Abstract Title Page

Not included in page count.

Title:

New Evidence on Self-affirmation Effects and Theorized Sources of Heterogeneity from Two Cohorts in a Large-scale Replication

Authors and Affiliations:

Paul Hanselman, University of California, Irvine
Christopher S. Rozek, University of Chicago
Jeffrey Grigg, Johns Hopkins University
Jaymes Pyne, University of Wisconsin-Madison
Geoffrey Borman, University of Wisconsin-Madison

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

One approach to reducing persistent racial/ethnic achievement gaps is to tackle their social-psychological dimensions, including the negative consequences of stereotype threat and other identity threats in school (Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002). Initial research suggested that a particularly promising approach is brief self-affirmation writing exercises for 7th grade students; studies in individual schools reported this strategy reduced racial achievement gaps in grade point average by as much as 40% (Cohen et al. 2006; Cohen et al. 2009; Sherman et al. 2013). However subsequent evidence has been mixed, with one large scale replication finding almost no benefits (Dee, 2015) and another finding small, if cost-effective, impacts (Borman, Grigg, & Hanselman, 2015).

It is important to distinguish between two potential sources of variability in self-affirmation treatment effects. One is explainable heterogeneity based on the delivery, population served, and context of these interventions. Because social-psychological interventions (including self-affirmation exercises) are hypothesized to target specific individual-context processes, their effects may be quite different in different settings (see Yeager & Walton, 2011). Some previous research supports this claim for self-affirmation (Hanselman et al., 2014). However, another possible explanation is that heterogeneity is fundamentally not explainable, such as would be expected from random differences due to sampling variability.

The practical implications of these two explanations of effect variability are quite different. If the first is true, then the intervention should be developed to target settings where it is most useful. However, if we cannot explain variable effects, then the practical value of the intervention is less clear, and more work is needed on the theoretical pathways and moderators of impacts.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

The key problem with differentiating these two explanations is that it is difficult to isolate relevant differences between different research efforts. Many of the theoretical moderators of treatment impacts— including features of implementation, student population, and school context— are unmeasured and confounded across research projects. The contribution of this paper is to consider heterogeneity in a setting that controls many of these factors by assessing differences in a within-study replication across two implementations of the intervention. Following up on a successful large scale replication of self-affirmation (Borman et al., 2015), we apply the same procedures to the subsequent cohort of 7th grade students in the same school district. The key questions we address are:

1. What was the effect of the self-affirmation intervention in a second population of students (cohort 2)?
2. Were estimated effects in cohort 2 substantively and significantly different from the impacts in cohort 1?

3. Do hypothesized moderators of self-affirmation effects related to the delivery, population, and social context of the intervention explain differential effects?

Setting:

Description of the research location.

The research was conducted in the regular 11 middle schools in Madison, WI in 2011-2014.

Population / Participants / Subjects:

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

The study includes 2109 individuals (939 in cohort 1, 1170 in cohort 2) who were 7th grade students in 2011-2012 or 2012-13. The sample was 50% female, 36% academically stereotyped minority (African American or Hispanic), and 45% eligible for free or reduced price lunch, with an average Grade Point Average of 3.3 (out of 4) in the year prior to the study.

Intervention / Program / Practice:

Description of the intervention, program, or practice, including details of administration and duration.

The self-affirmation intervention procedure followed Cohen et al. (2006). Seventh grade students completed a short (15-20 minute) writing prompt as part of their language arts or homeroom class activities three or four times during the school year. All exercises were identical on the cover sheet, which included each student's name. The exercises in the subsequent pages varied by randomly assigned condition (non-consented students completed the procedural/neutral control prompts). The treatment condition, which took slightly different forms throughout the year, prompted treated students to reflect on values (such as friends, family, music, or sports) that were important to them.

There were two randomly assigned control conditions: one focused on values, in which students are asked to select least important values from the same list presented to treatment students and explain why they may be important to someone else, and a second devoted to various procedural writing prompts, such as explaining how to open a locker (we also refer to these prompts as "neutral," as they do not explicitly concern values). The latter control branch was introduced after the first administration of the first cohort, so all control students in the first cohort received the "Least Important Values" prompt for the first exercise. Because we found no evidence of differences between control conditions in either cohort or evidence that these differences explain differential impacts, we combined both control groups in our main analyses.

Teachers administered the activities but were not aware of experimental condition of any individual student or the ultimate research hypotheses. (The study was described as a general investigation of writing activities.)

Research Design:

Description of the research design.

The research is an experimental design. Students were individually assigned to complete self-affirmation or comparison activities. Randomization was blocked so that treatment-comparison proportions were equal within each school.

Tests for balance on observable pre-randomization characteristics demonstrated baseline equivalence between the treatment and comparison group. In addition, attrition was modest (10%) and similar across experimental conditions.

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

Student outcome data were collected from administrative records through the end of 8th grade (the year following the intervention). Data include demographic characteristics, prior achievement variables, and academic outcomes. The primary outcome of interest was overall grade point average (GPA) in grade 8, reported on a 4 point scale. GPA is the main outcome in the self-affirmation literature because it summarizes both achievement and engagement in school, but of which may be benefited by the intervention. We supplement these analyses with tests of effects on standardized achievement outcomes in mathematics and language arts.

Treatment impact estimates were based on the following general multilevel model of treatment effects:

$$Y_{ijt} = \beta_0 + \beta_1(Treatment_i) + \beta \boldsymbol{X_i} + \eta_j + \varepsilon_i \qquad (1)$$

In this model, Y_ijt is the observed outcome in grade t for student i in school j, Treatment_i is the randomly assigned self-affirmation treatment status for student i, X_i is a vector of pre-treatment covariates (grade 6 outcome, gender, limited English proficiency, special education, and free lunch eligibility), η_j is the residual component for school j, and ε_i is the residual for student i. Because the treatment was randomly assigned to each student, β_1 provides an unbiased estimate of the effect of the self-affirmation writing exercises without additional controls, but we included a pretreatment achievement measure and additional covariates in X_i0 to increase the precision of this estimate.

Because self-affirmation is hypothesized to benefit marginalized students, we focus on estimating the treatment impacts within each cohort for the subsample of students identified as African American or Hispanic.

To address research question 3, we conducted a series of post hoc tests of potential explanations of different effects across cohorts, based on three hypothesized types of moderators of self-affirmation effects: implementation, individual characteristics, and social context (see summary in Table 1).

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

Research Question 1:

We found no evidence of self-affirmation treatment impacts in the second cohort of students. The estimated effect size on grade 8 GPA was -0.072 (se = 0.058) and statistically insignificant. This compares to an estimate of 0.152 (se = 0.070) in the first cohort. This result is robust to different models (covariates), contrast (comparison condition), and subsamples (that may be more sensitive to stereotype threat). Figure 1 summarizes estimates and 95% confidence intervals across a variety of specifications.

Research Question 2:
Estimated impacts on grade 8 GPA for cohort 2 were significantly different from the positive estimate for cohort 1 (p = 0.013). This result was substantively robust to all considered model and subsample specifications, although not always statistically significant, due to reduced precision (see Figure 1).

Research Question 3:
We found no evidence that any of the hypothesized mediators explained the differential effects of self-affirmation between the two cohorts. Student responses suggested similar engagement with the activities, changes in the student population were minor and did not moderate effects, and we found no evidence that changes in individual school contexts explained the results. In the interest of space, we present a summary of considered tests in Table 1.

Conclusions:

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

A limitation of this study is that, like any single study, we cannot identify to what extent our results are driven by sampling variability, or if so which cohort is more indicative of likely effects of the intervention. Post hoc power analyses (Gelman & Carlin, 2014) suggest that chance is not a plausible explanation for our results if the true effects of self-affirmation are as large as reported in Cohen et al. (2006). However, if the true effect is smaller, such as the average effect observed across both cohorts in the current study (d=0.07), then even in a large-scale replication like this one has relatively low power. More independent replication work is clearly need to build robust evidence of the effects of these interventions.

We highlight three key implications of this study:

First, our analyses demonstrate the value of tests of moderators to assess theory about where, and ultimately how, specific interventions are successful. The tests conducted here provide important, if indirect, evidence about the hypothesized influences of the implementation, individual, and context characteristics on self-affirmation impacts.

Second, our results point to the need to develop the theory and evidence about how and where self-affirmation works. Because we tested a comprehensive list of proposed moderators of self-affirmation and failed to explain the variation in our findings between cohorts, we conclude that the current cadre of moderators offered by the literature is insufficient.

Third, our results imply practical limitations of self-affirmation as a tool to improve student performance and close achievement gaps. If variability in impacts cannot be predicted with the information available to educators, then the practical value of these interventions is unclear.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Borman, G. D., Grigg, J., & Hanselman, P. (2015). An Effort to Close Achievement Gaps at Scale Through Self-Affirmation. *Educational Evaluation and Policy Analysis*, 0162373715581709. http://doi.org/10.3102/0162373715581709

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. Science, 313(5791), 1307-1310. doi: 10.1126/science.1128317

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap. Science, 324(5925), 400-403. doi: 10.1126/science.1170769

Dee, T. S. (2015). Social Identity and Achievement Gaps: Evidence from an Affirmation Intervention. Journal of Research on Educational Effectiveness, 8(2), 149-168. doi: 10.1080/19345747.2014.906009

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspectives on Psychological Science, 9(6), 641-651. doi: 10.1177/1745691614551642

Hanselman, P., Bruch, S. K., Gamoran, A., & Borman, G. D. (2014). Threat in Context: School Moderation of the Impact of Social Identity Threat on Racial/Ethnic Achievement Gaps. Sociology of Education, 87(2), 106-124. doi: 10.1177/0038040714525970

Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., Cohen, G. L. (2013). Deflecting the Trajectory and Changing the Narrative: How Self-Affirmation Affects Academic Performance and Motivation Under Identity Threat. Journal of Personality and Social Psychology, 104(4), 591-618. doi: 10.1037/a0031495

Steele, C. M., & Aronson, J. (1995). Stereotype Threat and the Intellectual Test-Performance of African-Americans. Journal of Personality and Social Psychology, 69(5), 797-811.

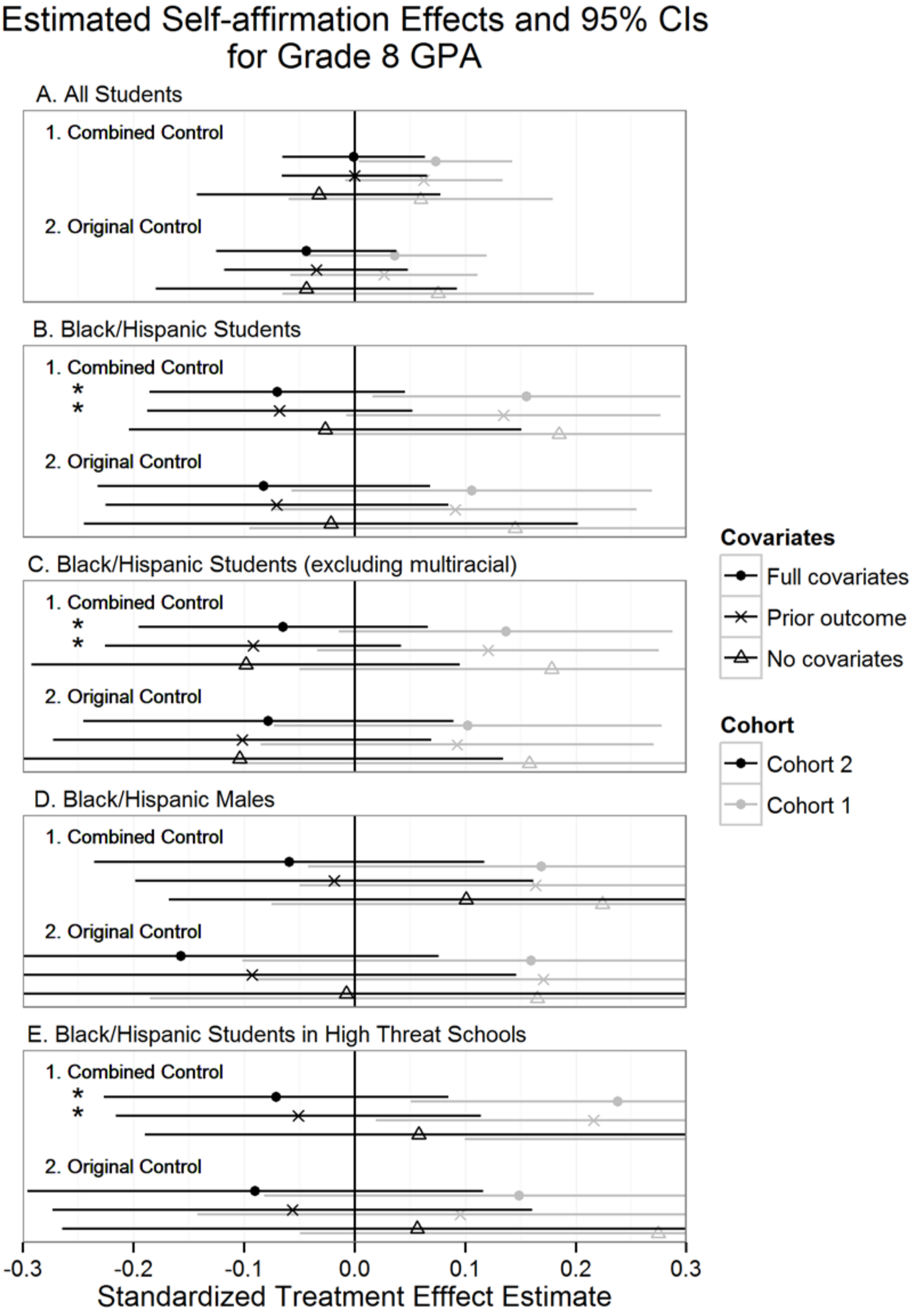
Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. Advances in Experimental Social Psychology, 34, 379-440. doi: 10.1016/s0065-2601(02)80009-0

Yeager, D. S., & Walton, G. M. (2011). Social-Psychological Interventions in Education: They're Not Magic. Review of Educational Research, 81(2), 267-301. doi: 10.3102/0034654311405999

Appendix B. Tables and Figures

Not included in page count.

Figure 1. Estimated Self-affirmation Treatment Effects on Grade 8 GPA by Cohort, Sample, Comparison Group, and Included Covariates



GPA = Overall Grade Point Average; CI = Confidence Interval
Note: Each estimate was calculated from a separate multilevel model (students nested within schools) of intention to treat effect of the self-affirmation writing activities. Full covariates specifications include: grade 6 GPA, gender, special education status, Limited English Proficiency designation, and eligibility for free or reduced price lunch. Prior outcome is grade 6 GPA. In the “Original Control” condition, students wrote about a least important value in each of the first two interventions. The “Combined Control” group includes these students as well as those who were assigned at least one writing prompt that did not explicitly mention values. For readability, the displayed range is restricted to effect sizes of absolute value 0.3 or less. Asterisks indicate that the estimated effects are statistically significantly different between cohorts ($p < 0.05$), based on a pooled model. The primary result, reported in abstract text, is the estimate for Black/Hispanic sample with combined control condition and full covariates (Panel B1 circles). Other results assess whether patterns were different for subpopulations and comparisons where self-affirmation benefits are hypothesized to be stronger and more consistent, as described in the text. Because the cohort difference persists across all specifications (although less precise in smaller subsamples), these tests provide no evidence that hypothesized moderators explain the difference.

Table 1. Summary of Tested Hypotheses

Hypothesized Explanation for Difference in Effects	**Empirical Test**	**Result**
Different effects due to features of the intervention implementation		
Providers	Consistent benefits for teachers implementing in both cohorts	No
	All changes in benefits are due to teachers implementing in both cohorts (due to fatigue)	No
Control group	Consistent benefits when compared to students in the original control condition.	No
Stealth	Teachers report more violations of protocol in second cohort: describing the activity as externally imposed research	No
Awareness of Purported benefits	Teachers report more violations of protocol in second cohort: describing the activity as "good for you"	No
Timing	Intervention more likely to miss key stressful periods in second cohort	No
Engagement with the prompt	Students complete fewer exercises in second cohort	No
	Students write fewer words in second cohort	No
	Impact on self-affirming writing is different in second cohort	No
Different effects due to individual characteristics		
Racial group	Consistent benefits for all Black and Hispanic students	No
	Consistent benefits for non-multiracial Black and Hispanic students	No
Race and gender	Consistent benefits for male minority students	No
Prior achievement and other administrative characteristics	Consistent benefits when populations are re-weighted across cohorts on observable characteristics	No

Unobserved receptivity to self-affirmation	Magnitude of different benefits for unobserved populations are plausible	No
Social context differences		
Broad (district) racial and academic climate	A higher share of racial minorities in the second cohort	No
	Lower racial achievement differences in the second cohort	No
School racial and academic climate	More consistent benefits in “high threat” schools with few minority students and large gaps	No
	Differential benefits explained by one or two schools	No