Abstract Title Page
Not included in page count.

Title: An Effort to Close Achievement Gaps at Scale through Self-Affirmation

Authors and Affiliations: Geoffrey D. Borman and Jeffrey Grigg, University of Wisconsin—Madison

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

The key policy issue we propose to address is closing the academic performance gaps between African American and Latino students and their White counterparts, and between girls and boys in mathematics. 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*. Steele and Aronson (1995), who coined the term, have referred to 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. As this threat persists over time, Steele and Aronson argue that it may have the further effect of pressuring these students to protectively disidentify with achievement in school and related intellectual domains.

Steele and Aronson (1995) tested their reasoning in a series of laboratory experiments. Their 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 women 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 Yeager & Walton, 2011 for a recent review). Specifically, individuals can manage threatening situations by shoring up their self-concept—a strategy known as "compensatory self-inflation" (Greenberg & Pyszczynski, 1985) or "self-affirmation" (Steele & Lui, 1983; Lui & Steele, 1986).

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

One recent field-based trial of self-affirmation exercises, in particular, cries out for replication because it is so simple and 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; What Works Clearinghouse, 2010) reported that brief self-affirmation

tasks aimed at affirming students' personal values reduced the Black–White GPA gap by as much as 40%, or 0.50 *SDs*, by improving the African American students' performance. We assessed the achievement impacts of this stereotype threat–reduction intervention relative to a control group condition on the test scores of nearly 1,000 middle-school students in the Madison (WI) Metropolitan School District. With recent National Assessment of Educational Progress results identifying Wisconsin's achievement gaps as among the largest in the nation (Vanneman, Hamilton, Baldwin-Anderson, & Rahman, 2009), the proposed work may have important implications for education policy and practice in the state. Moreover, these remarkably cost-effective intervention strategies—each involving simple 15-minute writing exercises—could easily be replicated and implemented across the United States, with the potential to help close the persistent achievement gaps within other school systems across the country.

Setting:
Description of the research location.

The research was conducted in the Madison, WI school district during the 2011-2012 academic year. All eleven of the district's middle schools participated in the study. To date, this is the first district-wide, "scale-up" evaluation of the impacts of self-affirmation writing exercises.

Population / Participants / Subjects:
Description of the participants in the study: who, how many, key features, or characteristics.

Approximately 1,700 seventh grade students are enrolled in the district; all were invited to participate. Parent consent and student assent was ultimately obtained for 1,049 students (61% of the total district enrollment), 943 of whom were enrolled in the study prior to the first administration of the writing exercise at their school. Test scores from both sixth and seventh grade were available for 881 students in reading and for 886 students in mathematics, representing 52% of the district's total seventh grade enrollment. The analytical sample is 50% female, 39% eligible for free/reduced lunch, 13% eligible for special education services, 10% Asian, 17% Black, 17% Hispanic, and 56% White.[1] This is one of very few studies to simultaneously estimate the impact of self-affirmation writing for four ethnic groups.

Half of the consented students were randomly assigned to receive the self-affirmation writing exercise, and half of the students were assigned to a similar exercise that asked students to write about things that might be important to other people. The characteristics of the intervention and comparison groups are reported in Table 1. The two groups are balanced on the whole, especially with respect to demographic characteristics, but the differences in the prior year test score are large enough to justify controlling for students' prior year test score in the analysis.

Intervention / Program / Practice:
Description of the intervention, program, or practice, including details of administration and duration.

The intervention was a self-affirmation writing exercise developed by Geoffrey Cohen and his colleagues (Cohen et al., 2006). We produced personalized copies with the student's name on a

[1] The average demographic characteristics of middle school students in this district are: 48% female, 52% eligible for free/reduced lunch, 17% eligible for special education, 10% Asian, 23% Black, 17% Hispanic, and 44% White.

cover sheet in order to maintain the fidelity of the random assignment. The assignments were distributed by classroom teachers and completed during the school day, either in homeroom or in Language Arts; each administration took about 15-20 minutes. Students completed the assignments quietly and independently, and teachers were not informed of the condition to which their students were assigned.

Students completed up to four writing exercises during the 2011-2012 academic year. The timing of the exercise administration, which was intended to precede major assessments, is shown in Figure 1. On the first page of the exercise, students were presented with a list of things that could be important to them or to other people (e.g., Friends and Family, Music, Religion, etc.). Students in the intervention condition were asked to select two or three items that were *most* important to them. Students in the comparison condition were asked to select two or three items that were *least* important to them. The intervention group was then asked to write a brief essay explaining how those things were important to them, and the comparison group was asked to write a brief essay about how the things that are not important to them might be important to someone else. The second writing exercise introduced school performance to the list of potentially important things. The third writing exercise reviewed the list of potentially important things, and the fourth writing exercise identified one of the items that students selected early in the year and asked them to reflect on it.

Research Design:
Description of the research design.

Within each of the 11 middle schools, half of the consented seventh grade students were randomly assigned to the intervention group. Consequently, students in both the intervention and comparison groups shared classrooms. Non-consented students were given an expository writing assignment to complete. All exercises

The current analysis focuses on the students who were enrolled in the study and randomly assigned to either the intervention or comparison group prior to the first administration of the writing exercise at their school. The standardized test we use as an outcome was conducted between the third and fourth writing exercises.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

The outcome reported here is student performance on the state accountability test in reading and mathematics, which was administered in November 2011, after the second writing exercise was assigned (see Figure 1). Student test scores—including scores in reading and mathematics in the previous year—and demographic information were provided by the school district.

The current analysis is the estimated impact of assignment to the intervention, or the intention-to-treat (ITT) analysis. We estimate the impact of assignment using the following multilevel model of students within the eleven schools:

$$Y_{ij} = \alpha + \beta T_{ij} + \sum \gamma X_{ij} + \sum \delta\left(T_{ij} * X_{ij}\right) + \sum \theta X'_{ij} + u_j + \epsilon_{ij}$$

The model includes assignment to the self-affirmation exercise (β), student gender and ethnicity indicators (γ); a set of interactions that estimate differential treatment impacts by ethnicity and gender (δ); controls for prior year performance, disability status, and free/reduced lunch eligibility (θ); a school-level random intercept (u_j); and a student error term (ϵ_{ij}). Because the intervention is hypothesized to impact students who are subject to stereotype threat, we expect the treatment impacts will be revealed in the interaction terms (δ) rather than in the overall average treatment effect estimate (β).

Findings / Results:
Description of the main findings with specific details.

The results of the analysis are reported in Table 2. As expected, we did not find an overall average effect of assignment to the self-affirmation condition. We did find, however, that gender and ethnic group moderated the impact of self-affirmation. Although we did not find the positive impact for Black students that we expected, we did find that the 8.7 point (0.15 *SD*) advantage that girls had over boys in the comparison group was not apparent in the self-affirmation group. Moreover, Hispanic students in the comparison group scored on average 10.9 and 9.0 points (0.18 *SD*) lower on the reading and mathematics test, and this gap was not apparent in the group that was assigned to complete the self-affirmation exercises.[2] The variation in achievement by gender and ethnicity also appears to be slighter in the self-affirmation condition, as is shown in Figures 2 and 3, which report estimates from a somewhat different model that reports estimates for the intersection of gender and ethnicity.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

Consistent with the conference theme, our large-scale replication produced results that contradicted the earlier findings of Cohen and colleagues (Cohen et al., 2006; Cohen et al., 2009). We find no evidence that the self-affirmations closed the Black-White achievement gap. Our initial results focused on high-stakes achievement test scores as the outcome, while Cohen and his colleagues had previously focused on students' in-school grades. However, Steele and Aronson (1995) contended that evaluative, high-stakes standardized tests, in particular, prime stereotype threat and interfere with the intellectual functioning of students. Our contradictory results have fueled development of a student survey, which we are fielding during the current academic year, to help us develop better understandings of the potential moderators and mediators of the observed effects. We are also engaged in additional analyses of grades, attendance, and behavioral outcomes. These results will allow us to examine whether the observed impacts may be sensitive to the particular outcome domain under consideration. Finally, our presentation will provide a discussion of other theoretical, contextual, and practical reasons for why the prior impacts did not replicate in this case. Our first-year results have been instructive to us in developing our ongoing intervention plans and should offer other researchers interested in the topic with key issues to consider for future replications.

[2] The finding for Hispanic students is robust to alternative specifications of the model, including omitting the sixth grade test score and using the larger sample. The finding for boys is not statistically significant in the alternative specifications due to the larger standard error.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

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.

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.

Greenburg, J., & Pyszczynski, T. (1985). Compensatory self-inflation: A response to the threat to self-regard of public failure. *Journal of Personality and Social Psychology*, 49(1), 273-280.

Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology*, *99,* 545–560.

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/

Liu, T. J., Steele, C. M. (1986). Attributional analysis as self-affirmation. *Journal of Personality and Social Psychology*, 51(3), 531-540.

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.

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, C. M., & Aronson, J. (2004). Stereotype threat does not live by Steele and Aronson (1995) alone. *American Psychologist,* 59(1), 47-48.

Steele, C. M., Liu, T. J. (1983). Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology*, 45(1), 5-19.

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

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

Yeager, D.S., & Walton, G.M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research, 81,* 267-301.

Appendix B. Tables and Figures

Not included in page count.

Table 1: Balance between the Self-Affirmation and Comparison Groups

Variable	Treatment Mean	Comparison Mean	p-value
Female	0.48	0.52	0.313
Free/Reduced Lunch	0.38	0.40	0.546
Disability Status	0.13	0.12	0.800
Asian	0.11	0.09	0.183
Black	0.18	0.16	0.502
Hispanic	0.16	0.18	0.307
White	0.55	0.57	0.580
2011 Reading	513.5	507.2	0.102
2011 Math	528.6	521.2	0.055

Notes: n = 886 (881 for reading). Standard deviations are 57.1 in Reading and 57.7 in Math.

Table 2: Results of Multilevel Models Estimating the Impact of Self-Affirmation Writing on 7th Grade Standardized Test Scores

	Reading	Mathematics
Self-Affirmation Exercise	1.440	-2.357
	(1.471)	(2.149)
Prior Year Test	0.788***	0.681***
	(0.028)	(0.026)
Disability Status	-9.342*	-9.016**
	(4.287)	(3.449)
Free/Reduced Price Lunch	-12.235***	-9.103***
	(2.868)	(1.804)
Female	8.694**	-1.898
	(3.101)	(1.813)
Interaction: Female x Self-Affirmation	-8.641*	1.620
	(3.532)	(2.945)
Asian	8.877	5.032
	(5.947)	(3.641)
Interaction: Asian x Self-Affirmation	-3.191	0.737
	(5.091)	(3.479)
African-American	-4.716	-5.212
	(3.149)	(2.930)
Interaction: African-American x Self-Affirmation	0.549	1.811
	(5.404)	(3.090)
Hispanic	-10.887**	-8.957**
	(3.759)	(3.198)
Interaction: Hispanic x Self-Affirmation	9.367*	11.663***
	(4.601)	(3.297)
Constant	127.927***	189.077***
	(15.616)	(14.504)
Variance Estimates		
Schools	1.233***	0.467
	(0.261)	(0.399)
Students	3.388***	3.137***
	(0.082)	(0.047)
Observations: Students	881	886
Observations: Schools	11	11

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

Figure 1: Study Design, Including the Sequence of Interventions and the Timing of the Outcome (State Accountability Assessment)

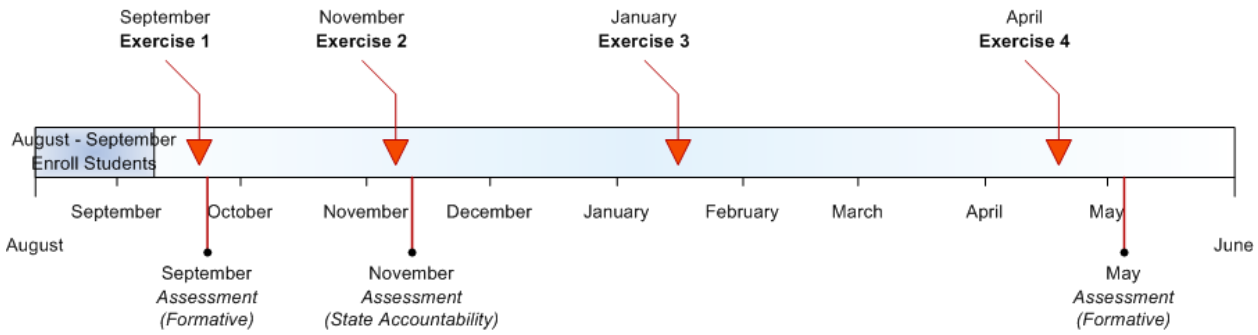

Figure 2: Self-Affirmation Impact Estimates in Reading by Gender and Ethnic Group

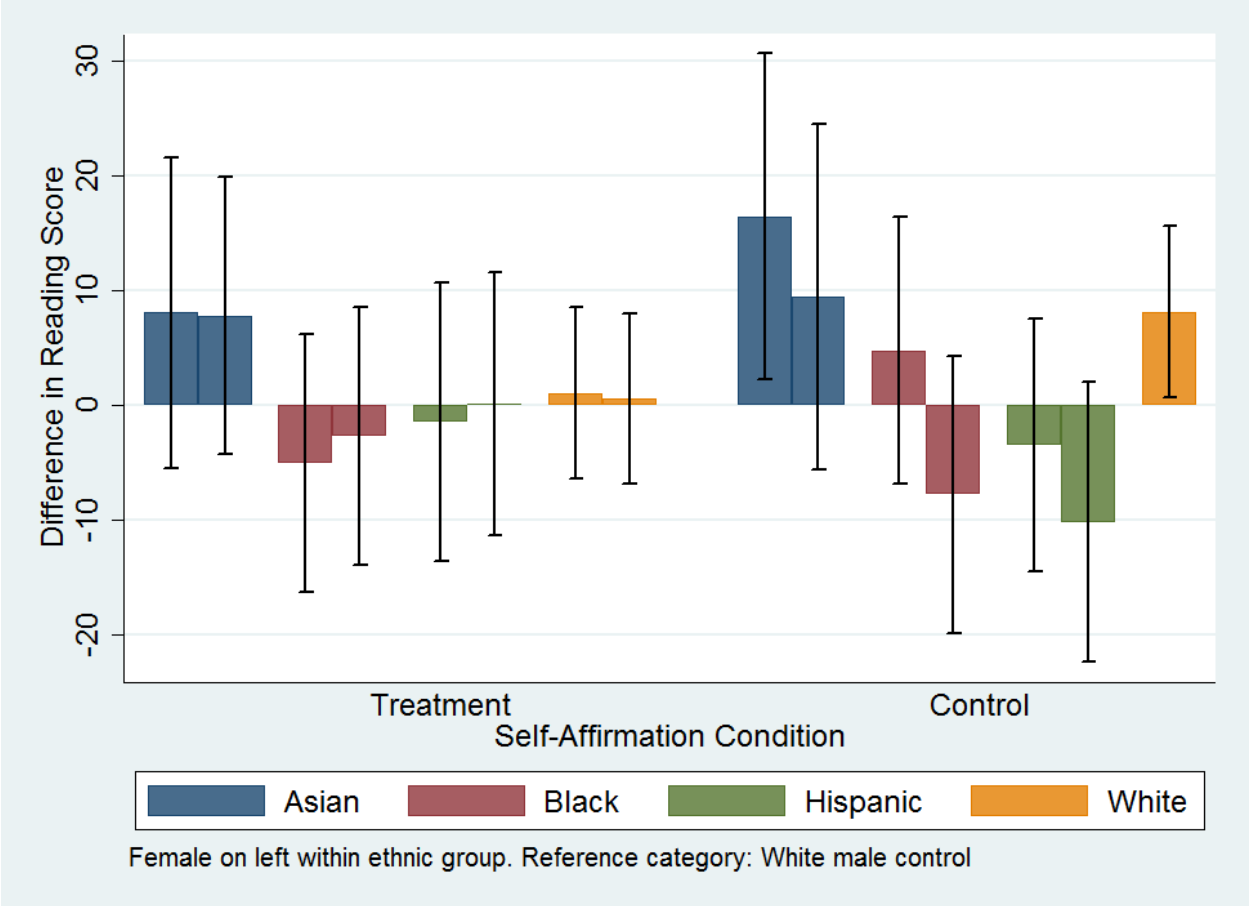

Figure 3: Self-Affirmation Impact Estimates in Mathematics by Gender and Ethnic Group