, and Good 2002; Good et al. 2003; Dweck 1999). Second, interventions of this type should be implemented even longer-term (beyond a single year) in order to further address the potential depth or strength of negative self-concept with which some students may struggle, especially since interventions and any gains arising from them seem to "accumulate in a recursive process" (Wisconsin Center for Education Research 2014). Finally, the limitations of implementation experienced in this study should be avoided if possible – namely, researchers carefully ensure the proper temporal ordering of the interventions (before exams) and ensure that previous achievement is more equalized between treatment and control groups. As mentioned in the "Limitations" section, the higher baseline achievement of the control group at one of the high schools points to the possibility that the results may actually be more conservative than if the groups had started with *more* similar baseline achievement. In other words, results in this study may underestimate the actual effects of the intervention on treatment vs. control students, and more similar baseline achievement could help analyze the treatment effect more carefully.

The current study implemented four waves of such intervention in order to influence positive, durable changes to students' self-concept over time, although for the purposes of this study only the first and fourth waves were analyzed (because implementation of these waves occurred before standardized math exams). While interventions that stave off vulnerability to stereotype threat have been shown in the past to increase the math success of minority women and girls, the interventions in this study did not have significant effects on these students' math scores. However, although the overall efficacy of the affirmations for improving math achievement was not convincingly demonstrated in this instance, positive (albeit statistically insignificant) intervention effects for girls and Hispanics in the treatment group suggest the utility of further investigation into affirmation interventions and their effects on standardized testing, particularly for preventing early decreases in math performance and associated negative feedback effects.

More importantly, the significant effect of the interventions among high-achieving students suggests that those young people already well-poised to move into upper-level STEM involvement (but perhaps still facing the social-psychological pressures of stereotype threat) might benefit from such intervention over time. If they receive such intervention, and experience the potentially durable benefits to their academic self-concept, engagement, and achievement, many more high-achieving minority students could move into the advanced ranks of math, and STEM more generally.

5.1. Future Stereotype Threat Interventions

Future studies of this type should heed this implication, targeting high-performing minority students for increased intervention. Future work will also need larger overall sample (or census) sizes and cell sizes, less attrition across waves, and consistently timed dosage across all groups.

Aside from the interesting relationships discovered between track and performance, the race/ethnic distribution of students across tracks is particularly troubling. Among students in the population studied, 29% of white students were in the general track, compared to 51% of black students and 51% of Hispanic students (Kelly 2009). Further studies on stereotype threat intervention should not only seek to hone intervention tools and better control the study environment, they should also more precisely account for tracking within and between schools, to better understand the relationship between tracking and performance, as well as the interaction of student track, intervention status, and performance.

6. APPENDIX

6.1. Tables, Figures, and Graphs

Table 1.1: Average PSAT Scores, HISD 2012-2013

Sub-Group	Avg. (20-80)
All Students	36.9
All African American Students	34.9
All Hispanic Students	36.2
All White Students	44.4

All Women	36.8
African American Women	35.3
Hispanic Women	36.1
White Women	43.8
All Men	36.9
African American Men	34.5
Hispanic Men	36.4
White Men	45.1

Table 1.2: HISD Profile, 2012-2013

Race/Ethnicity	# of students	% pop
Hispanic/Latino	127,091	62.7%
African American	49,781	24.6%
White	16,517	8.2%

Table 2.1: PSAT Math Comparison Groups

(all students in wave 1)

Group	T	C	All
Hispanic women	142	138	280
White women	25	30	55
Hispanic men	135	126	261
White men	31	34	65
TOTAL	**333**	**328**	**661**

Table 2.2: STAAR Math Comparison Groups (with group percentages)

(all students in wave 4)

Group	T	C	All	%AfAm	%Hisp	%White
Af Am women	24	35	59	**11.5%**		
Hispanic women	95	83	178		**34.6%**	
White women	14	14	28			**5.4%**
Af Am men	28	33	61	**11.9%**		
Hispanic men	80	78	158		**30.1%**	
White men	13	17	30			**5.8%**
TOTAL	**254**	**260**	**514**	**22.9%**	**64.7%**	**11.2%**

Table 2.3: PSAT Mean Score Comparisons

Hispanic women in Treatment (T), compared to:

H1	Hispanic women in C	-0.47
H2.1	Hispanic, white women in T & C	-1.44 *
H2.2	Hispanic, white men in T & C	-1.32 *
H2.3	White women in T & C	-1.34 *
H2.4	Hispanic men in T & C	-1.10
H2.5	White men in T & C	-1.29 *

****p < 0.001, **p < 0.01, *p < 0.05*

Table 2.4: STAAR Mean Score Comparisons

Af Am and Hispanic women in Treatment (T), compared to:		
H1	Af Am, Hispanic women in C	-1.43
H2.1	All women in T & C	-4.07 ***
H2.2	All men in T & C	-1.13
H2.3	White women in T & C	-8.47 ***
H2.4	Af Am, Hispanic men in T & C	0.78
H2.5	White men in T & C	-8.92 ***

****p < 0.001, **p < 0.01, *p < 0.05*

Table 2.5: STAAR Mean Score Comparisons

African American women in Treatment (T), compared to:		
H1	Af Am women in C	-2.19
H2.1	All women in T & C	-2.95
H2.2	All men in T & C	-3.00
H2.3	White women in T & C	-3.02
H2.4	Af Am men in T & C	-2.32
H2.5	White men in T & C	-3.02

****p < 0.001, **p < 0.01, *p < 0.05*

Table 2.6: STAAR Mean Score Comparisons

Hispanic women in Treatment (T), compared to:		
H1	Hispanic women in C	2.19
H2.1	All women in T & C	0.02
H2.2	All men in T & C	-0.28
H2.3	White women in T & C	-0.29
H2.4	Hispanic men in T & C	-0.34
H2.5	White men in T & C	-0.29

****p < 0.001, **p < 0.01, *p < 0.05*

Table 3.1: Regression Variables, Wave 1

Variable Name	Variable Description	DV/IV	Variable Type	Variable Range	Measurement Details
psatmath	PSAT Scale Score	DV	num	20-68	Used PSAT scale scores taken from HISD data. Matched students on id using HERC stereotype threat data.
hisp1	Hispanic Students vs. White Students	IV	dummy	0-1	Created dummy variable to distinguish Hispanic students as group of interest (1) and White students as reference group (0)

					for wave 1.
female	Females vs. Males	IV	dummy	0-1	Created dummy variable to distinguish females as group of interest (1) and males as reference group (0).
eco_dis	Economically Disadvantaged Control	IV	dummy	0-1	Used dummy variable from HERC data.
track	Track Control	IV	num	0-1	Created 0 track, which is AP/pre-AP/IB, and 1, which is all other.
hs1	School Control	IV	dummy	0-1	Created dummy variable to distinguish students in HS1 as group of interest (1) and students in HS3 as reference group (0).

Table 3.2: Regression Variables, Wave 4

Variable Name	**Variable Description**	**DV/IV**	**Variable Type**	**Variable Range**	**Measurement Details**
rescal13	Algebra I Scale Score	DV	num	1367-4942	Used STAAR scale scores taken from HISD data. Matched students on id using HERC stereotype threat data.
hisp4	Hispanic Students vs. White Students	IV	dummy	0-1	Created dummy variable to distinguish Hispanic students as group of interest (1) and White students as reference group (0)

					for wave 4.
afam4	African American Students vs. White Students	IV	dummy	0-1	Created dummy variable to distinguish African American students as group of interest (1) and White students as reference group (0) for wave 4.
female	Females vs. Males	IV	dummy	0-1	Created dummy variable to distinguish females as group of interest (1) and males as reference group (0).
eco_dis	Economically Disadvantaged Control	IV	dummy	0-1	Used dummy variable from HERC data.
track	Track Control	IV	num	0-1	Created 0 track, which is AP/pre-AP/IB, and 1, which is all other.
hs1	School Control	IV	dummy	0-1	Created dummy variable to distinguish students in HS1 as group of interest (1) and students in HS3 as reference group (0).
hs2	School Control	IV	dummy	0-1	Created dummy variable to distinguish students in HS2 as group of interest (1) and students in HS3 as reference group (0).

Table 4.1: Wave 1 Regression Models

Model 1: PSAT math scores regressed on race, gender, treatment, economic disadvantage, class track, and specific school environment.
Model 2: PSAT math scores regressed on race, gender, treatment, economic disadvantage, class track, specific school environment, and race x gender x treatment interaction terms. (*HERC 2013*)

Variables	Model 1	Model 2
Hispanic	-1.894	-2.381
(Reference = white)	(0.959)	(1.540)

Female	0.092	0.790
(Reference = male)	(0.555)	(1.787)
Treatment	0.132	-1.094
(Reference = control)	(0.554)	(1.773)
Economic disadvantage	-0.764	-0.713
(Reference = not econ dis)	(0.836)	(0.838)
Track	-6.952***	-6.951***
(Reference = AP, pre-AP/-IB)	(0.611)	(0.613)
Mostly Hispanic School (HS1)	-0.741	-0.726
(Reference = HS3)	(0.711)	(0.713)
Hispanic x Female	----	-0.441
	----	(1.987)
Hispanic x Treatment	----	1.937
		(1.979)
Female x Treatment	----	0.362
		(2.595)
Hispanic x Female x Treatment	----	-1.270
		(2.871)
Observations	694	694
Prob > F	0.000	0.000
R2	0.232	0.235
Constant	44.287***	44.462***

***p ≤ 0.001; **p ≤ 0.01; *p ≤ 0.05

Table 4.2: Wave 4 Regression Models – AfAm & White Students

Model 1: STAAR Algebra I scores regressed on race, gender, treatment, economic disadvantage, class track, and specific school environment.
Model 2: STAAR Algebra I scores regressed on race, gender, treatment, economic disadvantage, class track, specific school environment, and race x gender x treatment interaction terms. (*HERC 2013*)

Variables	**Model 1**	**Model 2**
African American	-30.616	-117.988

(Reference = white)	(70.544)	(106.242)
Female	42.890	-94.848
(Reference = male)	(47.395)	(112.288)
Treatment	-33.066	-109.114
(Reference = control)	(47.025)	(112.308)
Economic disadvantage	-53.908	-61.149
(Reference = not econ dis)	66.578	(67.395)
Track	-330.913***	-329.175***
(Reference = AP, pre-AP/-IB)	(52.222)	(52.727)
Mostly AfAm School (HS2)	-219.113	-209.870
(Reference = HS3)	(63.810)	(64.718)
AfAm x Female	----	166.235
		(135.385)
AfAm x Treatment	----	71.505
		(137.927)
Female x Treatment	----	134.927
		(164.881)
AfAm x Female x Treatment	----	-112.437
		(201.352)
Observations	225	225
Prob > F	0.000	0.000
R2	0.349	0.355
Constant	3952.557***	4024.360***

***p ≤ 0.001; **p ≤ 0.01; *p ≤ 0.05

Table 4.3: Wave 4 Regression Models – Hisp. & White Students

Model 1: STAAR Algebra I scores regressed on race, gender, treatment, economic disadvantage, class track, and specific school environment.
Model 2: STAAR Algebra I scores regressed on race, gender, treatment, economic disadvantage, class track, specific school environment, and race x gender x treatment interaction terms. (*HERC 2013*)

Variables	Model 1	Model 2
Hispanic	-78.171	-144.667

(Reference = white)	(66.439)	(102.615)
Female	-33.082	-83.998
(Reference = male)	(35.206)	(122.338)
Treatment	-25.391	-101.345
(Reference = control)	(35.078)	(122.315)
Economic disadvantage	31.659	35.678
(Reference = not econ dis)	(58.575)	(58.679)
Track	-261.340***	-262.090***
(Reference = AP, pre-AP/-IB)	(38.736)	(38.760)
Mostly Hispanic School (HS1)	-4.879	-7.251
(Reference = HS3)	(45.700)	(45.739)
Hispanic x Female	----	117.314
		(133.632)
Hispanic x Treatment	----	150.586
		(134.190)
Female x Treatment	----	141.338
		(179.758)
Hispanic x Female x Treatment	----	-280.050
		(195.299)
Observations	475	475
Prob > F	0.000	0.000
R2	0.113	0.120
Constant	3946.883***	3977.019***

***p ≤ 0.001; **p ≤ 0.01; *p ≤ 0.05

Figure 1: Example Treatment Exercise

DIRECTIONS:

There are a lot of things that are important to people—things that make their lives better, happy, or special. These might be activities that they **enjoy**, things that they are **good at**, or relationships that they **care about**.

In the space below, please describe something that is important to YOU in your life.

Feel free to write as much as you want. Focus on your thoughts and feelings, and don't worry about spelling or how well it is written.

Figure 2: Example Control Exercise

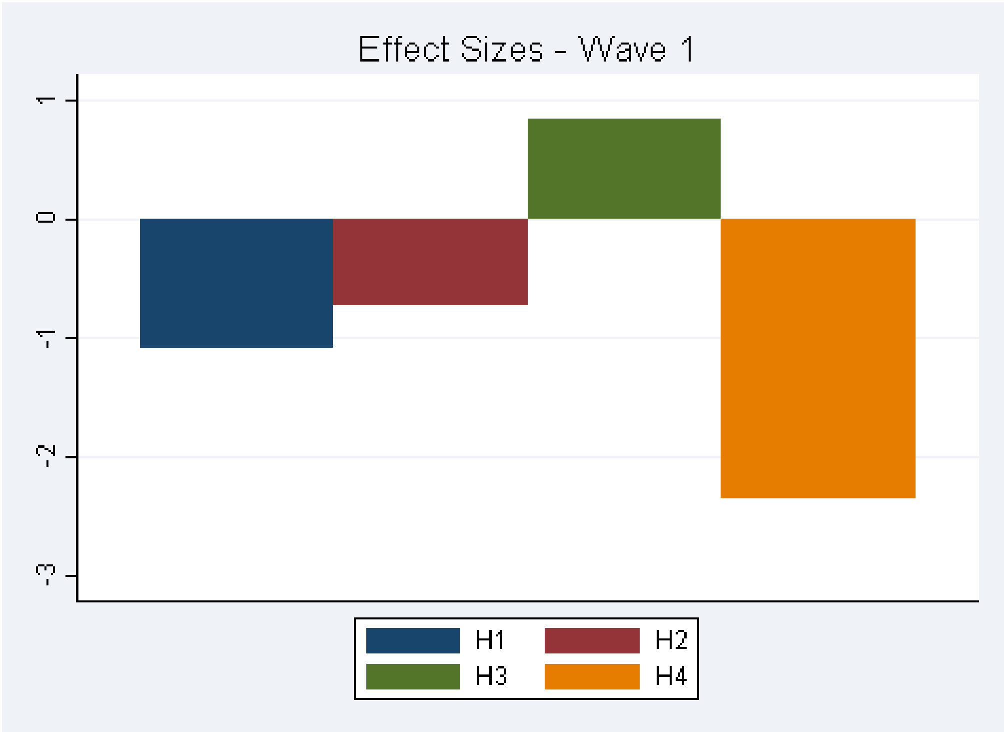
DIRECTIONS:

Online newspapers cost less and allow people to get the news faster. However, research has shown that reading the print version of the news makes it easier to remember. Some people think we should only use online newspapers. Others think we should only use print newspapers.

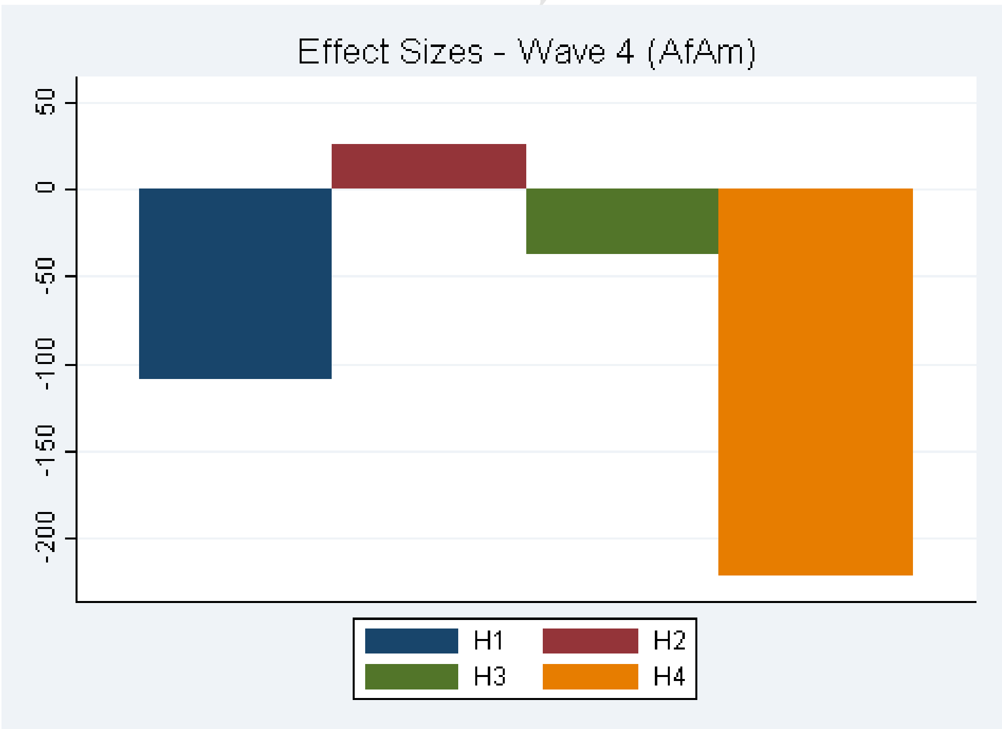
Do you think we should use online or print newspapers? Please write **3 REASONS** why you feel the way you do.

Feel free to write as much as you want. Focus on your ideas, and don't worry about spelling or how well it is written.

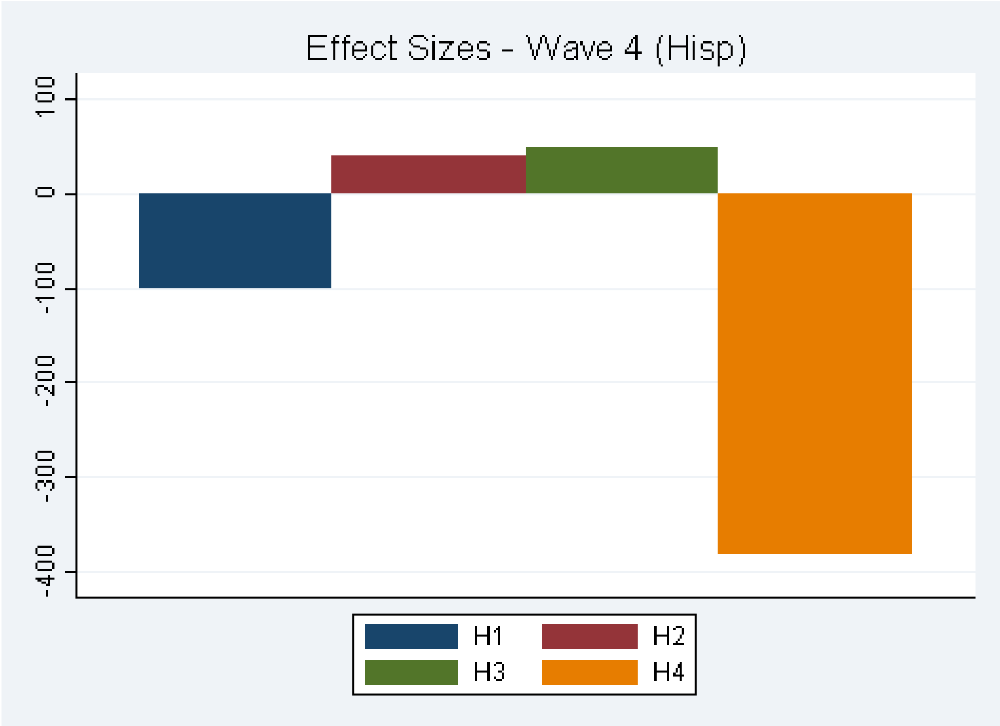
Graph 1



Graph 4a



Graph 4b



6.2. Limitations, cont'd

The HERC research team took several preventative steps to guard against potential implementation fidelity issues arising from complex study design. As mentioned above, detailed teacher training was offered to standardize implementation practices. Additionally, exercises were administered by teachers, and not an unfamiliar presence, thus guarding against any bias arising from the presence of an outsider. Finally, each teacher and school liaison was given a $100 gift card to thank teachers, and to incentivize proper implementation protocol.

HERC team members also took every opportunity to avoid issues before implementation (e.g., contracting a copying company, cross-checking the copy job against school enrollment lists). Another concern at the front end was the nonresponse that would inevitably arise from a student absences, or lack of participation. Few preventative measures could have been taken on HERC's part to ensure success on this front.

Several problems arose during wave 1 in October 2012. First, the contracted copying company did not finish the copy job in time, forcing a last-minute switch to another company. Second, schools released updated rosters days before the first wave, forcing secondary re-stratification and re-randomization due to a perceived large influx of unexpected students into the population under study. Once the dust settled, the HERC team found that new students were in fact not meant to be included at all, as most were enrolled in remedial, special education, or other English courses (i.e., not grade level courses). From these issues arose several others, which affected implementation.

Problems encountered during wave one informed the much more successful implementation of wave two. First, HERC edited rosters based on wave one and conferred with school liaisons to ensure that rosters were up to date and correct several weeks in advance of the second wave. Second, the copying company that successfully finished the first wave was contracted for wave two, ensuring that sorting and school delivery were finished in a timely and organized manner. The first page of the control exercise was printed incorrectly, and students were asked to circle values that are "most important" to them before completing the written portion of the exercise, rather than values that were not important to them. However, the remainder of the control packet was printed correctly and students in the control group did not write about the values, as did students in the treatment group. Thus, the control group is considered to have received a small "dose" of the treatment in that students circled personal values. Exercises were copy-edited by multiple HERC members, to help ensure that no further "dosage" or other issues would arise, and elimination of this error in subsequent exercises should have added to HERC's ability to test a cumulative treatment effect over all waves.

Additionally, problems were encountered with the incorrect placement of student exercises into classroom packets for HS1. However, the HERC liaison and school liaison for HS1 caught the problem in time, and re-organized exercises to ensure that implementation would be seamless for teachers. HERC is unsure at this time whether this organizational error occurred at the copying company or during packet sorting by HERC members.

Earlier issues with waves one and two prompted more successful implementation at waves three and four, but issues did still arise. During wave three, one of the teachers at HS3, did not administer the exercise to an entire class, thereby eliminating all of the students in that class from analysis of that wave.

Wave four saw only minor implementation problems – first, a teacher at HS1 distributed the exercises of absent students in the treatment group to new students who were not on the roster, and who therefore should have filled out a blank control exercise; and second, a new English teacher joined the staff at HS1, into whose classes multiple students (previously in other classes) were transferred. This mainly affected data entry, and not the exercise implementation itself, as HERC was aware of the transition well in advance.

6.2.1. PSAT Results[17]

African American students were not included in PSAT analyses (wave 1)[18], as the high school with a majority African American population (HS2) did not receive the treatment before the PSAT was administered. Nevertheless, the HERC team decided to administer the exercises at HS2 early in the week following the PSAT, as previous research shows a cumulative effect arising from multiple intervention waves. Students at HS2 received the treatment after the PSAT, which makes them eligible for the overall treatment analysis across all waves of the intervention.

[17] Effect sizes for wave 1/PSAT can be found in the Appendix in Graph 1.
[18] Please note that PSAT scale scores range from 20 to 68.

African American students thus appear in wave 4 regression models. Wave 1 only includes white and Hispanic students at HS1 and HS3. Here the treatment group experienced an insignificant effect (β = -1.094) compared to the control group (H1). Additionally, girls in the treatment group experienced an insignificant effect (β = -0.732) compared to girls in the control group (H2). As for race, Hispanics in the treatment group experienced an insignificant effect (β = 0.843) compared to Hispanics in the control (H3). Finally, Hispanic girls in the treatment experienced an insignificant effect (β = -2.364) compared to Hispanic girls in the control (H4). Due to the statistical insignificance of all findings, there is insufficient evidence to reject any of the null hypotheses for wave 1.

PSAT analyses were meant to act as a preliminary sort of “baseline” of comparison for students’ STAAR Algebra I exam scores, in order to better test the effect of the intervention over time as the PSAT was administered in the fall, and the STAAR in the spring. Because African Americans students’ post-PSAT intervention exposure, wave 1 and wave 4 scores may only be analyzed comparatively for Hispanic students (i.e., the PSAT outcome is invalid for African American students, and therefore does not appear in the results). However, all students included in wave 4 analysis were exposed to both waves of the intervention. First wave exposure simply occurred slightly later for African American students at HS2 (where most of the study’s African American students are enrolled) than for other students in the study. In other words, wave 4 (STAAR) analyses are completely valid for all students, as all were exposed to the intervention at waves 1 and 4.

7. ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. Geoffrey Cohen of Stanford University and Dr. Geoffrey Borman of the University of Wisconsin for their time and recommendations. Their research instruments, after which ours were modeled, and their expert counsel, which was both thoughtful and gracious, guided the design and implementation of this project.

The authors also thank Dr. Holly Heard, HERC's Data Goddess, for her endless expertise and patience with implementation and analysis. Special thanks also to Drs. Dara Schifrer and Diego Torres, and to Sandra Alvear, for their invaluable contributions to the stereotype threat project.

The Houston Education Research Consortium (HERC) is directed by Dr. Ruth López Turley and housed within Rice University's Kinder Institute for Urban Research. HERC is funded by the Laura and John Arnold Foundation, but this particular project was not funded by any specific foundation grants. Any opinions expressed in this report reflect those of the authors.

8. REFERENCES

Alexander, Karl L., Doris R. Entwisle, and Carrie S. Horsey. 1997. "From First Grade Forward: Early Foundations of High School Dropout." Sociology of Education 70(2): 87-107.

Aronson, J., C.B. Fried, and C. Good. 2002 "Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence." *Journal of Experimental Social Psychology* 38:113-125.

Ayalon, H. 1995. "Math as a Gatekeeper: Ethnic and Gender Inequality in Course Taking of the Sciences in Israel." *American Journal of Education* 104:34-56.

Beilock, S. L., W.A. Jellison, R.J. Rydell, A.R. McConnell, and T.H. Carr. 2006. "On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened?" *Personality and Social Psychology Bulletin* 32:1059–1071.

Beilock, S. L., R.J. Rydell, and A.R. McConnell. 2007. "Stereotype threat and working memory: Mechanisms, alleviations, and spillover." *Journal of Experimental Psychology: General*, 136(2):256-276.

Blascovich, J., S.J. Spencer, D.M. Quinn, and C.M. Steele. 2001. "African-Americans and high blood pressure: The role of stereotype threat." *Psychological Science* 12:225-229.

Brown, S.V. 1994. *Underrepresented minority women in graduate science and engineering education*. Princeton, NJ: Educational Testing Service.

Brown, S.V. 1995. "Testing the double bind hypothesis: Faculty recommendations of minority women fellowship applicants." *Journal of Women and Minorities in Science and Engineering* 2(4):207-223.

Burrelli, J. 2009. Women of color in STEM education and employment. Paper presented at the Mini-Symposium on Women of Color in STEM, Arlington, VA. Retrieved November 24, 2012 (http://www.nsf.gov/od/oia/activities/ceose/mini-symp-pres/Women_of_color_stem_Oct2009/Oct27/JoanBurrelliv2.pdf).

Carolene, H.B. and A. Johnson. 2007. "Understanding the science experiences of successful women of color: Science identity as an analytic lens." *Journal of Research in Science Teaching* 44(8):1011-1245.

Catsambis, S. 1994. "The Path to Math: Gender and Racial-Ethnic Differences in Mathematics Participation from Middle School to High School." *Sociology of Education* 67(3):199-215.

Cech, E.A. and M. Blair-Loy. 2010. “Perceiving Glass Ceilings? Meritocratic versus Structural Explanations of Gender Inequality among Women in Science and Technology.” *Social Problems* 57(3):371-397.

Charles, M. and K. Bradley. 2009. “Indulging our Gendered Selves? Sex Segregation by Field of Study in 44 Countries.” *American Journal of Sociology* 114(4):924-76.

Cheryan, S. and G.V. Bodenhausen. 2000. “When positive stereotypes threaten intellectual performance: The psychological hazards of ‘model minority’ status.” *American Psychological Society* 11(5):399-402.

Cohen, G. L. and J. Garcia 2005. “‘I am us’: Negative stereotypes as collective threats.” *Journal of Personality and Social Psychology* 89(4):566-582.

Cohen, G. L., J. Garcia, N. Apfel, and A. Master. 2006. “Reducing the racial achievement gap: A social-psychological intervention.” *Science* 313:1307-1310.

Cohen, G.L., J. Garcia, V. Purdie-Vaughns, N. Apfel, and P. Brzustoski. 2009. “Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap.” *Science* 324(5925):400-403.

College Board. 2012. “PSAT/NMSQT: About PSAT/NMSQT.” Retrieved November 20, 2012 (http://www.collegeboard.com/student/testing/psat/about.html).

College Board. 2012. “PSAT/NMSQT: Scores and Review.” Retrieved November 20, 2012 (http://www.collegeboard.com/student/testing/psat/scores.html).

Collins, M. and M.L. Matyas. 1985. “Minority Women: Conquering both Sexism and Racism.” In J.B. Kahle (Ed.), *Women in Science*. London: Falmer Press.

Department of Research and Accountability, Houston Independent School District. 2013. “Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT) Report: Fall 2011-2012.” Retrieved September 28, 2012 (http://www.houstonisd.org/site/default.aspx?PageType=14&DomainID=8269&PageID=38527&ModuleInstanceID=82738&IsMoreExpandedView=True).

Department of Research and Accountability, Houston Independent School District. 2013. “Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT) Report: Fall 2012-2013.” Retrieved June 5, 2013 (http://www.houstonisd.org/site/default.aspx?PageType=14&DomainID=8269&PageID=38527&ModuleInstanceID=82738&IsMoreExpandedView=True).

Dweck, C.S. 1999. *Self-Theories: Their role in motivation, personality, and development*. Philadelphia, PA: The Psychology Press.

Espinosa, L. 2011. “Pipelines and Pathways: Women of Color in Undergraduate STEM Majors and College Experiences That Contribute to Persistence.” *Harvard Educational Review* 81(2):209-240.

Gonzales, P.M., H. Blanton, and K.J. Williams. 2002. “The effects of stereotype threat and double-minority status on the test performance of Latino women.” *Personality and Social Psychology Bulletin* 28(5):659-670.

Good, C., J. Aronson, and J.A. Harder. 2008. “Problems in the pipeline: Stereotype threat and women’s achievement in high-level math courses.” *Journal of Applied Developmental Psychology* 29:17-28.

Good, C., J. Aronson, and M. Inzlicht. 2003. “Improving Adolescents' Standardized Test Performance: An Intervention to Reduce the Effects of Stereotype Threat.” *Journal of Applied Developmental Psychology* 24(6):645-662.

Gresky, D.M., L.L. Ten Eyck, C.G. Lord, and R.B. McIntyre. 2005. "Effects of salient multiple identities on women's performance under mathematics stereotype threat." *Sex Roles* 53:703-716.

Hanselman, Paul, Sarah K. Bruch, Adam Gamoran, and Geoffrey D. Borman. 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.

Houston Education Research Consortium (HERC). 2013. Stereotype Threat Study.

Huguet, P. and I. Régner. 2007. "Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances." *Journal of Educational Psychology* 99:545-560.

Inzlicht, M., L. McKay, and J. Aronson. 2006. "Stigma as ego depletion: How being the target of prejudice affects self-control." *Psychological Science* 17(3):262-269.

Jones Taylor, V. and G.M. Walton. 2011. "Stereotype Threat Undermines Academic Learning." *Personality and Social Psychology Bulletin* 37(8):1055-1067.

Keifer, A.K. and D. Sekaquaptewa. 2007. "Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat." *Journal of Experimental Social Psychology* 43:825-832.

Kellow, T. J. and B.D. Jones. 2005. "Stereotype threat in African-American high school students: An initial investigation." *Current Issues in Education* 8(20).

Kelly, S. 2009. "The Black-White Gap in Mathematics Course Taking." *Sociology of Education* 82(1):47-69.

Krendl, A.C., J.A. Richeson, W.M. Kelley, and T.F. Heatherton. 2008. "The negative consequences of threat: A functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math." *Psychological Science* 19:168-175.

Lucas, Samuel R., and Aaron D. Good. 2001. "Race, Class, and Tournament Track Mobility." *Sociology of Education* 74(2):139-156.

Major, B., S. Spencer, T. Schmader, C. Wolfe, and J. Crocker. 1998. "Coping With Negative Stereotypes about Intellectual Performance: The Role of Psychological Disengagement." *Personality and Social Psychology Bulletin* 24:34-50.

Malcom, L. and S. Malcom. 2011. "The Double Bind: The Next Generation." *Harvard Educational Review* 81(2):162-172.

Malcom, S.M., P.Q. Hall, and J.W. Brown. 1976. "The Double Bind: The Price of Being a Minority Woman in Science." *American Association for the Advancement of Science* 76(3):1-70.

Martens, A., M. Johns, J. Greenberg, and Schimel. 2006. "Combating stereotype threat: The effect of self-affirmation on women's intellectual performance." *Journal of Experimental Psychology* 44:645-657.

McIntyre, R.B., R. Paulson., and C. Lord. 2003. "Alleviating women's mathematics stereotype threat through salience of group achievements." *Journal of Experimental Social Psychology* 39(1):83-90.

Miyake, A., L.E. Kost-Smith, N.D. Finkelstein, S.J. Pollock, G.L. Cohen, and T.A. Ito. 2010. "Reducing the Gender Achievement Gap in College Science: A Classroom Study of Values Affirmation." *Science* 330:1234-1237.

National Science Foundation. 2011. "Women, Minorities, and Persons with Disabilities in Science and Engineering." Retrieved November 21, 2012 (http://www.nsf.gov/statistics/wmpd/).

Oaks, Jeannie. 1994. "More Than Misapplied Technology: A Normative and Political Response to Hallinan on Tracking. *Sociology of Education* 67(2): 84-91.
Obama, B. 2009. "Remarks by the President at the National Academy of Sciences annual meeting." Retrieved November 29, 2012 (http://www.whitehouse.gov/the_press_office/Remarks-by-the-President-at-the-National-Academy-of-Sciences-Annual-Meeting/).
Ong, M. 2002. "Against the current: Women of color succeeding in physics," Doctoral dissertation, University of California, Berkeley. Retrieved November 30, 2012 from ProQuest Dissertations and Theses database (http://search.proquest.com.ezproxy.rice.edu/docview/304803810/13AB7FF1EBF7EBE054D/1?accountid=7064).
Ong, M., C.Wright, L. Espinosa, and G. Orfield. 2011. "Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering, and mathematics." *Harvard Educational Review* 81(2):172-208.
Osborne, J.W., and C. Walker. (2006). "Stereotype threat, identification with academics, and withdrawal from school: Why the most successful students of colour might be the most likely to withdraw." *Educational Psychology* 26:563-577.
Osborne, J.W. 2001. "Testing stereotype threat: Does anxiety explain race and sex differences in achievement?" *Educational Psychology* 26:291-310.
Osborne, J.W. 2007. Linking stereotype threat and anxiety." *Educational Psychology* 27:135-154.
Reyes, M.E. 2011. "Unique Challenges for Women of Color in STEM Transferring from Community Colleges to Universities." *Harvard Educational Review* 81(2):241-263.
Schimel, J., J. Arndt, K.M. Banko, and A. Cook. 2004. "Not all self-affirmations were created equal: The cognitive and social benefits of affirming intrinsic (vs. extrinsic) self." *Social Cognition* 22:75-99.
Schmader, T. and M. Johns. 2003. "Converging evidence that stereotype threat reduces working memory capacity." *Journal of Personality and Social Psychology* 85:440-452.
Sells, L.W. 1973. "High school mathematics as the critical filter in the job market." In R.T. Thomas (Ed.), *Developing opportunities for minorities in graduate education*. Berkeley: University of California Press.
Shapiro, J. R., and S.L. Neuberg. 2007. "From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions." *Personality and Social Psychology Review* 11(2)107-130.
Shapka, J.D. and D.P. Keating. 2003. "Effects of a Girls-Only Curriculum during Adolescence: Performance, Persistence, and Engagement in Mathematics and Science." *American Educational Research Journal* 40(4):929-960.
Sherman, D.K. and G.L. Cohen. 2006. "The psychology of self-defense: Self-affirmation theory." *Advances in Experimental Social Psychology* 38:183-242.
Shih, M., T.L. Pittinsky, and N. Ambady. 1999. "Stereotype susceptibility: Identity salience and shifts in quantitative performance." *Psychological Science* 10(2):80-83.
Spencer, S.J., C.M. Steele, and D.M. Quinn. 1999. "Stereotype threat and women's math performance." *Journal of Experimental Social Psychology* 35:4-28.
Steele, C. M. and J. Aronson. 1995. "Stereotype threat and the intellectual test performance of African-Americans." *Journal of Personality and Social Psychology* 69:797-811.

Tate, E.D. and M.C. Linn. 2005. “How does identity shape the experiences of women of color engineering students?” *Journal of Science Education and Technology* 14(5-6):483-493.
Texas Education Agency. 2012. “STAAR Resources.” Retrieved November 20, 2012 (http://www.tea.state.tx.us/student.assessment/staar/).
Texas Education Agency. 2012. “STAAR Statewide Summary Reports 2011-2012.” Retrieved November 20, 2012 (http://www.tea.state.tx.us/staar/rpt/sum/yr12/).
The White House, Office of the Press Secretary. 2009. “President Obama Launches ‘Educate to Innovate’ Campaign for Excellence in Science, Technology, Engineering & Math (Stem) Education.” Retrieved November 29, 2012 (http://www.whitehouse.gov/the-press-office/president-obama-launches-educate-innovate-campaign-excellence-science-technology-en).
Wilson, T.D. and P.W. Linville. 1985. “Improving the performance of college freshman with attributional techniques.” *Journal of Personality and Social Psychology* 49(1):287-293.
Wisconsin Center for Education Research 2013: Project by Borman, J. and Gamoran, A. Retrieved September 30, 2012 (http://www.wcer.wisc.edu/projects/projects.php?project_num=747).
Wisconsin Center for Education Research 2014: Project by Borman, J. and Gamoran, A. Retrieved December 1, 2015 (2013http://www.wcer.wisc.edu/news/coverstories/2014/Self-Affirmation_Exercises_Found_To_Boost_Some_Achievement.php).