Examination of a Self-Affirmation Intervention in St. Paul Public Schools

Geoffrey Borman

The Senior Urban Education Research Fellowship Series
Volume IX

Examination of a Self-Affirmation Intervention in St. Paul Public Schools

Geoffrey Borman

Fall 2012

TABLE OF CONTENTS

TABLES AND FIGURES

OVERVIEW

THE SENIOR URBAN EDUCATION RESEARCH FELLOWSHIP PROGRAM

Large urban public school districts play a significant role in the American education system. The largest 67 urban school systems in the country – comprising less than one half of one percent of the nearly seventeen thousand school districts that exist across the United States – educate about 14 percent of the nation's K-12 public school students, including over 20 percent of the nation's economically disadvantaged students, 28 percent of its African American students, about a quarter of its Hispanic students, and a quarter of its English Language Learners. Clearly, any attempt to improve achievement and to reduce racial and economic achievement gaps across the United States must involve these school districts as a major focus of action.

These school districts face a number of serious, systematic challenges. To better understand the problems in urban education and to develop more effective and sustainable solutions, urban districts need a program of rigorous scientific inquiry focusing on what works to improve academic outcomes in the urban context. Moreover, in order to produce such evidence and to move public education forward generally, the standards of evidence in education research must be raised in such a way as to bring questions regarding the effectiveness of educational interventions and strategies to the fore and to promote careful scrutiny and rigorous analysis of the causal inferences surrounding attempts to answer them.

It has been argued that, in order to move such an effort forward, a community of researchers, committed to a set of principles regarding evidentiary standards, must be developed and nurtured. We contend further that, in order to produce a base of scientific knowledge that is both rigorously derived and directly relevant to improving achievement in urban school districts, this community of inquiry must be expanded to include both scholars and practitioners in urban education.

Though a great deal of education research is produced every year, there is a genuine dearth of knowledge regarding how to address some of the fundamental challenges urban school districts face in educating children, working to close achievement gaps, and meeting the demands of the public for better results. Moreover, while there is a history of process-related research around issues affecting urban schools, relatively few studies carefully identify key program components, document implementation efforts, and carefully examine the effects of well-designed interventions in important programmatic areas on key student outcomes such as academic achievement. In sum, there is an absence of methodologically sound, policy-relevant research to help guide practice by identifying the conditions, resources, and necessary steps for effectively mounting initiatives to raise student achievement.

In order to address this need, the Council of the Great City Schools, through a grant from the Institute of Education Sciences, established the Senior Urban Education Research Fellowship (SUERF) program.

The Senior Urban Education Research Fellowship was designed to facilitate partnerships between scholars and practitioners focused on producing research that is both rigorous in nature and relevant to the specific challenges facing large urban school districts. We believe such partnerships have the potential to produce better, more practically useful research in at least three ways. First, by deepening researchers' understanding of the contexts within which they are working, the program may help them maximize the impact of their work in the places where it is needed the most. Second, by helping senior staff in urban districts become better consumers of research, we hope to increase the extent to which the available evidence is used to inform policy and practice, and the extent to which urban districts continue to invest in research. Third, by executing well-designed studies aimed at the key challenges identified by the districts themselves, we hope to produce reliable evidence and practical guidance that can help improve student achievement.

The primary goals for the Senior Urban Education Research Fellowship are to:

- promote high quality scientific inquiry into the questions and challenges facing urban school districts;

- facilitate and encourage collaboration, communication, and ongoing partnerships between senior researchers and leaders in urban school districts;

- demonstrate how collaboration between scholars and urban districts can generate reliable results and enrich both research and practice;

- produce a set of high quality studies that yield practical guidance for urban school districts;

- contribute to an ongoing discussion regarding research priorities in urban education; and

- promote the development of a "community of inquiry", including researchers and practitioners alike, committed to both a set of norms and principles regarding standards of evidence and a set of priorities for relevant, applied research in urban education.

The SUERF program benefitted greatly from the guidance and support of a Research Advisory Committee made up of experts and leaders from large urban school districts and the education research community. The committee included Dr. Katherine Blasik, Dr. Carol Johnson, Dr. Kent McGuire, Dr. Richard Murnane, Dr. Andrew Porter, and Dr. Melissa Roderick. This extraordinary group helped to identify and define the objectives and structure of the fellowship program, and we thank them for lending their considerable insight and expertise to this endeavor.

The following volume of the *Senior Urban Education Research Fellowship Series* documents the research of Dr. Geoffrey Borman working in collaboration with the St. Paul Public Schools. Both the research and reporting is the sole intellectual property of Dr. Borman, and reflects his personal experience and perspective.

Dr. Borman's examination of an intervention designed to address stereotype threat adds to a growing base of research documenting the social and psychological dimensions of student achievement and engagement in school. Although the intervention was not found to have the impact on African American male achievement that was expected, the pursuit of these types of targeted interventions and programs—as well as the concurrent effort to evaluate them—is a promising direction for both school districts and education researchers. As we improve our understanding of the specialized needs and vulnerabilities of minority and at-risk students— particularly African American males—we are better able to support these students socially and academically.

We hope you will find this report both interesting and relevant to your own work in education.

Thank you.

Michael Casserly
Executive Director
Council of the Great City Schools

ABOUT THE SENIOR URBAN EDUCATION RESEARCH FELLOW

Trained as a quantitative methodologist at the University of Chicago, Dr. Borman (Ph.D., 1997) is a Professor of Education and Sociology at the University of Wisconsin–Madison, the Co-Director of the University of Wisconsin's Predoctoral Interdisciplinary Research Training Program, and a Senior Researcher with the Consortium for Policy Research in Education. Professor Borman's main substantive research interests revolve around the social distribution of the outcomes of schooling and the ways in which policies and practices can help address and overcome educational inequality. His primary methodological interests include the synthesis of research evidence, the design of quasi-experimental and experimental studies of educational innovations, and the specification of school-effects models.

Over the past ten years, Borman has led or co-directed twelve major randomized controlled trials, which have included randomization and delivery of educational interventions at the student, classroom, school, and district levels. He has conducted three recent research syntheses, including a meta-analysis of the achievement effects of 29 nationally disseminated school reform models. Finally, other recent projects reveal the consequences of attending high-poverty schools and living in high-poverty neighborhoods and uncover some of the mechanisms through which social-context effects may be manifested.

Professor Borman has been appointed as a methodological expert to advise many national research and development projects, including the National Research Center on the Gifted and Talented and three of the nation's regional educational laboratories funded by the Institute of Education Sciences. He was also named to the 15-member Urban Education Research Task Force established to advise the U.S. Department of Education on issues affecting urban education. Borman serves on the editorial boards of seven academic journals, including the *American Educational Research Journal, Reading Research Quarterly,* and *Elementary School Journal.* His research has been funded by a variety of organizations, including the National Science Foundation, U.S. Department of Education, Office of Educational Research and Improvement, Institute of Education Sciences, American Educational Research Association Grants Program, Spencer Foundation, Open Society Institute, and Smith-Richardson Foundation, among others. Dr. Borman was the recipient of a 2002 National Academy of Education/Spencer Postdoctoral Fellowship Award, the 2004 Raymond Cattell Early Career Award from the American Educational Research Association, the 2004 American Educational Research Association Review of Research Award, and the 2008 American Educational Research Association Palmer O. Johnson Award. In 2009, Dr. Borman's significant contributions to the field of education research were recognized by his nomination and selection as a Fellow of the American Educational Research Association.

This work builds on prior successful collaborations between Professor Geoffrey D. Borman of the University of Wisconsin—Madison and the St. Paul Public Schools (SPPS) that have involved quasi-experimental and experimental evaluations of science magnet programs operating within the district. We have productively designed and implemented the quasi-experiment and randomized controlled trials and have completed both of the three-year evaluations. With the Senior Urban Education Research Fellowship, Dr. Borman and the leadership from the SPPS saw an opportunity to continue and extend this fruitful partnership.

The partnership between Professor Borman and SPPS addressed key methodological and substantive concerns within the district (and within most urban districts across the United States). The two key methodological issues are: (1) how to design high-quality quasi-experimental evaluations of already-existing policies and practices that are being implemented across the district; and (2) how to implement randomized controlled trials to evaluate and inform new policies and practices as they are being rolled out. From a substantive perspective, the projects focus on: (1) how to build teacher effectiveness and school-level capacity to improve student achievement; and (2) how to design new initiatives to help narrow existing achievement gaps between historically underserved students and their more privileged peers. These methodological and substantive shaped the interactions among Professor Borman and the curriculum and instruction and evaluation experts within the district. Beyond the two-year Fellowship, the experience should also build an enduring model of a proactive research process by which school district administrators can design evaluations to inform decision-making.

The Fellowship offered Dr. Borman the opportunity to gain new insights into the research needs of urban districts and provided the SPPS with improved empirical understandings of the impacts of its reform efforts. No matter how technically sound research activities might be, though, if they do not address the issues and questions that are of concern to education policymakers and practitioners, the research will not be used to inform education policy and practice. Professor Borman's role in working with the SPPS has been to work with the district leadership to identify the most pressing issues in need of evaluation, to develop rigorous research designs that fit the realities of the district context, and to develop information from the studies that is relevant and responsive to the stakeholders' initial questions and concerns. As Coburn, Honig, and Stein's (2006) lessons for increasing districts' data use suggest, two crucial factors include collaboration with external organizations and partners that can facilitate access to the "right evidence" and developing structures or processes to fund and support the search for evidence. The Fellowship provided the structured opportunity and resources necessary to bring both rigor and relevance to SPPS's most pressing questions. Two of the most important issues raised by evaluators, curriculum and instruction specialists, and other district leaders in the SPPS are related to (1) building teacher effectiveness and school capacity to improve achievement, and (2) closing achievement gaps. This report addresses the second of these two central policy issues.

EXECUTIVE SUMMARY

Following an established literature of laboratory findings and recent tentative findings from a handful of "real-world" field trials, this study assessed the impact of a 15-minute self-affirmation writing exercise on the achievement test scores of seventh- and eighth-grade students attending three St. Paul Public Schools (SPPS) middle schools. The affirmation exercises are intended to buffer students from the negative influence of "stereotype threat" (Steele & Aronson, 1995), or the apprehension individuals experience when confronted with a personally relevant stereotype that threatens their social identity or self-esteem. This largely unconscious anxiety can interfere with thinking and performance on standardized tests and other evaluative activities, and prevent students from performing up to their full potential. Stereotype threat has been shown to impede the academic performance of minority students in multiple content areas and of girls and women in science and mathematics.

In collaboration with SPPS, we distributed writing exercises to 1,628 middle school students. Half of the students were randomly selected to complete an affirmation exercise and half of the students completed a similar exercise that did not elicit self-affirmation. In a prior implementation in a single Northeastern middle school, African-American students who completed the affirmation exercise increased their grade point average by as much as 40% relative to students who completed the alternative writing exercise (Cohen et al., 2009; Cohen et al., 2006). We hypothesized that the affirmation exercise would be beneficial to all students to whom it was assigned, and that ethnic minority students and girls would especially benefit.

Implementation analyses revealed that 1318 (81 percent) of all seventh- and eighth-grade students who attended the three middle schools completed at least some of the writing exercise. "Intention-to-treat" analyses of students who were assigned the exercise revealed no impact on reading test scores and a positive overall impact on mathematics test scores, controlling for prior achievement.

After statistically accounting for student demographics—including ethnicity and gender— the main effect was no longer statistically significant.

However, stereotype threat and affirmation theory predict that disadvantaged subgroups of students benefit more than majority students. While subgroup and interaction analyses revealed no significant treatment impacts for ethnic minority students, they did reveal a positive treatment impact for girls of approximately one-tenth of a standard deviation on mathematics test scores. The statistical models showed that, on average, girls who completed the alternative exercise underperformed boys with similar demographic and prior achievement profiles on the mathematics achievement test, but that girls who completed the affirmation writing exercise performed as well as similar boys on the mathematics achievement test.

In sum, the experiment did not replicate the positive impacts for African-American students found in prior research (Cohen et al., 2006), but we did find that the affirmation writing exercises positively impacted girls' math test scores. Girls generally perform as well as or better than, boys on homework assignments and course grades in math and science classes (College Board, 2006; Shettle et al., 2007), but girls are still underrepresented in math and science fields (National Science Foundation, 2006) and boys tend to outscore girls when tested on the same content in high-pressure situations, such as standardized tests with time limits. These findings suggest that the affirmation writing exercises may help girls demonstrate their ability on standardized tests in mathematics.

INTRODUCTION

INTRODUCTION

CLOSING ACHIEVEMENT GAPS THROUGH SELF-AFFIRMATION

The key policy issue we addressed is closing the reading and math test score gaps between African American and Latino students and White students, and the math test performance gap between girls and boys. Of the various models and theories of these social inequalities that have been advanced in the literature, one particularly compelling line of research concerns the idea of *stereotype threat*. Claude Steele and Joshua Aronson, who coined the term in 1995, have defined stereotype threat as the apprehension individuals experience when confronted with a personally relevant stereotype that threatens their social identity or self-esteem. Steele and Aronson proposed that the phenomenon could help explain group differences in performance on standardized tests and in school. Stereotype threat is predicated on the notion that people often fear behaving in a way that fits the negative cultural image associated with a group stereotype, thereby marking them as inferior. This largely unconscious fear elicits anxiety and other counterproductive responses that can severely interfere with thinking and performance on standardized tests or other evaluative activities in the classroom.

Steele and Aronson (1995) tested their reasoning in a series of laboratory experiments. In the prototypical study in this series, African American and White college students were given a difficult section of a verbal Graduate Record Examination. Half of the participants were led to believe that the purpose of the test was to measure their intellectual ability; the others were told that the test was merely a non-evaluative laboratory exercise. Everything else about the situation was identical for the two groups, including the items on the test, the room in which they took the test, the experimenter, and so on. The results were striking. African American students who believed the test was being used to diagnose their abilities performed significantly worse than their peers in the non-diagnostic group. The difference in the way the test was described had no statistically significant effect on the White test takers; they performed equally well in both conditions.

These results have now been replicated by more than 300 independent laboratory studies and a growing number of field studies testing the stereotype threat phenomenon with females and mathematics, Latinos and verbal problem solving, and so on. More recent research has provided clear evidence that stereotype threat effects can and do occur in real-world environments (e.g., Cole, Matheson, & Anisman, 2007; Good, Aronson, & Inzlicht, 2003; Good, Rattan, & Dweck, 2007; Huguet & Régner, 2007; Keller, 2002; Keller & Dauenheimer, 2003; Kellow & Jones, 2005; Roberson, Deitch, Brief, & Block, 2003). A recent meta-analysis of stereotype threat studies, combining data from nearly 20,000 students (Walton & Spencer, 2009), suggested that standard measures of ability underestimate the true abilities of Black and Latino students by approximately one fifth of a standard deviation.

However, a number of research programs suggest that interventions aimed at reducing stereotype threat can attenuate its effects in school-based contexts (Cohen, Garcia, Apfel, & Master, 2006; Good et al., 2003; Walton & Cohen, 2007), yielding significant gains in test scores (see Aronson & McGlone, 2009; Aronson et al., 2008 for reviews).

The goal of our project was to combat the deleterious impacts of stereotype threat on academic performance. We assessed the impacts of a self-affirmation intervention relative to a neutral control group condition on the test scores of 1,628 middle-school students in the St. Paul Public School District. The stereotype threat–reduction intervention we studied involved a 15-minute writing exercise, which engages students in a values-based self-affirmation exercise. It derives from the work of Geoffrey Cohen, a psychologist from Stanford University, and his colleagues (Cohen et al., 2006; Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009). The control condition also involves an expressive writing exercise, but it asks students to consider and write about the values of others and is, therefore, not self-affirming. With the closing of achievement gaps as a significant local and national priority, this work has far-reaching implications for education policy and practice. This remarkably

cost-effective intervention strategy—involving a simple 15-minute writing exercise—could be replicated and implemented across the United States, with the potential to help significantly close the persistent achievement gaps within other school systems across the country.

Stereotype Threat Reduction: The Research

Self-affirmation Intervention. One general means of protecting the self from perceived threats and the consequences of failure is to encourage affirmation of self-worth. This can be done by asking individuals to think about the characteristics, skills, values, or roles that they value or view as important (Schimel, Arndt, Banko, & Cook, 2004). Frantz, Cuddy, Burnett, Ray, and Hart (2004), for example, showed that Whites who were given the opportunity to affirm their commitment to being nonracist were less likely to respond in a stereotypical fashion to an implicit measure of racial associations that had been described as indicative of racial bias. Martens, Johns, Greenberg, and Schimel (2006) provided evidence that encouraging women to self-affirm eliminated performance decrements that typically arise when stereotypes about gender differences in mathematics and spatial ability are invoked.

These effects are not limited to the laboratory. Indeed, one field-based trial of self-affirmation exercises in particular cries out for replication because it is relatively simple yet effective and has, as a consequence, garnered a great deal of both attention and skepticism. In results published in *Science*, Cohen and colleagues (Cohen et al., 2009; Cohen et al., 2006) reported that brief self-affirmation tasks aimed at affirming students' personal values reduced the Black–White grade-point-average (GPA) gap by as much as 40%, by improving African American students' performance. Specifically, having middle-school students write essays affirming their personal values as little as just *once* during the school year (treatment students randomized to additional exposure to the writing exercises did not achieve any additional benefits) improved academic performance in four core academic subjects: science, social studies, math, and English. The two suburban Northeastern middle schools in which the research took place were approximately half African

American and half White in racial/ethnic composition. The studies examined effects separately for White and African American students and for low- and high-achieving students. The largest gains were concentrated among the lowest performing African American students.

In these studies, seventh graders were placed at random into intervention and comparison groups near the start of the school year. Both groups were given structured 15-minute writing assignments a total of three to five times over two academic years, seventh- and eighth-grade. These assignments primed the students to think about important qualities and values; students selected three values (e.g., sports talent, family, sense of humor) and explained in writing why they were personally important. Compared to control group students (who wrote about other non-self-affirming topics, such as their daily routine or why values they considered unimportant might be important to others), the affirmation group performed significantly better as reflected in both grades and test scores.

These types of affirmations, which Creswell et al. (2005, 2007) have shown reduce physiological anxiety, appear to be particularly salient for reducing the negative consequences of stereotype threat among the lowest performing minority students. The 2006 study by Cohen and colleagues showed short-term effects during a single 3-month period. The 2009 study by the Cohen research team demonstrated the long-term effects of the intervention, as the short-term effects and gains were maintained across the two-year follow-up (Cohen et al., 2009). This research—which was reviewed in February 2010 by the What Works Clearinghouse (WWC) and deemed consistent with WWC evidence standards (see WWC, 2010)—demonstrated that African American students who completed the writing exercises about their values increased their average seventh- and eighth-grade GPA by a quarter of a letter grade (0.24 points), a change that was statistically significant and equivalent to an effect size of approximately $d = 0.30$. Among low-achieving African American students, the effect was somewhat larger, an increase in average seventh- and eighth-grade GPA of 0.41 points. In addition, the intervention reduced

the likelihood that low-achieving African American students would be assigned to a remedial program or retained in grade (5 percent *versus* 18 percent). The intervention did not have a statistically significant effect on the academic outcomes of White students.

The condition to which each student was randomized was not known to the teachers. The fact that teachers were blind to the students' assigned condition alleviates some concern that teacher expectancy effects might explain some of the differences observed on GPA. Cohen et al. (2009) concluded: "Findings suggest that because initial psychological states and performance determine later outcomes by providing a baseline and initial trajectory for a recursive process, apparently small but early alterations in trajectory can have long-term effects" (p. 400). By enhancing students' feelings of personal worth, the authors surmised, the exercise changed their perception of bias at school and shifted how they interpret their academic successes and failures. These steps particularly protected African American students who had been struggling in school. Rather than feeling discouraged and plunging further into a downward spiral, it was as though the psychologists gave Black students an inoculation against the threat inherent in the stereotypes. Beyond the key effects on GPA, unpublished data from the first author suggests that the impact of the intervention on African Americans' standardized test scores was between $d = 0.20$ and $d = 0.30$, or the equivalent of an 8 to 12 percentile point advantage (G. Cohen, personal communication, June 9, 2010).

Cohen and colleagues' study of this intervention was fielded in only two Northeastern suburban middle schools with three relatively small cohorts ($N = 133$, 149, and 134), and only two replications have been attempted. One—a 2009 study involving medical students in the United Kingdom—found that the achievement gap did significantly close, but by lowering scores of White students rather than raising those of Blacks (Woolf, McManus, Gill, & Dacre, 2009). These results from the U.K. certainly do not generalize well to students in U.S. K–12 education systems. A second, IES-funded replication study in Philadelphia middle schools by Thomas

Dee was recently completed, and preliminary results are forthcoming. Thus, this modest, remarkably cost-effective intervention begs for replication and further empirical and theoretical research. If its surprisingly strong results can be replicated and better understood, the intervention could hold tremendous potential for closing achievement gaps in the St. Paul school system and beyond.

The theory behind the intervention we studied is that it acts like a *catalyst* (Purdie-Vaughns et al., 2009). As Garcia and Cohen (in press) argue, the interventions "permit the positive forces in school to assert a fuller impact." That is, the affirmations reduce stress, allowing people's skills to be more completely displayed and the institution's resources to have greater impact. In addition, by reducing stress and test anxiety, the self-affirmation can free up psychological resources that, in the short term, can help improve student performance on academic tasks, including performance on high-stakes standardized tests.

Consistent with this theory, we hypothesized that the intervention would have proximal (and longer term) impacts on stereotype threat vulnerability. We expected to find proximal impacts on the Minnesota state reading and math tests, which were administered in seventh and eighth grade following the students' participation in the self-affirmation and control writing exercises.

As these proximal, relatively "quick wins" accumulate, the developers note that *recursive processes acting like chain reactions* then carry forward the initial effects of the intervention (Cohen et al., 2009). As Purdie-Vaughns et al (2009) argue, a small improvement early in the year due to the intervention might, for example, give children a little extra confidence, and this confidence might lead to further gains in performance, in a potentially repeating cycle that sustains their performance for a long time. Likewise, small early improvements in children's performances may cause their teachers to see them as more able and worthy of attention and mentoring. Purdie-Vaughns and her colleagues suggest that the effects of teacher expectancies could then assert themselves, acting as channels that carry forward and amplify the effects of the intervention. Consistent with these ideas,

we suggest that broader and more distal outcomes, including a successful transition to high school, or ninth grade could be of importance during the years beyond the project. We hope to explore ways in which we might study whether broad indicators of school success—including positive impacts on grade point average, and reductions in both retentions in grade and special education placements—might persist during the years beyond this study.

The Psychology Behind the Intervention. As we have discussed, the affirmation exercises are designed to help buffer students from the deleterious effects of a personally relevant stereotype that threatens their social identity or self-esteem. The classic work of Steele and Aronson (1995), and the many other researchers who have followed in this tradition, demonstrates that stereotype threat can undermine the performances of groups that have historically underperformed during testing situations and other school-based and classroom-based academic situations. In a review, Aronson (2002) noted that perceptions of negative stereotypes lead many individuals to engage in activities such as self-handicapping (Smith, 2004), challenge avoidance (Good, Aronson, & Inzlicht, 2003), self-suppression (Steele, 1997; Pronin, Steele, & Ross, 2002), and disidentification or disengagement with the task or the context in which the task is to be performed (Steele, 1997; Aronson, 2002; Major et al., 1998). In addition to these poor academic performance correlates, stereotype threat has also been linked to high blood pressure among African Americans (Blascovich et al., 2001), altered career and/or professional aspirations and belonging (Steele, James, & Barnett, 2002), and social distancing, particularly from the stigmatized social group of which the participants are members (Pronin, Steele, & Ross, 2002).

These psychological and behavioral outcomes that have been found among African American, Latino, and women students are not typically the result of negative stereotypes being communicated directly to them from others within the given social context. Rather, these behaviors typically result from exposure to a context in which historically (a) the performance of a given group

is evaluated and compared with that of others, (b) such performance has been valued by the group and the larger society, and (c) the performance of one's group has been consistently negatively evaluated and, thus, stereotyped more than other groups.

In a recent article by Schmader, Johns, & Forbes (2008), the authors developed an integrated process model of stereotype threat effects on performance. They argued that stereotype threat disrupts performance via three distinct, yet interrelated, mechanisms: (a) a physiological stress response that directly impairs prefrontal processing; (b) a tendency to actively monitor performance; and (c) efforts to suppress negative thoughts and emotions in the service of self-regulation. Providing empirical evidence to support their assertions, Schmader and colleagues concluded that these mechanisms combine to consume executive resources needed to perform well on cognitive and social tasks, with the active monitoring mechanism disrupting performance on sensorimotor tasks directly. In summary, stereotype threat is a phenomenon that generally is not recognized by the affected individual, but is manifested through various other behavioral, physiological, and psychological outcomes that impede students' abilities to perform academically to their full potential.

The theory of change supporting the intervention we studied suggests that students who are provided the needed psychological resources through self-affirmation to help buffer them from stereotype threat will experience fewer of the physiological and psychological mechanisms outlined above that are associated with poor academic performance. The self-affirmation exercise helps students overcome stereotype threat by protecting them from perceived threats and the consequences of failure. Affirming one's self-worth in other valued domains helps students move past the group differences that they may associate with academic inferiorities and, instead, emphasize positive traits that affirm students' overall self-worth.

INTRODUCTION

Indeed, we all face failure and self-threat in a variety of ways throughout our day-to-day lives. These may include a poor performance on the job or in class, not achieving one's goals, information that challenges the validity of long-held beliefs, illness, the defeat of one's political party in an election or of one's favorite sports team in a playoff, scientific evidence suggesting that one is engaging in risky health behavior, negative feedback at work or in school, rejection in a romantic relationship, real and perceived social slights, interpersonal and inter-group conflict, the loss of a loved one, and so on. In the course of a given day, the potential number of events that could threaten people's "moral and adaptive adequacy," or their sense of themselves as good, virtuous, successful, and able to control important life outcomes (Steele, 1988), seems limitless and likely to exceed the small number of events that affirm it. A major undertaking for most people is to sustain self-integrity when faced with the inevitable setbacks and disappointments of daily life. How do individuals adapt to such threats and defend self-integrity?

Research suggests that people have a "psychological immune system" that initiates protective adaptations when a real or impending threat is perceived (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998). When such a threat is recognized, people typically respond in one of three ways in order to cope. First, people sometimes respond by accommodating to the threat. That is, they accept the failure or the threatening information and then use it as a basis for changing their attitudes or behavior. However, when the threatened domain concerns an important part of one's identity, the need to maintain self-integrity can make it difficult to accept the threatening information and to change one's attitude or behavior in response.

A second reaction, thus, involves ameliorating the threat *via* underline direct psychological adaptations. Some direct adaptations preserve the fundamental informational value of the event while also changing one's construal of that event—such as framing a failure as a learning opportunity (Dweck & Leggett, 1988). Typically, though, direct psychological adaptations are defensive in nature, in that they involve dismissing, denying, or avoiding the threat in some way. Sherman and Cohen (2002) refer to these responses as defensive biases. Although a defensive bias can restore self-integrity, the rejection of the threatening information can lessen the probability that the person will learn from the potentially important information transmitted through the event.

Yet, these normal adaptations can be "turned off" through a different psychological adaptation to threat that does not depend on distorting or denying the threatening event to render it less significant. Self-affirmation theory suggests a third alternative, a different kind of psychological adaptation— one that, under many circumstances, enables both the restoration of self-integrity and adaptive behavior change. In this instance, people respond to a threat using the indirect psychological adaptation of affirming other self-resources unrelated to the provoking threat. Such "self-affirmations" may include reflecting on important aspects of one's life irrelevant to the threat, or engaging in an activity that makes salient important values unconnected to the threatening event. In this way, indirect psychological adaptations, like self-affirmation, allow people to focus on domains of self-integrity unrelated to the threat (Cohen et al., 2006).

When self-affirmed in this manner, people can come to realize that their overall self-worth does not hinge on the evaluative outcomes or implications of the immediate situation. As a result, they have less need to distort or re-construe the provoking threat and can respond to the threatening information in a more open and even-handed manner. For instance, a student-athlete who performs poorly at football practice may remind himself that, regardless of the negative event related to football, he is still a good student. The student-athlete's reflection on other valued aspects of the self not immediately related to the events on the football field—in this case, being a good student—can help buffer his overall self-integrity from the threat of a bad day in another domain. Similarly, when consciously or unconsciously threatened by stereotype threat when engaged in an evaluative academic activity such as a test, a student may reflect on other domains that they value and find affirming—such as being a good friend, being creative, or being a good athlete. By doing so, students threatened by such stereotypes may reduce their anxiety, restore their self-integrity, and, ultimately, perform better on the academic task at hand.

No interventions to reduce stereotype threat have been implemented in the St. Paul district to date, but district leaders agreed to work with our team to implement a randomized trial of the affirmation writing intervention in the participating schools. Based on both long-established laboratory findings and newer, more tentative findings from a handful of field trials, we expect the self-affirmation intervention to hold considerable promise for unseating some of the maladaptive academic behaviors identified by researchers as undermining African American and Latino students' academic progress and discouraging girls' achievement in math and science.

Underlying these interventions is an assumption that the underperformance of these students is significantly rooted in environmental and cultural factors. Over time, these environmental forces may lead to less cognitive development among minority students relative to their White counterparts or the disengagement of girls from science and math, but these deficits are taken to be the result, not the origin, of the learning and performance gap.

Evidence from the 2007 National Assessment of Educational Progress (NAEP) suggests that the achievement gaps in Minnesota are as large or larger than the gaps found across the other states of the nation (Vanneman, Hamilton, Baldwin-Anderson, & Rahman, 2009). Inequalities characterize student outcomes in this locale, as they do throughout the country, and policy makers and educators in the school district are highly motivated to address them. The randomized trial may help guide future decisions within the district about the dissemination of the self-affirmation intervention and others like it to reduce the academic performance gaps that separate negatively stereotyped students from their non-stereotyped classmates.

METHODOLOGY

METHODOLOGY

SPPS officials recruited four middle schools to participate in the study, and three schools eventually participated. Within each of the three participating schools, we assigned at random half of the seventh- and eighth-grade students to a receive a self-affirming writing exercise in which students selected two or three values or characteristics from a list that were important to them, such as being a good athlete or being a good friend. The students were then asked to write a brief expressive essay about the values they selected. The other half of the students received a similar intervention in which they were asked to identify two or three values from the list that were not important to them, and then to write about why the values or things may be important to someone else. These exercises were administered as a replication of the protocol previously used by Cohen et al., (2009; 2006). In collaboration with SPPS stakeholders, this study was designed to understand whether the self-affirmation writing exercise could close achievement gaps on the spring 2011 seventh- and eighth-grade MCA-II reading and math tests.

PRIMARY RESEARCH QUESTIONS

Our randomized trial of the self-affirmation intervention designed to combat stereotype threat addressed the following primary research questions:

- Does the intervention have an impact on seventh- and eighth-grade students' reading and math test performance relative to the control condition?

- Does the intervention close the reading and math achievement gaps between minority and White seventh- and eighth-grade students?

- Does the intervention close the math achievement gaps between seventh- and eighth-grade boys and girls?

- Are the observed impacts consistent across the three schools, or is there evidence that the effects vary by school context?

In addition, we present information from the self-affirmation exercises themselves to help describe the nature of the students' responses, including the types of values that students chose to write about.

DATA

Dependent Variables. The outcomes, or dependent variables, are the MCA scores in reading and mathematics. Our initial census included students who were administered modified achievement assessments (MCA-Modified) for persistently low achievers who receive special education services as well as students with cognitive disabilities who were administered an alternative test (the Minnesota Test of Academic Skills, or MTAS). We omit students who took the MCA-Modified and the MTAS from this analysis because the tests are scaled differently than the MCA.

The MCA is scaled to provide a valid comparison from year to year in a given grade and subject (e.g., grade seven reading current *versus* prior year) (MN Interpretive Guide 2010-2011: 9). The scale for seventh grade is from 701 to 801 and the scale for eighth graders is from 801 to 899. Moreover, the mathematics scale was changed somewhat in 2010-2011. Consequently, we normalize the scale scores using the sample mean and standard deviation within each grade to produce a standardized outcome variable with a mean of zero and a standard deviation of 1.

Independent Variables. The explanatory, or independent, variables include, most importantly, the student's assignment to receive the self-affirmation (treatment) writing exercise or the non-affirming control exercise. We used a block randomized design, in which random assignment of students was conducted within school and grade level. This method assured that all schools, and grade levels within them, would have the same proportions of students participating in each of the two experimental conditions.

In addition, using extant district data, we obtained the prior year MCA score in the same subject areas and student demographic variables, including the gender, free/reduced lunch eligibility, limited English proficiency, special education status, and race/ethnicity. The racial/ethnic categories in the SPPS administrative data include American Indian, African-American, Asian-American, Hispanic, and Caucasian. With the exception

of the American Indian group, the ethnic categories are substantial enough to conduct analyses within the groups. Descriptive statistics for the full sample and the analytic sample are reported in Table 1.

SAMPLE

Study Schools. SPPS has ten middle schools. Each school has a designated attendance area, and eight of the ten schools draw students from throughout the city through the SPPS magnet program. The two remaining schools draw only from the attendance area in which they are located but demand for them is substantial and they do not admit all applicants. Of the three schools in our sample, two are citywide magnet schools and one was a local attendance school. School #1 is a science magnet with single-sex instruction, School #2 is an arts magnet, and School #3 is the most traditional.

Student Sample. Based on the rosters provided by the schools and district, we identified a census of 1628 students from the seventh and eighth grades at the three participating schools. We randomized students individually within schools and grade to either of two writing exercises. That is, within the three middle schools, each had a participating seventh- and eighth-grade group. As a result, across the three schools and two grade levels, there were a total of six distinct randomization pools in which half of the students were assigned to each exercise. Assigning students within grade level by school, we obtained 816 treatment, 50.1%, and 812 control, 49.9%, students. These 1628 students represented our group of students selected to receive the exercises. In other words, this group represented our "intention-to-treat" sample.

PROCEDURE

The exercises were generously provided by Geoffrey Cohen. The first exercise was the affirmation writing exercise (most important values) and the second exercise was the "control" condition (least important values). The participating teachers were provided a general overview of the study, including a script for introducing the writing exercise to students and suggested responses to questions students might ask. All student and teacher materials were the same as those used by Cohen et al. (2006). The teachers were blind both to the overall intent of the study and the specific treatment status of individual students. Students received their writing assignments in closed envelopes, and these assignments were designed to be self-directed. The treatment and control assignments, which are described in more detail below, also had nearly identical three-

TABLE 1. AVAILABLE DATA

	READING	%	MATH	%
ORIGINAL SAMPLE	**1,628**	**100%**	**1,628**	**100%**
2011 TEST SCORE	**1,577**	**97%**	**1,589**	**98%**
2011 STANDARD FORMAT TEST	**1,488**	**91%**	**1,508**	**93%**
2010 PRIOR YEAR STANDARD FORMAT TEST	**1,319**	**81%**	**1,322**	**81%**
DEMOGRAPHICS AVAILABLE	**1,318**	**81%**	**1,322**	**81%**

page layouts printed in booklet form. The delivery of the sealed student packets, the design of the booklets, and the similarity of the two exercises imply that the status of individual students could not be easily observed by teachers or other students. The student packets were sorted by classroom, and delivered by the researchers to language arts coordinators at the schools on March 8, 2011. All seventh- and eighth-grade students were given one of the two exercises. The exercises were given to the students by their classroom teachers and completed during language arts class. The students completed the exercises just before the Minnesota Comprehensive Assessment (MCA) testing window opened during the last week of March.

In both the treatment and the control conditions, each three-page booklet began by asking students to read a list of "personal values." This list included 11 values: Sports Ability; Sense of Humor; Religious Values; Relationships with Friends and Family; Politics; Music; Membership in a Social Group (such as a community or school club); Independence; Creativity; Being Smart or Getting Good Grades; and Being Good at Art.

Students assigned to the treatment condition were asked to circle the two or three values "most important" to them. On the second page, the treatment students were then asked to write a "few sentences" to "describe why these values are important" to them. In the control condition, students were asked to circle the two or three values from the list "least important" to them. The expressive essay portion of the exercise then asked control students to write about why *someone else* might find those values important. The third and final page of the exercise presented treatment students with four statements (e.g., "These values have influenced my life" and "I care about these values") to which they were to indicate their level of agreement. For control students, the third page asked them to indicate their level of agreement with four statements about how others viewed these values (e.g., "These values have influenced some people" and "Some people care about these values."). For treatment students, the expressive writing exercise and their responses to the four statements were intended to reinforce the affirmation of the students' selected values.

RESULTS

RESULTS

Our results speak primarily to the effects of the intervention on students' MCA reading and math scores. However, several other relevant questions emerge that may shape our interpretation of these effects. First, how effectively were the writing assignments implemented? Second, how representative are the final analytical samples of seventh and eighth graders at the three participating schools, and as a result, how might we generalize these results to all SPPS seventh and eighth graders?

We begin by examining the extent to which students actually completed the writing exercise and, by implication, experienced the treatment as intended by the study design. Next, we consider how our analysis sample compares to the original group of 1628 students across the three schools. Specifically, did data attrition limit the internal or external validity of our results? Finally, we consider the nature of the impacts of the intervention on students' math and reading MCA scores.

Implementation. For the total of 1628 seventh- and eighth-grade students, 1410 packets, 706 treatment (50.1 percent) and 704 control (49.9 percent), were returned by the three schools. Of the undistributed 218 packets, 162 were not distributed due to students' absences (74 treatment and 88 control), and 56 were not distributed for other reasons (36 treatment and 20 control). Of the 1410 distributed packets, 1318 were returned with at least one value selected from a student, 652 (49.5 percent) treatment and 666 (50.5 percent) control. These data indicate that a fairly high percentage of the students who were randomly assigned at baseline did meaningfully engage in the writing assignment. That is, 1410, or approximately 87 percent, of the 1628 packets were distributed to the intended students. Of those 1410 receiving the packets, approximately 94 percent had complete responses to the first-page questions on values and completed the subsequent short writing assignment.

The following selection of excerpts provides some fairly typical examples of how students responded to the self-affirming treatment exercise and the non-self-affirming control exercise:

> Treatment girl: "The values I picked are very important to me. Being smart and getting good grades mean a lot to me because being smart doesn't always mean being smart at school. It could mean making smart choices…"

> Treatment girl: "The values I chose are important to me because when you have a sense of humor, people will look at you as a lively person and everybody will like you for who you are…"

> Treatment boy: "I think music is very important to me because music is awesome it sometimes relates to life or it is kinda relates to life in so many ways. And music is like your free to do things and it just releases stress!…"

> Control girl: "Politics is not important to me but politics are important to my mom. I think this important to her because I am really young and really don't care about politics…"

> Control boy: "The least important value for me is being good at art and having good sense of humor. These 2 value may be important for someone else because having these values could help them with their hopes and dreams. Like as a comedian or an artist…"

> Control boy: "Politics I don't really even care about but I don't even know what they are. Politics can be important to other people who know what they are because they care what is happening in the world…"

In addition to these responses, Figures 1 through 4 provide more information about the substance of the students' responses. Figure 1 displays the frequencies of treatment and comparison students who circled varying frequencies, from zero to six, of values. Treatment students tended to select more values than comparison students, and more uniformly selected three values.

Figure 2 shows the percentages of treatment and comparison students who selected each of the 11 "values" listed on the exercises. Most treatment students—71 percent—indicated that friends and family was the value that was most important to them, while art (52 percent) and politics (45 percent) were most frequently identified by comparison students as areas that were unimportant to them.

Figures 3 and 4 show the values selected by participating girls and boys, respectively.[1] These figures suggest some differences of note. For instance, boy treatment students indicated the value of sports more often than girls, and girls indicated the value of music more often than boys.

FIGURE 1. NUMBER OF VALUES SELECTED BY TREATMENT AND COMPARISON STUDENTS

COMPARISON (EXERCISE 2)

TREATMENT (EXERCISE 1)

N=1410

Number of Values Chosen

Percent of Students

---
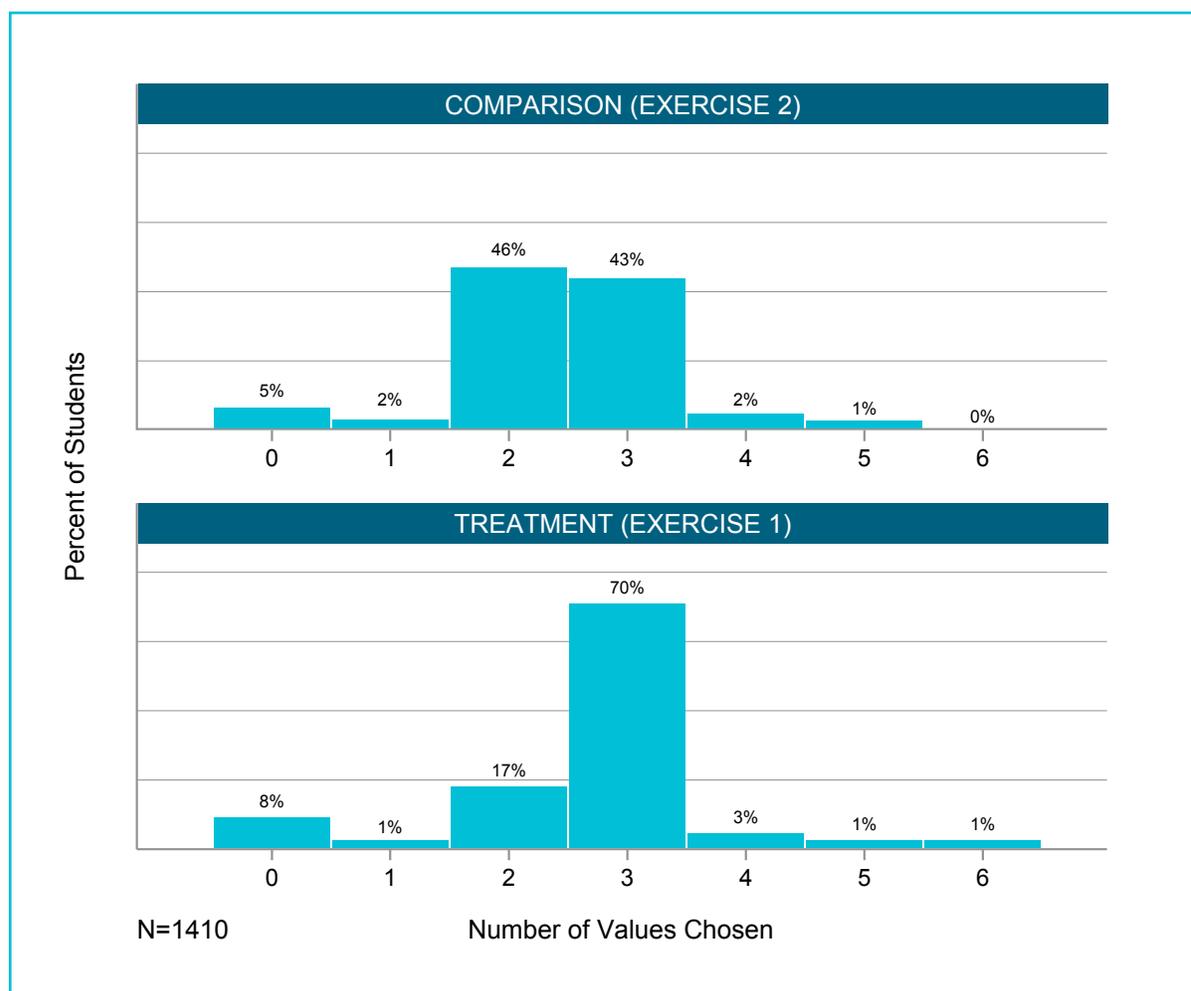
1 Of the 1318 students who submitted complete responses, 5 students could not be located in the district's testing files, which was used to identify boys and girls. These 5 students were omitted from the data presented in Figures 3 and 4.

RESULTS (CONT'D)

FIGURE 2. VALUES SELECTED BY TREATMENT AND COMPARISON STUDENTS

COMPARISON (EXERCISE 2)

27% | 12% | 27% | 8% | 45% | 20% | 30% | 11% | 9% | 12% | 52%

TREATMENT (EXERCISE 1)

35% | 29% | 16% | 71% | 2% | 38% | 4% | 16% | 23% | 46% | 9%

Sports | Social Group | Religion | Independent
Humor | Creative | Friends & Family | Smart/Good Grades
Politics | Art | Music

N=1318
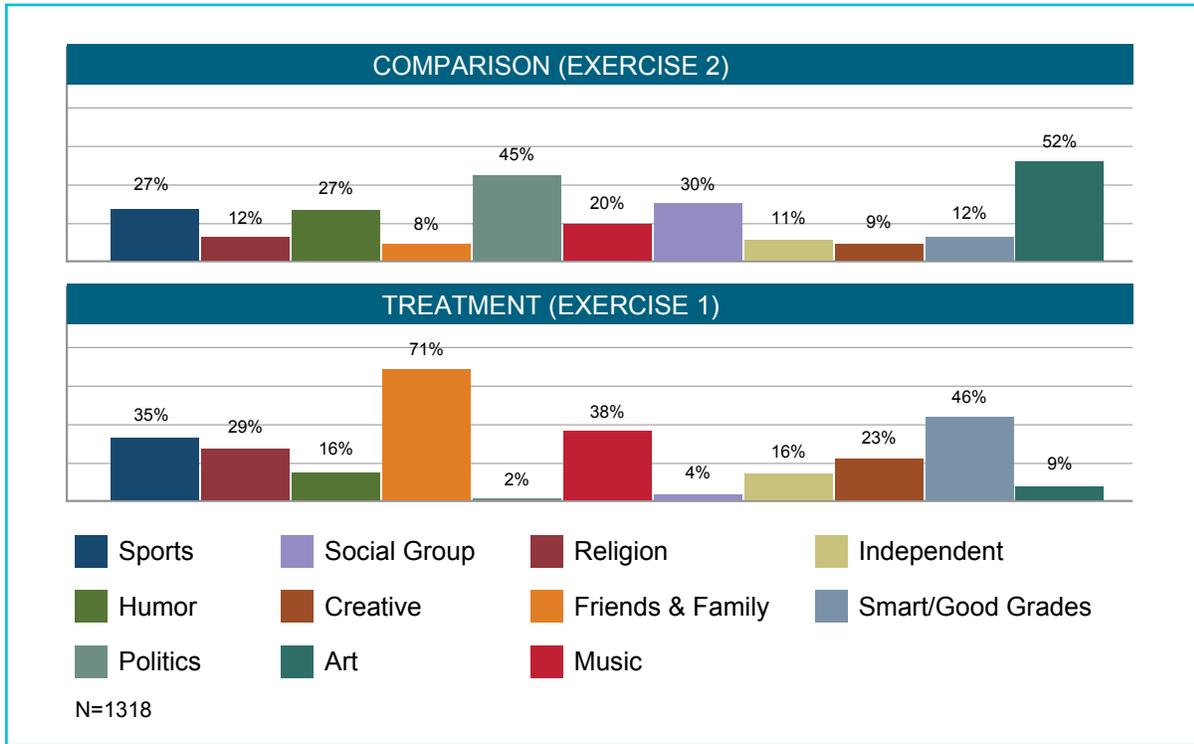
FIGURE 3. VALUES SELECTED BY FEMALE TREATMENT AND COMPARISON STUDENTS

COMPARISON (EXERCISE 2)

29% | 12% | 29% | 10% | 48% | 15% | 30% | 10% | 10% | 12% | 49%

TREATMENT (EXERCISE 1)

22% | 27% | 16% | 76% | 3% | 46% | 3% | 20% | 24% | 48% | 7%

Sports | Social Group | Religion | Independent
Humor | Creative | Friends & Family | Smart/Good Grades
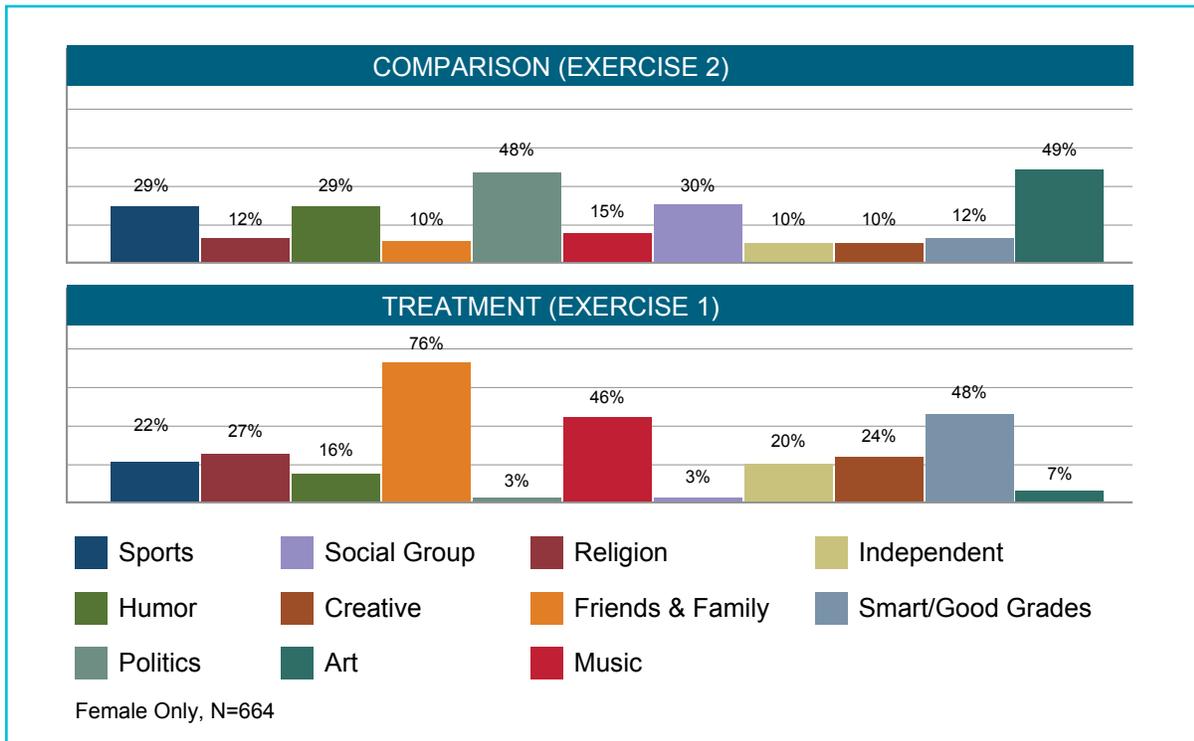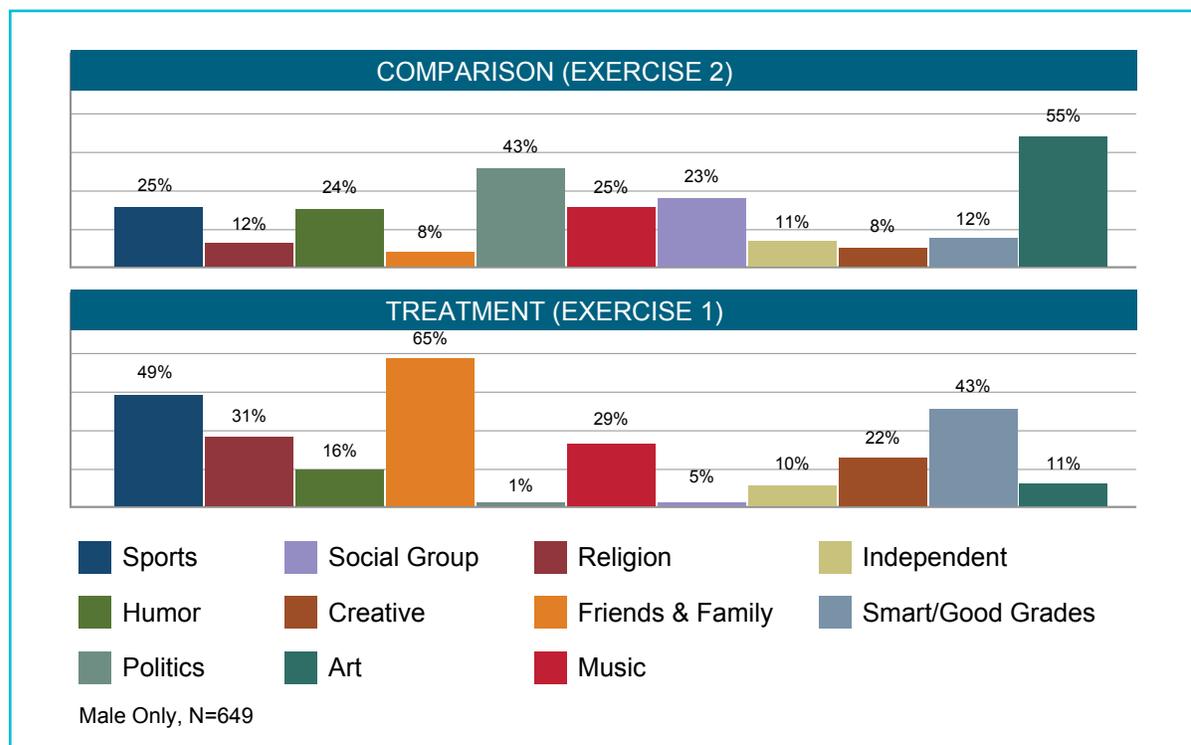Politics | Art | Music

Female Only, N=664

FIGURE 4. VALUES SELECTED BY MALE TREATMENT AND COMPARISON STUDENTS

Analytical Sample. With respect to how our final analytical sample compares to the original group of 1628 students targeted for the study, we consider both the reasons for and rates of data attrition (i.e., students dropping from our sample due to missing data) and the demographic and academic characteristics of our final sample. The reasons for missing data are reported in Table 1. Test scores from 2011 were available for 97 to 98 percent of the students in the initial sample, but 89 students took an alternate form of the reading test and 81 students took an alternate form of the math test, leaving 91 percent of the original sample in reading and 93 percent of the original sample in math. Prior year scores on the standard form of the test were available for 81 percent of the initial sample. One student in the reading sample with otherwise complete data was missing demographic information, leaving 1,318 students with reading test scores and 1,322 students with mathematics test scores. These represent our final analytical samples for the respective analyses of reading and math MCA test scores.

With regard to the demographics and academic backgrounds of the students with and without complete data for analysis, Table 2 reveals that the analytic sample is higher achieving, slightly less poor, receives fewer special education services, and is more likely to attend School #3 than the overall sample. These differences are due to restricting our analysis to the standard form of the MCA and to requiring that students have a prior year test score. The characteristics of students in the analytical sample at baseline, prior to the writing exercises, are shown in Table 3. The two samples are balanced in all respects but one: there is a statistically significantly greater proportion of girls in the treatment group than control group (53 percent versus 46 percent).

RESULTS (CONT'D)

TABLE 2. DESCRIPTIVE STATISTICS OF FULL AND ANALYTIC SAMPLES

VARIABLE	FULL SAMPLE MEAN	OBSERVATIONS	ANALYTIC SAMPLE MEAN	OBSERVATIONS
2011 READING SCORE	0.00	1,488	0.05	1,318
PRIOR YEAR READING SCORE	0.00	1,394	0.08	1,318
2011 MATH SCORE	0.00	1,508	0.10	1,307
PRIOR YEAR MATH SCORE	0.00	1,387	0.07	1,311
TREATMENT GROUP	0.50	1,628	0.49	1,318
GRADE 8	0.46	1,628	0.46	1,318
FEMALE	0.49	1,609	0.49	1,318
FRL ELIGIBLE	0.75	1,607	0.72	1,318
LIMITED ENGLISH PROFICIENT	0.20	1,607	0.29	1,318
SPECIAL EDUCATION	0.30	1,607	0.14	1,318
SCHOOL 1	0.42	1,628	0.38	1,318
SCHOOL 2	0.11	1,628	0.10	1,318
SCHOOL 3	0.47	1,628	0.52	1,318
AMERICAN INDIAN	0.02	1,604	0.02	1,318
ASIAN	0.25	1,604	0.26	1,318
AFRICAN-AMERICAN	0.32	1,604	0.30	1,318
HISPANIC	0.15	1,604	0.14	1,318
CAUCASIAN	0.26	1,604	0.28	1,318

Notes: Analytic sample reported for Reading (Math sample N = 1322)

TABLE 3. BALANCE BETWEEN TREATMENT AND COMPARISON GROUPS

VARIABLE	TREATMENT MEAN	COMPARISON MEAN	P-VALUE
PRIOR YEAR READING SCORE	0.09	0.06	0.574
PRIOR YEAR MATH SCORE	0.04	0.09	0.264
GRADE 8	0.46	0.46	0.888
FEMALE	0.53	0.46	0.008**
FRL ELIGIBLE	0.72	0.73	0.717
LIMITED ENGLISH PROFICIENT	0.28	0.3	0.464
SPECIAL EDUCATION	0.13	0.15	0.208
SCHOOL 1	0.38	0.38	0.88
SCHOOL 2	0.10	0.10	0.489
SCHOOL 3	0.52	0.52	0.698
AMERICAN INDIAN	0.02	0.02	0.652
ASIAN	0.26	0.26	0.832
AFRICAN-AMERICAN	0.29	0.32	0.378
HISPANIC	0.14	0.14	0.936
CAUCASIAN	0.29	0.26	0.299

Notes: Statistics from analytical sample in reading (T = 652, C = 666) except for prior year math score (T = 685, C = 664). Test score p-values from two-sample t tests; all other p-values from two-sample tests of proportions.

Results of Impact Analyses. We used ordinary least squares (OLS) regression analysis to assess the experimental effects of the intervention. Put simply, we estimated the non-directional achievement score difference (two-tailed hypothesis test) between treatment (self-affirmation) and control (non-affirming) groups. These analyses respond to the question: What impact did the stereotype threat–reduction exercise have relative to a neutral writing exercise?

The analyses are performed on data from the entire sample of 1,318 students with reading test scores and 1,322 students with mathematics test scores, but possibly more important questions are to be answered through the analysis of data from student subgroups—notably, African American, Latino, and White students and male and female students. These subgroup analyses will allow us to address questions concerning the relative effects of the intervention for subgroups for whom stereotype vulnerability is more salient and to speak to the reduction of performance gaps by examining the interventions' effects for these subgroups as compared to those for non-threatened White students and for non-threatened male students in the math domain.

The model we formulate to examine the potential effect of assignment to one condition versus another (intention-to-treat) is the following regression model:

$$Y_i = a + \beta_p T_i + \sum \beta_q X_{qi} + \varepsilon_i$$

where Y_i is an individual student's MCA scale score in reading or mathematics, T_i is a dichotomous indicator of assignment to the self-affirmation exercise or control, β_p is the effect of that assignment on a student's test score, and X_{qi} are a set of additional covariates. The covariates include dummy indicators of the schools and student grade level, which together formed the randomization blocks, as well as a prior-year MCA reading or math score. In addition, we model students' special education status, free or reduced lunch status (as a proxy for poverty), limited English proficient status, gender, and ethnicity.

We use these covariates to help us predict, or explain, random variability in the reading and math scores. By predicting this random variation in our achievement outcomes, we are able to reduce the unexplained differences, or "noise," across the students in our sample and improve our statistical power and precision to detect the difference that is at the heart of our analysis: the "signal" of the effect of the treatment.

Also, as in prior analyses by Cohen et al. (2006), we test for ethnicity-by-treatment and gender-by-treatment interactions. With the sample of three schools, we also test for possible school-by-treatment interactions. These analyses, respectively, tell us whether the impacts of the self-affirmation are different across: (1) students of different racial/ethnic backgrounds; (2) boys and girls; and (3) the three participating schools.

The results are reported in Table 4 for the entire sample and in Table 5 for separate male and female subgroups. In Table 4, we present a series of five OLS regression models by each outcome, math and reading. The five models within each domain begin with only a treatment assignment indicator, Model 2 adds the prior year test score, Model 3 adds the students' pre-intervention GPA, Model 4 adds the school indicators, and Model 5 adds the complete set of student covariates.

As the results in Table 4 indicate, there is no evidence of a statistically significant treatment effect without taking into account prior performance (the pretest), but after controlling for prior performance the magnitude of the positive effect of the self-affirmation writing exercise increases and the standard error of the impact estimate decreases. The covariate-adjusted results through Model 4 suggest a positive overall impact of the self-affirmation exercise on mathematics achievement but not for reading. However, after controlling for student demographics, including ethnicity and gender, the positive effect in mathematics is no longer statistically significant. After finding limited evidence of a main effect for the overall sample, we turned to analyses of potential differences in treatment effects across gender and racial/ethnic subgroups

TABLE 4. RESULTS OF MAIN EFFECTS MODELS

	READING					MATH				
	MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5	MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5
TREATMENT CONDITION	0.038 (0.054)	0.013 (0.030)	0.007 (0.029)	0.007 (0.029)	-0.004 (0.029)	0.022 (0.053)	0.072* (0.029)	0.058* (0.027)	0.056* (0.026)	0.050 (0.026)
PRIOR YEAR TEST SCORE (STANDARDIZED)		0.848*** (0.016)	0.751*** (0.017)	0.743*** (0.018)	0.684*** (0.021)		0.847*** (0.015)	0.716*** (0.017)	0.701*** (0.017)	0.654*** (0.019)
PRE-INTERVENTION GPA (STANDARDIZED)			0.192*** (0.017)	0.192*** (0.017)	0.184*** (0.019)			0.220*** (0.017)	0.217*** (0.017)	0.222*** (0.018)
SCHOOL INDICATORS	NO	NO	NO	YES	YES	NO	NO	NO	YES	YES
DEMOGRAPHICS	NO	NO	NO	NO	YES	NO	NO	NO	NO	YES
CONSTANT	0.033 (0.038)	-0.022 (0.021)	-0.024 (0.020)	0.006 (0.025)	0.090* (0.044)	0.085* (0.037)	0.005 (0.020)	0.005 (0.019)	0.082*** (0.023)	0.192*** (0.039)
BIC	3690.3	2189.3	2070.7	2078.8	2095.1	3663.0	2034.6	1876.8	1834.9	1864.9
OBSERVATIONS	1318	1318	1318	1318	1318	1322	1322	1322	1322	1322

*Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

As shown in Table 5, the apparent effect of treatment for the overall sample in mathematics is driven almost entirely by girls, who were slightly overrepresented in the treatment group. The OLS regression analysis for the female subgroup of 654 students with complete math achievement data suggests there is a statistically significant positive effect of the self-affirmation of approximately one tenth of a standard deviation, or about four percentile points, on girls' MCA math test scores.

No statistically significant or practically significant effects were found for boys in math and reading or for girls in the reading domain. As the results in Table 5 suggest, the most consistent predictors of math and reading achievement were special education status, prior year test scores, and pre-intervention GPA. In all cases, students assigned to special education were predicted to perform at a lower level on the MCA than non-special education students, and students with higher prior test scores and higher GPAs tended to also score higher 2011 MCA math and reading scores.

After controlling for the additional variables in the models, it is noteworthy that there were no statistically significant differences among student racial/ethnic groups on either MCA outcome. We performed similar subgroup analyses by student race/ethnicity and found no statistically significant impacts of the self-affirmation for either of the hypothesized stereotyped groups--African American and Hispanic students. (see Table 6).

In Table 7, we employ gender-by-treatment and school-by-treatment interaction terms to test the robustness of our models of the overall impacts from Table 4 and to further explore potential treatment heterogeneity. These analyses include the full complement of variables used in Model 5 from Table 4, and first add the gender-by-treatment interaction in the results displayed in the left-hand column for both reading and math, and the school-by-treatment interaction in the right-hand columns. The results in Table 7 suggest no evidence of a main effect of assignment to the self-affirmation exercise after including the female-by-treatment interaction term. Instead, a statistically significant gender-by-treatment interaction is displayed for the MCA math outcome. In other words,

there is no overall effect of the self-affirmation for the entire sample of boys and girls. Rather, the positive impacts depend on your gender—boys do not appear to benefit, but girls do.

Also of interest is the statistically significant negative coefficient for female students for the math MCA outcome. This suggests that when comparing the typical boy and girl with the same GPAs and prior test scores who are, otherwise, demographically similar, we would expect that the girl would not perform as well on the 2011 MCA as the boy. This phenomenon is not necessarily unusual. For instance, at the undergraduate level, women typically earn higher grades than predicted by their performance on ability or achievement tests, and men earn lower than expected grades (e.g., Hunter, Schmidt, & Rauschenberger, 1984; Jenson, 1980; Linn, 1973, Stricker Rock, & Burton, 1993). As Duckworth and Seligman (2006) note, these phenomena have been termed *underprediction* and *overprediction*, respectively, and have been investigated extensively at the undergraduate level but less thoroughly among younger students. In our case, our models show that females' MCA math scores are lower than what we would expect given their prior GPAs. In other words, GPA underpredicts the math MCA score for girls.

It is also of interest that this underprediction occurs for girls only in math, which is a negatively stereotyped domain for females. This underprediction may signal that girls are experiencing stereotype threat because males are often expected to outperform females on standardized math tests and girls' anxiety about underperformance on the test could be interfering with their ability to obtain a higher score that is more reflective of their true potential (Steele, 1997). This finding is also interesting in light of the fact that one of the participating schools had separate schools-within-a-school for boys and girls. Although the results in Table 7 suggest that the treatment impacts were statistically equivalent across the three participating schools, the positive impact of the self-affirmation on math test scores was strongest for the girls attending the school divided by gender.

TABLE 5. RESULTS WITHIN GENDER

	READING		MATH	
	MALE	**FEMALE**	**MALE**	**FEMALE**
TREATMENT CONDITION	-0.022 (0.042)	0.019 (0.039)	-0.005 (0.039)	0.104** (0.035)
PRIOR YEAR TEST SCORE (STANDARDIZED)	0.705*** (0.029)	0.656*** (0.030)	0.704*** (0.027)	0.583*** (0.027)
PRE-INTERVENTION GPA (STANDARDIZED)	0.178*** (0.027)	0.194*** (0.026)	0.175*** (0.026)	0.283*** (0.025)
GRADE 8	0.026 (0.043)	0.039 (0.040)	0.026 (0.040)	0.040 (0.035)
SCHOOL 1	-0.015 (0.049)	-0.124** (0.045)	-0.191*** (0.046)	-0.140*** (0.040)
SCHOOL 2	0.002 (0.082)	-0.050 (0.065)	0.083 (0.075)	-0.009 (0.058)
FREE/REDUCED LUNCH PROGRAM	-0.022 (0.057)	-0.066 (0.057)	-0.013 (0.052)	-0.099* (0.050)
SPECIAL EDUCATION	-0.230*** (0.059)	-0.286*** (0.077)	-0.139* (0.054)	-0.289*** (0.067)
LIMITED ENGLISH PROFICIENCY	-0.014 (0.067)	-0.088 (0.061)	-0.087 (0.061)	-0.014 (0.053)
AMERICAN INDIAN	-0.144 (0.155)	0.009 (0.153)	-0.091 (0.141)	-0.167 (0.136)
ASIAN OR PACIFIC ISLANDER	-0.014 (0.075)	-0.117 (0.072)	-0.066 (0.069)	-0.089 (0.063)
BLACK (NON-HISPANIC)	-0.065 (0.065)	-0.087 (0.061)	-0.100 (0.060)	-0.001 (0.054)
HISPANIC	0.047 (0.084)	-0.084 (0.075)	-0.063 (0.078)	0.014 (0.067)
CONSTANT	0.031 (0.060)	0.225*** (0.063)	0.209*** (0.054)	0.131* (0.053)
OBSERVATIONS	667	651	668	654

*Notes: Standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001*

RESULTS (CONT'D)

TABLE 6. SUMMARY OF ESTIMATED IMPACT OF ASSIGNMENT TO SELF-AFFIRMATION
EXERCISE BY STUDENT SUBGROUP

RACE/ETHNICITY	READING	MATH
BLACK [$N_R = 401$, $N_M = 403$]	0.023 (0.049)	0.073 (0.054)
ASIAN [$N_R = 341$, $N_M = 342$]	-0.066 (0.050)	-0.075 (0.046)
HISPANIC [$N_R = 187$, $N_M = 186$]	0.055 (0.084)	0.053 (0.074)
GENDER	READING	MATH
MALE [$N_R = 667$, $N_M = 668$]	-0.022 (0.042)	-0.005 (0.039)
FEMALE [$N_R = 651$, $N_M = 654$]	0.019 (0.039)	0.104** (0.035)

Notes: N_r: number of observations in reading; N_m: number of observatiosn in mathematics. Outcome is reported in standard deviations of 2012 test score. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Controls for prior test score, GPA, grade level, FRL status, LEP status, special education status, school attended, gender (for ethnicity models) and ethnicity (for gender models) not shown.

TABLE 7. TREATMENT INTERACTIONS WITH STUDENT GENDER AND SCHOOL

	READING		MATH	
TREATMENT CONDITION	-0.027 (0.040)	-0.012 (0.048)	-0.006 (0.037)	-0.023 (0.044)
PRIOR YEAR TEST SCORE (STANDARDIZED)	0.684*** (0.021)	0.684*** (0.021)	0.654*** (0.019)	0.655*** (0.019)
PRE-INTERVENTION GPA (STANDARDIZED)	0.184*** (0.019)	0.184*** (0.019)	0.222*** (0.018)	0.221*** (0.018)
FEMALE	0.037 (0.041)	0.036 (0.041)	-0.089* (0.038)	-0.087* (0.038)
FEMALE x TREATMENT	0.046 (0.057)	0.047 (0.058)	0.113* (0.052)	0.109* (0.052)
SCHOOL 1	-0.065 (0.033)	-0.046 (0.045)	-0.171*** (0.030)	-0.181*** (0.041)
SCHOOL 1 x TREATMENT		-0.039 (0.061)		0.022 (0.056)
SCHOOL 2	-0.011 (0.051)	-0.004 (0.072)	0.036 (0.046)	-0.019 (0.066)
SCHOOL 1 x TREATMENT		-0.015 (0.099)		0.105 (0.090)
SCHOOL 2	-0.260*** (0.046)	-0.260*** (0.046)	-0.191*** (0.041)	-0.191*** (0.041)
SCHOOL 2 x TREATMENT		-0.015 (0.099)		0.105 (0.090)
SPECIAL EDUCATION	-0.260*** (0.046)	-0.260*** (0.046)	-0.191*** (0.041)	-0.191*** (0.041)
CONSTANT	0.101* (0.046)	0.093 (0.048)	0.220*** (0.041)	0.229*** (0.043)
OBSERVATIONS	1318	1318	1322	1322

*Notes: Standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001*

DISCUSSION

DISCUSSION

This randomized trial of a self-affirmation intervention distributed across three SPPS middle schools provides some interesting and convincing evidence of the impacts of delivering an intervention intended to affirm girls' and minority students' beliefs in themselves. Our experiment did not replicate the positive impacts for African American students found by Cohen et al. (2006), but we did observe a positive impact on girls' MCA math test scores.

Though women now earn 58 percent of all bachelor's degrees, 59 percent of master's degrees, and 45 percent of all doctoral degrees, females continue to be underrepresented in math and science graduate degree programs and math and science career pathways (National Science Foundation, 2006). In general, girls do as well as, or better than, boys on homework assignments and course grades in math and science classes (College Board, 2006; Shettle et al., 2007), but boys tend to outscore girls when tested on the same content in high-pressure situations, such as standardized tests with time limits. What explains these differences and the continued underrepresentation among women in math and science graduate programs and careers?

Areas where researchers have found consistent gender differences concern children's and adolescents' beliefs about their abilities in math and science, their interest in math and science, and their perceptions of the importance of math and science for their futures. In general, researchers have found that girls and women have less confidence in their math abilities than males do and that from early adolescence, girls show less interest in math or science careers (Andre et al., 1999; Herbert & Stipek, 2005; Jacobs et al., 2002; Lanza et al., 2002; Simpkins & Davis-Kean, 2005; Wigfield et al., 1991). This gender difference is somewhat puzzling, given that males and females generally enroll in similar courses and display similar abilities —at least as measured by course grades. In other words, girls, particularly as they move out of

elementary school and into middle and high school and beyond, often underestimate their abilities in mathematics and science. However, it is important to note that not all girls have less confidence and interest in mathematics and science, and that girls, as well as boys, who have a strong self-concept regarding their abilities in math or science are more likely to choose and perform well in elective math and science courses and to select math- and science-related college majors and careers (Simpkins and Davis-Kean (2005), Updegraff and Eccles (1996)). As Halpern et al. (2007) suggest, this is important because it implies that positively impacting girls' beliefs about their abilities could alter their choices and performance. Theory and empirical research suggest that children's beliefs about their abilities are central to determining their interest and performance in different subjects, the classes they choose to take, and, ultimately, the career choices they make (Pajares, 2006).

This study also provides a potential model for how future policy adoptions could benefit from simultaneously adopted research efforts. From the outset, there was some uncertainty with regard to the potential of the self-affirmation writing exercise. By developing a pilot program wherein we implemented the intervention alongside our efforts to evaluate it, we were able to track the impact of the strategy from the beginning, and convincingly compare the outcomes of the sample of students experiencing the intervention to their peers. While the lack of positive results for minority students was disappointing, it was an important and useful exercise that helped us learn how a new strategy affected key outcomes within the district. Insofar as the study found positive effects for some students, it may also provide some convincing evidence to support district-wide or school-to-school decisions for broader or more regular adoptions of the affirmation exercises. Future studies may benefit from a larger sample, or observing the effects of this pilot over a longer term by continuing to track the progress of the participating students.

REFERENCES

REFERENCES

Andre, T., Whigham, M., Hendrickson, A., and Chambers, S. (1999). Competency beliefs, positive affect, and gender stereotypes of elementary students and their parents about science versus other school subjects. *Journal of Research in Science Teaching, 36,* 719-747.

Aronson, J. (2002). Stereotype threat: Contending and coping with unnerving expectation. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 279–301). New York: Academic Press.

Aronson, J. & Inzlicht, M. (2004). The ups and downs of attributional ambiguity: Stereotype vulnerability and the academic self-knowledge of African-American students. *Psychological Science, 15,* 829–836.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. L., (2006). A self-affirmation intervention to reduce the racial achievement gap. *Science, 313, 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,* 400–403.

Cole, B., Matheson, K., & Anisman, H. (2007). The moderating role of ethnic identity and social support on relations between well-being and academic performance. *Journal of Applied Social Psychology, 37,* 592–615.

College Board. (2006, August 29). *College Board announces scores for new SAT with writing section.* Retrieved November 15, 2006, from http://www.collegeboard.com/press/releases/150054.html.

Creswell, J. D., Welch, W. T., Taylor, S. E., Sherman, D. K., Gruenewald, T. L. & Mann, T. (2005). Affirmation of personal values buffers neuroendocrine and psychological stress responses. *Psychological Science, 16, 846–851.*

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98(1)*, 198-208.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24,* 645–662.

Good, C., Rattan, A., & Dweck, C.S. (2007). *Theories of intelligence influence females' sense of belonging, intent to continue, and achievement in math.* New York: Columbia University.

Halpern, D., Aronson, J., Reimer, N., Simpkins, S., Star, J., and Wentzel, K. (2007). *Encouraging girls in math and science* (NCER 2007-2003). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ncer.ed.gov.

Halpern, D.F., Benbow, C.P., Geary, D.C., Gur, R.C., Hyde, J.S., and Gernsbacher, M.A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*(1), 1–51.

Herbert, J., and Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Applied Developmental Psychology, 26,* 276–295.

Hunter, J. E., Schmidt, F. L., & Rauschenberger, J. (1984). Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 41–99). New York: Plenum Press.

Jacobs, J.E., Lanza, S., Osgood, D.W., Eccles, J.S., and Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73,* 509–527.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 47,* 193–198.

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29,* 371–381.

Kellow, T. J., & Jones, B. D. (2005). Stereotype threat in African-American high school students: An initial investigation. *Current Issues in Education,* 8(20). Retrieved from http://cie.ed.asu.edu/volume8/number20/

Linn, R. L. (1973). Fair test use in selection. *Review of Educational Research, 43,* 139–161.

Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance. The role of psychological disengagement. *Personality and Social Psychology Bulletin, 24*, 34–50.

Martens, A., Johns, M., Greenberg, J., & Schimel (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42,* 236–243.

McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology, 27,* 486–493.

National Science Foundation. (2006d). *Women, minorities, and persons with disabilities in science and engineering.* Retrieved June 25, 2006, from http://www.nsf.gov/statistics/wmpd/.

Pajares, F. (2006). Self-efficacy during childhood and adolescence: Implications for teachers and parents. In F. Pajares and T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 339–367). Greenwich, CT: Information Age Publishing.

Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology, 40*, 152–168

Roberson, L., Deitch, E. A., Brief, A. P., & Block, C. J. (2003). Stereotype threat and feedback seeking in the workplace. *Journal of Vocational Behavior, 62,* 176–188.

Schimel, J., Arndt, J., Banko, K. M., & Cook, A. (2004). Not all self-affirmations were created equal: The cognitive and social benefits of affirming the intrinsic (vs. extrinsic) self. *Social Cognition, 22,* 75–99.

Shettle, C., Roey, S., Mordica, J., Perkins, R., Nord, C., Teodorovic, J., Brown, J., Lyons, M., Averett, C., and Kastberg, D. (2007). *The nation's report card: America's high school graduates* (NCES 2007-467). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Shih, M., Bonam, C., Sanchez, D., & Peck, C. (2007). The social construction of race: Biracial identity and vulnerability to stereotypes. *Cultural Diversity and Ethnic Minority Psychology, 13, 125–133.*

REFERENCES

Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review, 16(3),* 177–206.

Simpkins, S.D., and Davis-Kean, P.E. (2005). The intersection between self-concept and values: Links between beliefs and choices in high school. *New Directions for Child and Adolescent Development, 110,* 31–47.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613–629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69, 797–811.*

Steele, J., James, J. B., & Barnett, R. (2002). Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. *Psychology of Women Quarterly, 26,* 46–50.

Stricker, L. J., Rock, D. A., & Burton, N. W. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. *Journal of Educational Psychology, 85,* 710–718.

Updegraff, K.A., and Eccles, J.S. (1996). Course enrollment as self-regulatory behavior: Who takes optional high school math courses? *Learning and Individual Differences, 8,* 239–259.

Vanneman, A., Hamilton, L., Baldwin Anderson, J., & Rahman, T. (2009). *Achievement gaps: How black and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress* (NCES 2009-455). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/studies/2009455.pdf

Wainer, H., and Steinberg, L.S. (1992). Sex differences in performance on mathematics section of the Scholastic Aptitude Test: A bidirectional validity study. *Harvard Educational Review, 62*(3), 323–336.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*, 82–96.

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, 1132-1139.*

What Works Clearinghouse. (2010). *WWC quick review of the article "Recursive processes in self-affirmation: intervening to close the minority achievement gap."* Retrieved from http://ies.ed.gov/pubsearch/pubsinfo. asp?pubid=WWCQRRP0210

Wigfield, A., Eccles, J.S., Mac Iver, D., Reuman, D.A., and Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27,* 552–565.

Woolf, K., McManus, I. C., Gill, D., & Dacre, J. (2009). The effect of a brief social intervention on the examination results of UK medical students: A cluster randomised controlled trial. *BMC Medical Journal, 9*, 1.