

Self-Affirmation Effects Are Produced by School Context, Student Engagement With the Intervention, and Time: Lessons From a District-Wide Implementation

Geoffrey D. Borman¹, Jeffrey Grigg² , Christopher S. Rozek³, Paul Hanselman⁴, and Nathaniel A. Dewey²

¹Department of Educational Leadership & Policy Analysis, University of Wisconsin–Madison; ²School of Education, Johns Hopkins University; ³Department of Psychology, University of Chicago; and ⁴Department of Sociology, University of California, Irvine

Psychological Science
2018, Vol. 29(11) 1773–1784
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0956797618784016
www.psychologicalscience.org/PS


Abstract

Self-affirmation shows promise for reducing racial academic-achievement gaps; recently, however, mixed results have raised questions about the circumstances under which the self-affirmation intervention produces lasting benefits at scale. In this follow-up to the first district-wide scale-up of a self-affirmation intervention, we examined whether initial academic benefits in middle school carried over into high school, we tested for differential impacts moderated by school context, and we assessed the causal effects of student engagement with the self-affirming writing prompted by the intervention. Longitudinal results indicate that self-affirmation reduces the growth of the racial achievement gap by 50% across the high school transition ($N = 920$). Additionally, impacts are greatest within school contexts that cued stronger identity threats for racial minority students, and student engagement is causally associated with benefits. Our results imply the potential for powerful, lasting academic impacts from self-affirmation interventions if implemented broadly; however, these effects will depend on both contextual and individual factors.

Keywords

schools, minority groups, intervention

Received 11/9/17; Revision accepted 5/21/18

Over the past two decades, a new class of social psychological interventions, such as mind-set (Blackwell, Trzesniewski, & Dweck, 2007), self-affirmation (Cohen, Garcia, Apfel, & Master, 2006), utility-value (Hulleman & Harackiewicz, 2009; Rozek, Svoboda, Harackiewicz, Hulleman, & Hyde, 2017), and social-belonging (Walton & Cohen, 2011) interventions, has demonstrated promise as a cost-effective strategy to improve educational outcomes. These “wise interventions” attempt to change how students construe failure, stereotypes, belonging, or other challenges they may encounter in school (Walton, 2014). Despite initial successes in small field trials, more evidence is needed to establish the broader efficacy of these approaches, especially across necessarily variable conditions (Borman, 2017).

In the present study, we focused on a self-affirmation intervention consisting of a series of 20-min exercises

in which students wrote expressively about important personal values. This intervention resulted in both a remarkable 40% decrease in the Black–White grade-point-average (GPA) gap at a middle school in the northeastern United States (Cohen et al., 2006) and, subsequently, a decade of efforts to replicate and scale up the intervention (Cohen & Sherman, 2014; Harackiewicz & Priniski, 2018). Self-affirmation interventions are believed to promote the academic performance of racially stigmatized students by short-circuiting a phenomenon known as *stereotype threat*, which occurs when individuals worry about confirming negative stereotypes regarding

Corresponding Author:

Jeffrey Grigg, Johns Hopkins University School of Education, 2800 N. Charles St., Baltimore, MD 21218
E-mail: jgrigg1@jhu.edu

a group to which they belong (Steele & Aronson, 1995). Students affected by stereotype threat perform below expectations because of diminished cognitive resources brought on by a physical stress response, heightened vigilance regarding personal performance, and the need to suppress negative thoughts induced by the stereotype (Schmader, Johns, & Forbes, 2008). Stereotype threat was first identified in African American students who underperformed on achievement tests when made cognizant of negative stereotypes about the academic capabilities of their racial group (Steele & Aronson, 1995). It has since been confirmed by numerous additional studies suggesting that standard measures underestimate the true abilities of Black and Hispanic students by approximately one fifth of a standard deviation (Walton & Spencer, 2009).

Recent research suggests that self-affirmation reduces stress for stigmatized group members through perspective broadening (Critcher & Dunning, 2015). By prompting students to write about primarily nonacademic values or attributes, self-affirmation exercises remind them that the overall self-concept consists of more than just the academic domain. In so doing, the intervention can broaden one's perspective beyond feeling like a negatively stereotyped student in school. When the threat is put into perspective, it decreases in prominence, and its effect on students' cognitive resources also decreases (Critcher & Dunning, 2015). Consequently, the academic performance of stereotype-threatened students increases after a self-affirmation exercise (Cohen & Sherman, 2014).

By fostering a less-stressful academic setting, a self-affirmation exercise prior to an assignment or test allows stereotype-threatened students to perform up to their potential. When the exercises are given at targeted times during the school year, such as the first week of classes or the week prior to a high-stakes test (Cook, Purdie-Vaughns, Garcia, & Cohen, 2012), then proximal improvements, such as a better grade or higher test score, can reinforce a more positive academic identity. When these quick wins accumulate, threatened students may develop an emerging narrative of personal adequacy (Cohen & Sherman, 2014), teachers may begin to see students as more able (Purdie-Vaughns et al., 2009), and eventually, students may even spontaneously recommit to their self-affirmations when faced with future threats (Brady et al., 2016).

In these ways, the effects of a brief intervention can persist and build for months or even years (Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009; Goyer et al., 2017; Tibbetts et al., 2016). In the absence of relief from stereotype threat, students may experience the opposite cycle, in which negative performances during crucial periods reinforce negative attitudes and expectations. The surprisingly large effect

observed from self-affirmation interventions, as well as other types of wise interventions, has been attributed to diverging recursive cycles of positive and negative performance over time (Cohen et al., 2009; Harackiewicz & Priniski, 2018).

This research raises several practical and theoretical questions about self-affirmation as a strategy to improve educational outcomes, especially if it were to be adopted widely. Indeed, some attempts to replicate the effects of a self-affirmation intervention in other settings have been unsuccessful (e.g., Bratter, Rowley, & Chukhray, 2016; Dee, 2015; Hanselman, Rozek, Grigg, & Borman, 2017; Protzko & Aronson, 2016), and thus far, only the original research team has found long-term effects for secondary-school students (Goyer et al., 2017). Heterogeneous impacts may reflect contextual variation in the benefits of self-affirming or the challenges of triggering the key psychological mechanism when implementation varies. Given these possible contingencies of self-affirmation effects, a pressing question concerns their long-term academic impact. Can positive recursive processes drive lasting benefits when implemented at scale, or do initial benefits fade out? Do benefits persist as students transition into new educational environments, and under what kinds of individual and social circumstances? Moreover, to the extent that students respond to the self-affirmation activity differently when implemented broadly, what are the consequences for the long-term effects of the theorized self-affirming writing?

The present study addressed these questions with new data from a randomized controlled trial of a self-affirmation intervention implemented at scale in an entire urban school district in the midwestern United States. Prior research suggested that at the end of the first year of the trial, potentially stereotype-threatened Black and Hispanic students assigned to receive the self-affirmation exercises outperformed the control group in seventh and eighth grade (Borman, Grigg, & Hanselman, 2016; Hanselman et al., 2017). Additional analyses suggest that the impact was larger for students in schools hypothesized to be high-threat contexts, where Black and Hispanic students had a lower numerical presence and academic standing (Hanselman, Bruch, Gamoran, & Borman, 2014).

To begin, we investigated whether the initial impacts persisted across the transition to high school. Evidence of enduring effects in a new academic setting would suggest that the intervention impacted the students in a way that transcended their middle-school environment and that the intervention could lead to long-term changes in students' academic trajectories. Next, we examined whether potentially threatened students' educational contexts—the extent to which their middle-school setting had the hallmarks of a stereotype-threatening

environment—moderated the long-term effect. Finally, we aimed to determine how the level of student engagement with the intervention, assessed by coding the degree to which students demonstrated self-affirming writing during the exercises, influenced the average treatment effect of the intervention. Given that not all students engaged with the self-affirmation exercises to the same extent, this would suggest that the estimated intent-to-treat effect understated the potential benefits for students who engaged in the theorized perspective broadening through self-affirming writing (the impact of the treatment on the treated, or TOT). Put simply, this research question gauged the impact for students who wrote in a self-affirmed way as encouraged by the exercises. Overall, these research questions allowed us to understand more about in which contexts, for whom, and for how long self-affirmation exercises effectively narrow achievement gaps in schools.

In summary, we had three research questions:

1. Do intervention effects on potentially stereotype-threatened seventh-grade students' GPAs persist through the end of ninth grade?
2. Does school context moderate the lasting effects of the intervention? Specifically, are impacts greater for stereotype-threatened students who initially receive the intervention in high-threat middle schools?
3. To what extent does self-affirming writing—as induced by the intervention—improve academic performance?

Method

Data and participants

The data analyzed here were derived from the Madison Writing and Achievement Project, which was designed to test, at scale, the self-affirmation intervention originally fielded by Cohen and colleagues (2006, 2009). The trial targeted all 1,648 seventh-grade students enrolled in the Madison, Wisconsin, school district, and as many students as possible were enrolled in the study. Parental consent and student assent were obtained for 1,048 students (64%). Analysis of this sample suggested it was representative of the population of seventh graders in the Madison Metropolitan School District (Borman et al., 2016). Within the 11 middle schools in the district, each of the seventh-grade students who gave consent was randomly assigned to the self-affirmation (treatment) condition or a control (comparison) condition. Analytically, students were considered members of the middle school they attended at the time of randomization, even if they left the school during or after the intervention. The sample was sufficient to reliably detect

effect sizes (d s) as small as 0.154 among the subgroup of potentially threatened students ($n = 324$, $\alpha = .05$, $1 - \beta = 0.90$; Dong & Maynard, 2013).

Missing data and attrition

We obtained complete district records, but some GPA data were missing for students who arrived to or transferred from the district during the study period. Cases with incomplete data were dropped, yielding a final sample of 920 (88% of students who consented).¹ There were similar rates of attrition for the self-affirmation and comparison conditions for both the full sample (treatment: 13%, control: 12%), $\chi^2(1, N = 1,048) = 0.38$, $p = .54$, and the subsample of potentially threatened Black and Hispanic students (treatment: 19%, control: 15%), $\chi^2(1, N = 390) = 1.47$, $p = .23$. In the subsample of potentially threatened students who attended a middle school with a high-threat context, the treatment and comparison conditions manifested somewhat greater, yet not statistically significant, differential attrition (treatment: 28%, control: 18%), $\chi^2(1, N = 299) = 3.19$, $p = .07$. Analyses of the demographic characteristics and prior achievement of these students suggest that the students in the self-affirmation and comparison conditions remained statistically equivalent after accounting for attrition (see Table S1 in the Supplemental Material available online).

Procedure

The intervention, as closely as possible, followed the procedures used by Cohen and colleagues (2006) in prior experiments. The seventh graders completed up to four writing exercises during the school year. The first two exercises were nearly identical to those used in prior self-affirmation interventions. Each 20-min exercise involved selecting two or three important personal values (e.g., being with friends or family, having a sense of humor, being good at art) from a list and then, in a few sentences, reflecting on why these values were important to the student.² The comparison exercise involved selecting two or three values the student found personally unimportant and reflecting on why those values might be important to someone else.

The first exercise was given close to the start of the school year just prior to fall benchmark standardized testing, and the second exercise was given in the late fall just before state achievement testing. The third and fourth exercises were given prior to periods of standardized testing in the winter and spring. Four of the 11 schools opted out of the third exercise because of scheduling challenges around the winter holiday; students in all of the schools had the opportunity to complete the first, second, and fourth exercises. As suggested

by the original procedures used by Cohen et al. (2006), the third and fourth exercises differed slightly from earlier exercises in order to diminish perceived repetitiveness and maintain student interest.

The research team trained the classroom teachers prior to proctoring the writing exercises to prepare them to deliver the intervention with fidelity. The intervention was described as a writing exercise, and proctoring teachers were informed that the experience should be stress free for the students. The intervention was not to be introduced as an assessment, a research study, or an activity that would be beneficial to students, as specified by self-affirmation theory (Cohen & Sherman, 2014). Additionally, teachers were not told the purpose of the intervention, nor was there any mention of race or stereotype threat during their orientation. Despite these steps, implementation varied, and in year-end surveys completed by a subset of teachers, some reported describing the activities as beneficial or as research (Hanselman et al., 2017). As a result, findings may reflect a conservative picture of the effects of the intervention. Nevertheless, imperfect implementation is an important component of our project: Could self-affirmation exercises remain effective, even when implemented at scale by trained teachers with limited oversight by the researchers?

Teachers administered the writing exercises to students in either a homeroom or language arts class as a free-writing exercise, similar to other activities completed during these classes. Students received personalized and self-contained three- or four-page writing packets in both experimental conditions (self-affirmation and comparison), which were nearly indistinguishable from one another. Because of these precautions, randomly assigned students were blind to their own and others' experimental condition, as were their teachers. However, unlike in some prior studies, researchers were not physically in the rooms during implementation, which, again, tested whether the intervention is robust enough to be adopted widely in schools.

Measures

Stereotype threat. Demographic and academic data were collected from the district's student records. The primary construct of interest for students was their potential vulnerability to stereotype threat. This was operationalized on the basis of the student's parent-reported racial/ethnic group membership. Students reported as Black or Hispanic were designated as potentially vulnerable to stereotype threat. Multiracial students were considered potentially vulnerable if they reported Black or Hispanic as part of their racial/ethnic identity. This approach was consistent with prior research that has shown no impact of self-affirmation for White and Asian students and positive, statistically significant impacts of

similar magnitudes for Black (Cohen et al., 2006) and Hispanic (Sherman et al., 2013) students.

Covariates. We included pretreatment covariates known to predict GPA to increase the precision of impact estimates.³ Demographic characteristics, school fixed effects, and a measure of prior achievement were included to improve the precision of the impact estimates. Binary indicator variables for gender, eligibility for free and reduced-price lunch (a proxy for socioeconomic status), special education status, and limited English proficiency were generated from the demographic data. Sixth-grade cumulative GPA, a continuous variable on a 4-point scale, measured prior achievement. (See Table S1 for the characteristics of the overall sample and of each experimental condition.)

Outcomes. The primary outcome of interest was the students' ninth-grade cumulative GPA ($M = 2.92$, $SD = 1.02$). We also examined the trajectory of students' GPAs across the 3-year postintervention span using a latent growth model. Our outcome variable for the latent growth model was the students' GPAs on report cards from the beginning of seventh grade through the end of ninth grade. Grades were reported in middle school by quarter and in high school by semester, yielding up to 10 GPAs for each student. All GPAs were calculated on a 4-point scale by converting letter grades in classes to their 4-point equivalent and weighting them by credits earned. When we report effect sizes for ninth-grade outcomes (d), we use the standard deviation of the full sample (1.02).

High-threat school context. Our second hypothesis required a measure of each middle school's threat context. We measured context in the year of the intervention to focus on the fundamental moderation processes related to the key immediate effects of the self-affirming exercises. School-threat context was operationalized as a binary variable, referred to as high-threat context, meant to signal that a school's environment was more conducive to the presence of stereotype threat (Hanselman et al., 2014). Following previous research, we included two measures of school characteristics in this operationalization: the proportion of students in the school who were reported as Black or Hispanic and an index of the standardized test performance of Black and Hispanic students relative to their White and Asian peers. These measures were normalized and averaged to create a combined measure of a school's vulnerability to stereotype threat. After finding the mean vulnerability to stereotype threat among the 11 middle schools in the district, the 7 schools greater than the mean were coded as having high-threat context. Students were considered to have attended a high-threat middle school if they did so at the beginning of seventh grade when their data were captured for the experiment.⁴

Coding the writing exercises for self-affirming writing. Our third hypothesis, concerning the effect of students engaging in self-affirming writing during the exercises, required a measure of this particular type of student engagement. We operationalized self-affirming writing as a binary indicator of the combination of two necessary components: (a) The student selected a value from the list, and (b) the student affirmed a value (i.e., wrote about it being important, for example, because the student enjoyed it or was good at it). Trained coders determined whether students selected a value from the list on the basis of whether their written responses mentioned one of the intervention's listed values. Students were then coded as having affirmed a value if the coder determined that the written responses discussed a value in terms of its importance to them (e.g., using "like," "love," "care about," "good at," "best at"). This allowed us to detect whether students in the comparison condition spontaneously self-affirmed in response to the prompt. The coding scheme for these variables was initially generated on the basis of self-affirmation theory as part of a content analysis of a random subsample of the responses and refined during a round of pilot coding.⁵ Using the final coding scheme, two trained coders then rated each exercise; these coders were blind to experimental condition, exercise prompt, and the other coder's ratings. A third coder adjudicated all cases of disagreement. Coder agreement was substantial for the self-affirmation construct, $\kappa = 0.75$, $N = 3,592$, $p_0 = .88$, according to guidelines for interrater reliability (Landis & Koch, 1977).⁶

For each student, all four exercises were independently coded for self-affirming writing. Because the possible recursive benefits of self-affirmation are contingent on timely intervention, and there is currently limited empirical evidence that acts of self-affirmation accumulate (Cook et al., 2012), we considered a student to have engaged in self-affirming writing if he or she were coded as doing so on both the first and second exercises. By this definition, 76% of potentially threatened students in the self-affirmation condition and 2% in the comparison condition consistently demonstrated self-affirming writing. These rates were similar for potentially threatened students within high-threat school contexts.

Results

Hypothesis 1: do intervention effects persist across the transition to high school?

To test whether the impact of the intervention persisted through the end of ninth grade, we estimated a multi-level growth model (Muthén, 1997; Raudenbush & Bryk, 2002) for all students in the sample:

$$\begin{aligned} \text{GPA}_{it} = & b_{00} + b_{01}\text{treatment}_i + b_{02}\text{threatened}_i \\ & + b_{03}\text{treatment}_i \times \text{threatened}_i + b_{10}\text{year}_{it} \\ & + b_{11}\text{treatment}_i \times \text{year}_{it} + b_{12}\text{threatened}_i \times \text{year}_{it} \\ & + b_{13}\text{treatment}_i \times \text{threatened}_i \times \text{year}_{it} + b\mathbf{X}_i + \eta_i + \varepsilon_{it}. \end{aligned}$$

The outcome variable, GPA_{it} , represents the GPA of student i at time t , where time is measured in years prior to the end of ninth grade ($t = 0$ is the end of ninth grade) so that main effects in the model for treatment and potentially threatened students represent differences at the end of the time period. The impact of the intervention on end-of-year ninth-grade GPA for students not subject to stereotype threat is the coefficient b_{01} . The additional impact for potentially threatened students is the coefficient b_{03} . We expected the estimate for b_{01} to be null and the estimate for b_{03} to be positive, indicating that benefits for theoretically affected students persisted through the end of ninth grade. Beyond end-of-ninth-grade predictions, the growth model allowed us to further examine how postintervention student GPAs changed over time. The coefficient b_{10} represents the overall trend in postintervention student GPA for nonthreatened comparison students, and the coefficient b_{12} is the difference in this trend for potentially threatened comparison students. We expected the estimates for these coefficients to both be negative, representing a downward trend in GPA during this transition period overall and a steeper decline for potentially threatened students. The extent to which the intervention's effect on academic trajectory was stronger for threatened students is reflected by the coefficient b_{13} . A positive value for this coefficient would indicate that the intervention improved the academic trajectories of potentially threatened students, relative to similar students in the comparison condition. Finally, student covariates (i.e., sixth-grade GPA, gender, eligibility for free and reduced-price lunch, special education status, limited English proficiency, and indicator variables for schools), represented by the expression $b\mathbf{X}_i$, were included to increase the precision of the impact estimates by accounting for more variance in the outcome variable.

The impact of the intervention for potentially threatened students persisted through the end of ninth grade, $b_{03} = 0.177$, 95% confidence interval (CI) = [0.066, 0.289], $t(9161) = 3.11$, $p = .002$, $d = 0.174$. Postintervention GPAs declined for nonthreatened comparison students through this transition period, $b_{10} = -0.072$, 95% CI = [-0.089, -0.056], $t(9161) = -8.50$, $p < .001$, with potentially threatened students experiencing significantly greater declines, $b_{12} = -0.118$, 95% CI = [-0.146, -0.090], $t(9161) = -8.26$, $p < .001$. However, the intervention slowed these declines for potentially threatened

Table 1. Estimates From Latent Growth Models of Grade Point Average in Grades 7 Through 9

Predictor	Main: all students (Model 1)	Simple: all students (Model 2)	Simple: stereotype-threatened students (Model 3)
Treatment	0.036 (0.034)	0.100* (0.027)	0.206* (0.056)
Potentially threatened	-0.331* (0.045)	-0.099* (0.031)	
Treatment × Threatened	0.177* (0.057)		
Years (slope)	-0.072* (0.009)	-0.114* (0.007)	-0.190* (0.014)
Years × Threatened	-0.118* (0.014)		
Years × Treatment	0.017 (0.012)	0.034* (0.010)	0.061* (0.020)
Years × Treatment × Threat	0.043* (0.020)		
Number of observations	9,161	9,161	3,213
Number of students	920	920	324

Note: The table shows unstandardized regression coefficients (standard errors are given in parentheses). All models included covariates that are not shown here but can be found in Table S2 in the Supplemental Material. Main-effects models included interaction terms with student background to estimate effects for threatened students; simple-effects models directly estimated the effect for the subgroup of interest.

* $p < .05$.

students, $b_{13} = 0.043$, 95% CI = [0.004, 0.083], $t(9161) = 2.15$, $p = .031$, such that over time, treated students distanced themselves from their peers in the control condition. The full model estimates are presented in Table 1 and unadjusted GPA over time is shown in Figure 1.

As a check of the robustness of our growth model, we also considered alternate specifications of GPA trajectories. In addition to the simple linear model presented above, a linear model including grade intercepts, a linear model including grade intercepts and yearly slopes, and a quadratic model were tested. Results are

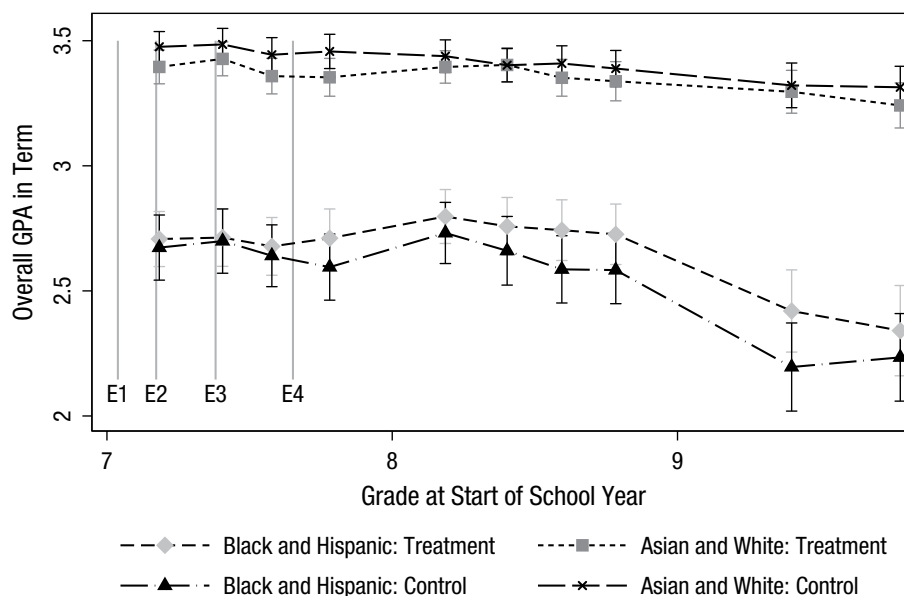


Fig. 1. Unadjusted grade-point-average (GPA) trends from Grades 7 through 9 for the four combinations of experimental condition (treatment vs. control) and stereotype-threat group (Black and Hispanic vs. Asian and White). Data points represent the end of the marking period. The labels "E1" through "E4" indicate median implementation time of each of the four exercises. Error bars represent 95% confidence intervals.

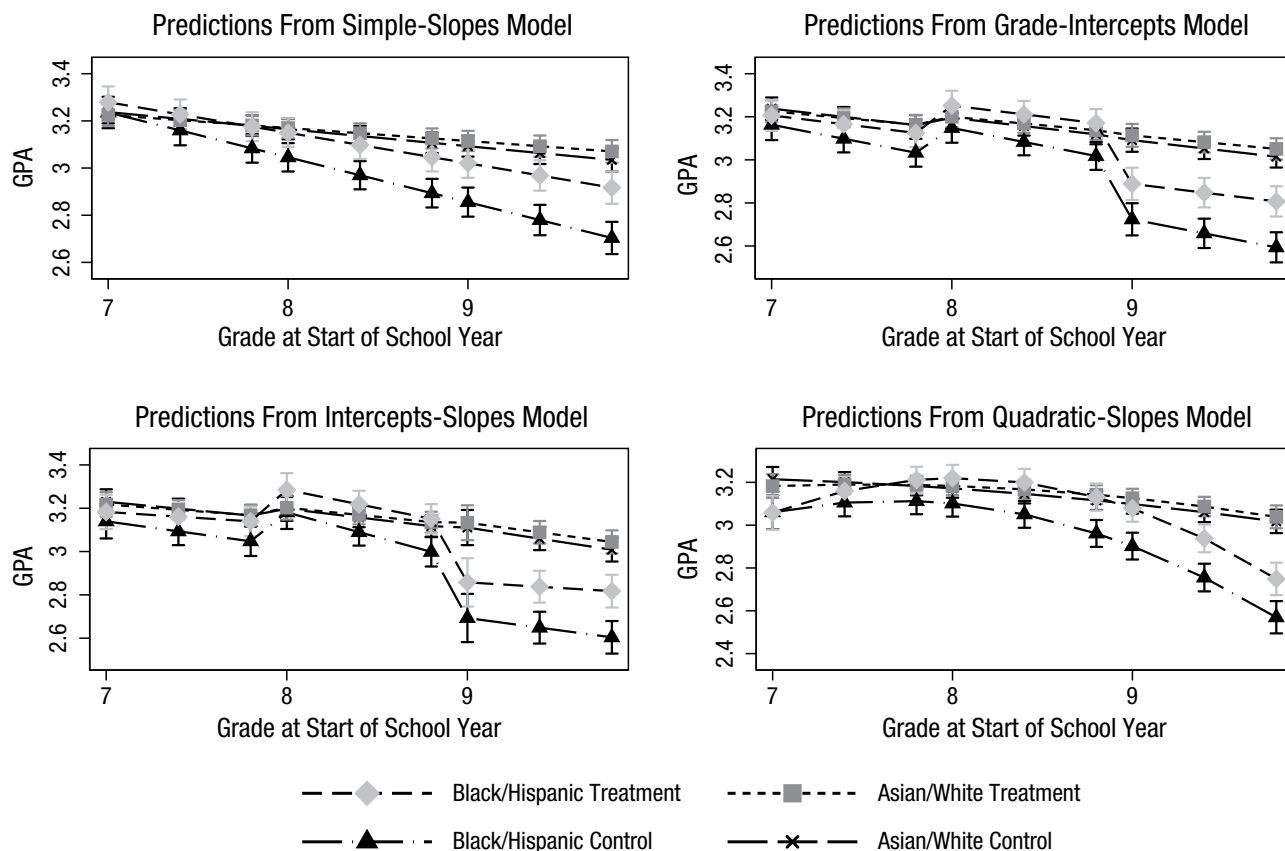


Fig. 2. Robustness checks for latent growth models of grade point average (GPA) in Grades 7 through 9. Results are shown for the four combinations of experimental condition (treatment vs. control) and stereotype-threat group (Black and Hispanic vs. Asian and White), separately for each of the four models. Error bars represent 95% confidence intervals.

summarized by predicted values from each model (Fig. 2). The more flexible specifications highlight notable aspects of GPA trends over time: Students tend to experience a decline throughout seventh grade, an uptick at the start of eighth grade, declines throughout that year, and a large drop after the transition to high school. However, the preponderance of evidence from these additional analyses supports the key conclusions across all models. Potentially threatened students fared worse in ninth grade than their nonthreatened peers, but assignment to the self-affirmation condition reduced this discrepancy by narrowing the growing differences between threatened and nonthreatened students.

Hypothesis 2: does school context moderate lasting intervention effects?

To test whether the ninth-grade impact of the intervention was greater for students who attended a middle school with a high-threat context, we estimated a second multilevel growth model. We focused on only the population of Black and Hispanic students in the sample for simplicity of presentation and because intervention effects were restricted to this student subgroup:

$$\begin{aligned} \text{GPA}_{it} = & b_{00} + b_{01}\text{treatment}_i + b_{02}\text{high threat}_i \\ & + b_{03}\text{treatment}_i \times \text{high threat}_i + b_{10}\text{year}_i + b_{11}\text{treatment}_i \\ & \times \text{year}_{it} + b_{12}\text{high threat}_i \times \text{year}_{it} + b_{13}\text{treatment}_i \\ & \times \text{high threat}_i \times \text{year}_{it} + b\mathbf{X}_i + \eta_i + \varepsilon_{it}. \end{aligned}$$

This is a similar model to that used in the analysis for our first hypothesis. However, the threatened_i variable, including its interactions, was replaced with the high-threat_i variable, because all students analyzed here were potentially threatened students. We were now interested in the interaction between the treatment and a high-threat school context. The outcome variable and covariates are the same as those described in our first hypothesis. The impact of the intervention on end-of-year ninth-grade GPA for potentially threatened students not in high-threat contexts is the coefficient b_{01} . We expected that the estimate for this coefficient would be positive, indicating that all Black and Hispanic students benefited from the intervention, even if they were not in a high-threat school context. The coefficient β_{03} is the additional impact of treatment for potentially threatened students in high-threat school contexts. We expected this estimate to be positive as well, indicating

Table 2. Estimates From Latent Growth Models of Grade Point Average in Grades 7 Through 9 for Potentially Stereotype-Threatened Students

Predictor	Main: threatened students (Model 1)	Simple: threatened students (Model 2)	Simple: threatened students in high-threat contexts (Model 3)
Treatment	0.082 (0.083)	0.206* (0.056)	0.320* (0.077)
High-threat context	−0.217 (0.165)		
Treatment × High Threat	0.232* (0.112)		
Years (slope)	−0.217* (0.021)	−0.190* (0.014)	−0.170* (0.019)
Years × High Threat	0.047 (0.028)		
Years × Treatment	0.045 (0.029)	0.061* (0.020)	0.078* (0.028)
Years × Treatment × High Threat	0.033 (0.040)		
Number of observations	3,213	3,213	1,755
Number of students	324	324	177

Note: The table shows unstandardized regression coefficients (standard errors are given in parentheses). All models include covariates that are not shown here but can be found in Table S3 in the Supplemental Material. Main-effects models included interaction terms with student background to estimate effects for threatened students; simple-effects models directly estimated the effect for the subgroup of interest.

* $p < .05$.

that potentially threatened students in high-threat school contexts benefited to a greater extent from the intervention than their potentially threatened peers in low-threat contexts.

The results suggest that school context strongly moderated the impact of the self-affirmation intervention. The estimated impact at the end of ninth grade for potentially threatened students who were not in high-threat contexts was positive but not statistically significant, $b_{01} = 0.082$, 95% CI = $[-0.080, 0.245]$, $t(3213) = 1.00$, $p = .319$, $d = 0.080$. The additional impact at the end of ninth grade for potentially threatened students in high-threat school contexts was substantial, $b_{03} = 0.232$, 95% CI = $[0.012, 0.451]$, $t(3213) = 2.07$, $p = .038$, $d = 0.227$. The simple effect for threatened students in high-threat contexts was statistically significant, $b = 0.320$, 95% CI = $[0.168, 0.471]$, $t(1755) = 4.15$, $p < .001$; $d = 0.314$. The full model estimates are presented in Table 2.

Hypothesis 3: to what extent does engaging in self-affirming writing—as induced by the intervention—predict academic performance?

We estimated the impact of self-affirming writing on the cumulative ninth-grade GPA of potentially threatened students using an instrumental-variables model. Instrumental-variables methods can be used with experimental data to

produce unbiased causal estimates of impacts for endogenous predictors (Angrist, Imbens, & Rubin, 1996). Self-affirming writing is endogenous, as factors that influence a student's writing, such as his or her inclination to follow instructions or tendency to spontaneously self-affirm, may also contribute to that student's academic performance. By using the randomized treatment condition as an instrument for self-affirmation, however, we can carve out a portion of the variation in self-affirming writing that is exogenous. By assuming that the intervention impacts student outcomes exclusively via self-affirming writing (and, by the same token, that students who did not engage in self-affirming writing were unaffected by the intervention), this analysis provides a causal estimate of the impact of the treatment for students who activated the planned psychological mechanism (i.e., self-affirming writing) as a result of the intervention. Specifically, the estimate corresponds to a causal contrast between the treatment students who engaged in self-affirming writing as a result of assignment to the intervention and those comparison students who would have self-affirmed had they been assigned to the treatment condition. Since there is no reason to believe that a brief writing exercise would have a long-term impact on a student's GPA outside of the planned psychological mechanism, this method appropriately tests the third hypothesis.

To perform this analysis, we estimated the following models using two-stage least squares regressions for

Table 3. Estimates From Models of Grade 9 Overall Grade Point Average for Potentially Stereotype-Threatened Students

Predictor	Simple: threatened students (Model 1)	Simple: threatened students in high-threat contexts (Model 2)	Simple IV: threatened students (Model 3)	Simple IV: threatened students in high-threat contexts (Model 4)
Treatment	0.186* (0.0874)	0.283* (0.117)		
Self-affirming writing			0.252* (0.108)	0.382* (0.156)
Number of students	324	177	324	177
Number of schools	11	7	11	7

Note: The table shows unstandardized regression coefficients (standard errors are given in parentheses). All models include covariates that are not shown here but can be found in Table S4 in the Supplemental Material. Models 1 and 2 are multilevel models, and Models 3 and 4 include school fixed effects. IV = instrumental variables.

* $p < .05$.

the population of Black and Hispanic students in the sample:

$$\text{self-affirm} = \alpha_0 + \alpha_1 \text{treatment} + \alpha X + \delta$$

$$\text{GPA}_{9\text{th}} = \beta_0 + \beta_1 \text{self-affirm} + \beta X + \varepsilon$$

In a two-stage least squares instrumental-variables regression, the endogenous explanatory variable (i.e., self-affirming writing) was regressed on the instrumental variable (i.e., treatment assignment) in the first stage. The predicted values of the endogenous variable from the first stage were then used in place of the actual values to predict the outcome variable (i.e., GPA_{9th}, the ninth-grade cumulative GPA) in the second stage, since these predicted values are exogenous. The model in the second stage addressed our third hypothesis. The coefficient β_1 is the average treatment effect of self-affirming writing induced by the intervention on the cumulative ninth-grade GPA of potentially threatened students; that is, the effect of the TOT. This estimate represents both the effect of self-affirming writing for those who engaged in it as well as the difference from those who did not produce self-affirming writing. The covariates were included in both stages of the instrumental-variables regression, as before, to improve precision.

The simple treatment effect for all potentially threatened students at the end of ninth grade, $b = 0.186$, 95% CI = [0.015, 0.358], $t(324) = 2.13$, $p = .033$; $d = 0.182$ was comparable in magnitude with the result for Hypothesis 1. The effect of self-affirming writing for potentially threatened students was statistically significant, $b_1 = 0.252$, 95% CI = [0.040, 0.464], $t(324) = 2.33$, $p = .020$; $d = 0.247$.⁷ The same was true for students who exhibited self-affirming writing in high-threat school contexts, $b_1 = 0.382$, 95% CI = [0.077, 0.688], $t(177) = 2.45$, $p = .014$; $d = 0.375$. These results suggest that there were substantial performance benefits for

students who engaged in self-affirming writing as a result of receiving the self-affirmation intervention. The full model estimates are presented in Table 3.⁸

Discussion

The present study demonstrated that the impact of a self-affirmation intervention can persist through the transition to high school. Black and Hispanic students in the treatment group benefited from the intervention beyond their middle-school experiences such that they now are on different academic trajectories from their untreated peers (see Fig. 2). Furthermore, this is the first replication of long-term effects of self-affirmation for secondary-school students by an independent research team and the first demonstration of these long-term effects at scale.

Throughout this crucial transition period from middle school to early high school, the racial-achievement gap observed in our study widened on average by 0.12 GPA points per year, conditional on covariates. Self-affirmation reduced this growth by one third to 0.08 GPA points per year, which through the end of ninth grade resulted in a 50% reduction of the residual racial-achievement gap. This is a substantial reduction in the residual racial-achievement gap that is posited to stem from identity-threat processes (Steele, 1997), and it also corresponds to a practically meaningful 18% reduction in the raw gap in ninth grade GPA, from 1.1 to 0.9 points. Further research is needed to determine whether improved academic trajectories and narrowed GPA achievement gaps can be maintained throughout high school and, then, whether these changes also persist across the next transition into postsecondary education (a long-term follow-up of the initial self-affirmation study by Goyer et al., 2017, suggests that this may be the case).

Furthermore, the present study assessed how contextual and individual factors might help explain

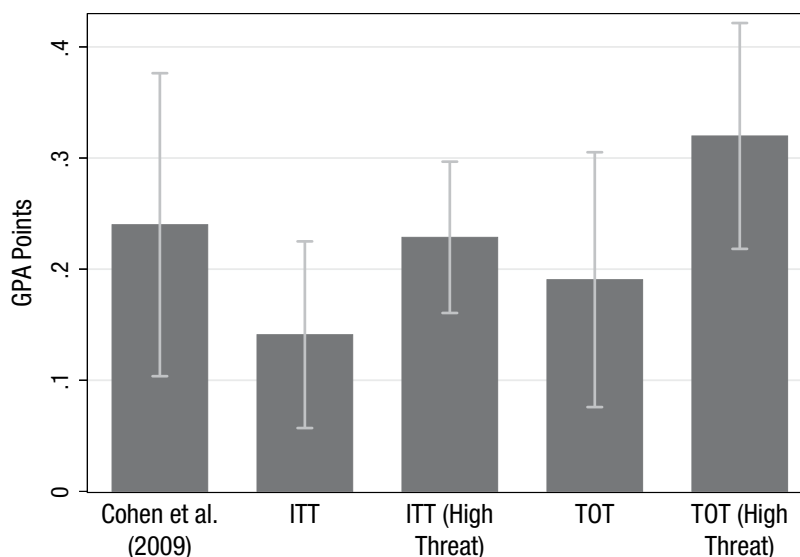


Fig. 3. Mean Grade 8 grade point average (GPA) for each of the four groups in the current study—intent-to-treat (ITT) overall, ITT in high-threat schools (Hypothesis 2), treatment on the treated (TOT) overall, and TOT in high-threat schools (Hypothesis 3)—and of the sample analyzed by Cohen, Garcia, Purdie-Vaughns, Apfel, and Brzustoski (2009). ITT estimates were calculated with school fixed effects for comparability. Error bars represent 95% confidence intervals.

heterogeneity in self-affirmation effects. With respect to context, the intervention primarily impacted Black and Hispanic students who attended middle schools where they were a small minority and had lower academic standing. These conditions were likely prone to producing stereotype threat among students (Borman & Pyne, 2016). Additionally, we found that the nearly three quarters of Black and Hispanic students who received the intervention and completed it as intended reaped substantial benefits from the intervention, suggesting that how students engage with the intervention matters, as well as school context. Accounting for both context and engagement with the intervention, treated students in high-threat contexts who self-affirmed in their writing—the TOT impact—experienced effects that were equal to or greater than those reported in the original published research by Cohen et al. (2009; see Fig. 3).

These three key findings suggest that long-term benefits from self-affirmation depend on local conditions (i.e., high-threat school contexts) and individual responses (i.e., self-affirming writing responses due to the prompts), both of which are likely to vary if these strategies are implemented widely. The initial field trials of self-affirmation interventions necessarily took place in a few schools with more highly involved researchers (Cohen et al., 2006), potentially preserving uniformly high fidelity. Scale-up studies, such as this one, are critical for assessing and examining potential factors that affect treatment efficacy, both for policymakers

who need to know where and for whom wise interventions work and for the advancement of theory through field tests of lab-discovered moderators.

This study sheds light on how self-affirmation interventions might function in real-world settings. Black and Hispanic students may vary in their exposure to stereotype threat, and students who might benefit from the exercises may not engage as expected. However, these results also suggest that greater understanding of who most needs the intervention and how to encourage a sufficient level of engagement can increase the effectiveness of self-affirmation interventions. Notwithstanding the need to refine how we target and deploy interventions, our study also adds to the burgeoning evidence that brief self-affirmation exercises, which can be substituted into a curriculum at almost no cost, can have substantial effects for certain students in particular school contexts. Consequently, they should be considered by local school administrators as part of a strategy to address racial-achievement gaps, and future research should continue to refine the growing intervention science in psychology.

Action Editor

Bill von Hippel served as action editor for this article.

Author Contributions

G. D. Borman developed the study concept and design. J. Grigg, C. S. Rozek, and P. Hanselman contributed to the study design. C. S. Rozek developed and supervised the coding

protocol. Data were collected by J. Grigg, C. S. Rozek, and P. Hanselman. P. Hanselman initially analyzed and interpreted the data, and N. A. Dewey performed additional analyses under the supervision of J. Grigg. J. Grigg and N. A. Dewey drafted the manuscript, and G. D. Borman, P. Hanselman, and C. S. Rozek provided critical revisions. All authors approved the final version of the manuscript for submission.

ORCID iD

Jeffrey Grigg  <https://orcid.org/0000-0003-1975-3232>

Acknowledgments

We thank Geoffrey Cohen, Joshua Aronson, and Valerie Purdie-Vaughns for advice during the design of this project; Jaymes Pyne, Alex Schmidt, Jennifer Corley, and Sadie Millen for assistance with the data; and Marcia Davis, Kelly Siegel-Stechler, and Dhathri Chunduru for reviewing the manuscript.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The research reported here was supported by grants from the U.S. Department of Education (R305A110136) and the Spencer Foundation (201500044). The content is the responsibility of the authors and does not necessarily represent the views of supporting agencies.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618784016>

Open Practices

Data and materials for this study have not been made publicly available, and the design and analysis plans for the experiments were not preregistered.

Notes

1. Thirty-six students (3%) were missing data in Grade 6, 45 more students (4%) were missing data in Grade 8, and an additional 47 students (4%) were missing data in Grade 9.
2. See Borman et al. (2016) for a detailed account of the procedure used in the trial.
3. The covariates increased the precision of the estimates in our data. For example, the standard error of the estimated interaction between student background and treatment assignment decreased from 0.093 to 0.057 (see Table S2 in the Supplemental Material).
4. We also examined the context of the high schools the students attended and found that half of them could be considered potentially threatening. In this setting, high school enrollment—like middle school enrollment—was determined by residence, so students followed a prescribed “feeder pattern.” For 76% of students in the sample, and 69% of potentially threatened

students, the context was consistent across the transition to high school (high-high or low-low). Student experiences undoubtedly vary within high schools more than they do within middle schools, but we lack data on the localized contexts students experienced (e.g., academic track) in high school.

5. Students were also prompted to explain why their chosen value was important to them. We did not include this explanation in our coding of self-affirming writing because writing about a nonthreatened aspect of the self alone should lead to the hypothesized mechanism of self-affirmation: a broadened perspective in threatening contexts.

6. Conditional on whether coders agreed that the student wrote about an item from the provided list, agreement was still substantial, $\kappa = 0.66$, $N = 3,592$, $p_0 = .84$.

7. We found that potentially threatened students in the intervention condition who did not engage in self-affirming writing were statistically indistinguishable from similar students in the comparison condition ($b = 0.074$, 95% CI = $[-0.22, 0.37]$, $p = .624$).

8. We also tested models for more and less restrictive definitions of self-affirming writing throughout the year. These different definitions led to different-sized effect estimates but produced the same basic conclusion that self-affirming writing had substantial impacts on GPA (see Table S5 in the Supplemental Material).

References

- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444–455.
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78, 246–263.
- Borman, G. D. (2017). Advancing values affirmation as a scalable strategy for mitigating identity threats and narrowing national achievement gaps. *Proceedings of the National Academy of Sciences, USA*, 114, 7486–7488.
- Borman, G. D., Grigg, J., & Hanselman, P. (2016). An effort to close achievement gaps at scale through self-affirmation. *Educational Evaluation and Policy Analysis*, 38, 21–42.
- Borman, G. D., & Pyne, J. (2016). What if Coleman had known about stereotype threat? How social-psychological theory can help mitigate educational inequality. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(5), 164–185.
- Brady, S. T., Reeves, S. L., Garcia, J., Purdie-Vaughns, V., Cook, J. E., Taborsky-Barba, S., . . . Cohen, G. L. (2016). The psychology of the affirmed learner: Spontaneous self-affirmation in the face of stress. *Journal of Educational Psychology*, 108, 353–373.
- Bratter, J. L., Rowley, K. J., & Chukhray, I. (2016). Does a self-affirmation intervention reduce stereotype threat in Black and Hispanic high schools? *Race and Social Problems*, 8, 340–356.
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *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.
- Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, 65, 333–371.
- Cook, J. E., Purdie-Vaughns, V., Garcia, J., & Cohen, G. L. (2012). Chronic threat and contingent belonging: Protective benefits of values affirmation on identity development. *Journal of Personality and Social Psychology*, 102, 479–496.
- Critcher, C. R., & Dunning, D. (2015). Self-affirmations provide a broader perspective on self-threat. *Personality and Social Psychology Bulletin*, 41, 3–18.
- Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness*, 8, 149–168.
- Dong, N., & Maynard, R. (2013). *PowerUp!*: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, 6, 24–67.
- Goyer, J. P., Garcia, J., Purdie-Vaughns, V., Binning, K. R., Cook, J. E., Reeves, S. L., . . . Cohen, G. L. (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. *Proceedings of the National Academy of Sciences, USA*, 114, 7594–7599.
- 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, 106–124.
- Hanselman, P., Rozek, C. S., Grigg, J., & Borman, G. D. (2017). New evidence on self-affirmation effects and theorized sources of heterogeneity from large-scale replications. *Journal of Educational Psychology*, 109, 405–424.
- Harackiewicz, J. M., & Priniski, S. J. (2018). Improving student outcomes in higher education: The science of targeted intervention. *Annual Review of Psychology*, 69, 409–435.
- Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, 326, 1410–1412.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Muthén, B. (1997). Latent variable modeling of longitudinal and multilevel data. *Sociological Methodology*, 27, 453–480.
- Protzko, J., & Aronson, J. (2016). Context moderates affirmation effects on the ethnic achievement gap. *Social Psychological & Personality Science*, 7, 500–507.
- Purdie-Vaughns, V., Cohen, G. L., Garcia, J., Sumner, R., Cook, J. C., & Apfel, N. (2009). Improving minority academic performance: How a values-affirmation intervention works (ID No. 15774). *Teachers College Record*. Retrieved from <http://www.tcrecord.org/content.asp?contentid=15774>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.
- Rozek, C. S., Svoboda, R. C., Harackiewicz, J. M., Hulleman, C. S., & Hyde, J. S. (2017). Utility-value intervention with parents increases students' STEM preparation and career pursuit. *Proceedings of the National Academy of Sciences, USA*, 114, 909–914.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336–356.
- 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, 591–618.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629. doi:10.1037/0003-066X.52.6.613
- 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.
- Tibbetts, Y., Harackiewicz, J. M., Canning, E. A., Boston, J. S., Priniski, S. J., & Hyde, J. S. (2016). Affirming independence: Exploring mechanisms underlying a values affirmation intervention for first-generation students. *Journal of Personality and Social Psychology*, 110, 635–659.
- Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, 23, 73–82.
- Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331, 1447–1451.
- 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.