Social Psychological and Personality Science
1-8

DOI: 10.1177/1948550616646426
spps.sagepub.com

Context Moderates Affirmation Effects on the Ethnic Achievement Gap

John Protzko[1] and Joshua Aronson[2]

Abstract

We attempted to replicate a self-affirmation intervention that produced a 40% reduction in the academic achievement gap among at-risk students. The intervention was designed as a protection against stereotype threat—, which creates stress and suppresses the performance, engagement, and learning of students stereotyped as intellectually inferior. In previous research, Black and Hispanic students who engaged in a values-affirmation exercise significantly improved their academic performance over the course of a school semester. We attempted to replicate these salutary effects in both an inner-city school and a more wealthy suburban school—contexts not tested in the original research. Despite employing the same materials, we found no effect of the affirmation on academic performance. We discuss these results in terms of the possibility that negatively stereotyped students benefit most from self-affirmations in environments where their numbers portray them neither as clearly "majority" nor minority.

Keywords

affirmation, stereotype threat, replication, achievement

Introduction

This report concerns the use of self-affirmation to improve the academic performance of at-risk students. Specifically, we consider the remarkable finding that a brief writing assignment designed to buffer social stressors in schools can produce impressive academic gains (Cohen, Garcia, Apfel, & Master, 2006; Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009). This work grew from two well-established literatures on self-affirmation (Steele, 1988) and stereotype threat (Steele & Aronson, 1995).

Self-affirmation, directing one's attention to valued aspects of the self, was first proposed as an alternative route to resolving cognitive dissonance (Steele, 1988). In a series of influential studies, Steele found that people were able to resist the normal urge to rationalize bad decisions or failure, if they were provided an opportunity to reflect upon an important source of self-worth (e.g., values, talents, and relationships). People who smoke cigarettes, for example, may not feel the need to distort the risks of smoking (it's really not that unhealthy) if they can focus on unrelated source of self-worth (e.g., "I'm great at my job"). Affirmations thus enable people to reduce the stress arising from inconsistencies and other self-image threats without distorting reality or changing their behavior, so long as they can cast themselves in a favorable light.

Self-affirmation effects are plentiful and diverse in both laboratory and field experiments (Sherman & Cohen, 2006). The approach is very straightforward: provide some participants in a threatening situation the opportunity to self-affirm and compare their responses to those of a control group. Self-affirmation appears to replenish self-regulatory capacity in situations that deplete it, thus increasing a person's ability to maintain performance, tolerate pain, resist temptations, or distractions. In one study, for example, participants had to write a story *without* using the letters *a* or *n*. This demanding task reduced the writers' subsequent pain tolerance relative to those given an easier task; the opportunity to self-affirm after the writing task eliminated this "ego-depletion effect" (Schmeichel & Vohs, 2009). Similarly, task failure tends to make people ruminate and rationalize (e.g., "the test was biased"); affirmations reduce this tendency (Koole, Smeets, van Knippenberg, & Dijksterhuis, 1999). Affirmations have even been shown to prevent the release of cortisol—a stress-induced hormone that interferes with learning (Creswell et al., 2005). Affirmations thus appear to be a reliable way to reduce stress and improve performance in threatening situations.

Affirmations and Stereotype Threat

One obvious place to apply the potential of self-affirmations is stereotype threat. Stereotype threat occurs when people are

[1] Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA
[2] Applied Psychology, New York University, New York, NY, USA

Corresponding Author:
John Protzko, Psychological and Brain Sciences, University of California, Santa Barbara, CA 93106, USA.
Email: protzko@gmail.com

confronted with the risk of confirming a negative ability stereotype tied to their social identity. Having a Black student indicate their ethnicity before taking a difficult test can induce anxiety and impair performance, for example, by reminding the student of a well-known stereotype of Blacks as intellectually inferior (Steele & Aronson, 1995). By reducing the inherent threat in such situations, self-affirmations can reduce the performance decrements resulting from stereotype threat (Sherman & Cohen, 2006). The value affirmation procedures have been brought into the field to test the proposition that long-term performance might benefit from affirmation in a potentially threatening school context. In the wake of the success of this minimal intervention—which cheaply and with little effort appears capable of improving learning among low performing groups—we undertook a conceptual replication of this finding.

The intervention (Cohen et al., 2006a, study 1) is remarkably simple. Students are asked to select from a list of values (e.g., happiness, sense of adventure, drive or grit, personal relationships, political views, arts and music abilities, popularity, etc.) and explain in writing why the value is personally important. The original intervention removed up to 40% of the ethnic achievement gap in a middle school by improving the performance of low achieving Black and Hispanic students. Subsequent research suggests that the affirmations replicate and produce gains that can persist for as long as 2–3 years (Borman, 2012; Cohen et al., 2009; Cook, Purdie-Vaughns, Garcia, & Cohen, 2012; Sherman et al., 2013) and can also reduce the gender-achievement gap in undergraduate physics by lifting the performance of women (Miyake et al., 2010). Self-affirmations appear to bolster performance by both buffering the threat and reducing defensiveness about low performance or negative feedback (Cohen & Sherman, 2014).

In addition to replicating these effects, our goal was to examine the robustness of the intervention across settings and to shed light on potential contextual moderators of the effect. Such considerations are weighty given two facts. First, to forecast our findings, some affirmation interventions do not replicate. Second, this intervention has been added to the What Works Clearinghouse of the U.S. Institute of Education Sciences, meaning it is likely to be taken up by schools in search of an evidence-based way to improve education outcomes. For these reasons, we believe research on when and where the procedure is likely to work underscores the utility of close analysis of both successful and unsuccessful replications. Self-affirmations are clearly promising in their potential to boost performance; we hope to shed light on how an eventual user can employ the intervention with the greatest chance of success.

Due to limitations placed on us by the local Department of Education, we were unable to conduct an exact replication of the Cohen, Garcia, Apfel, and Master (2006) procedure. Our intervention differed in that the affirmation was, by necessity, presented as coming from the research team. In the Cohen et al., intervention, the exercise was presented by the teacher as a class assignment, which may have led students to assume that their teacher was interested in knowing the contents of their affirmations (Cohen et al., 2006). Whether this was a critical difference is unclear and is taken up in the discussion. The difference we sought to explore, however, was the setting—specifically, the ethnic composition of the school. In the original intervention (Cohen et al., 2006; study 1) and subsequent replications (e.g., Sherman & Cohen, 2006; Cook et al., 2012; Sherman et al., 2013), the composition of the schools and classrooms—the degree of ethnic representation—is remarkably balanced across Black, Hispanic, White, and Asian students (roughly 50% Black and Hispanic and 50% White and Asian). By contrast, most U.S. school systems are de facto segregated, with either a majority or a minority of Black and Hispanic students (Sethi & Somanathan, 2004; Eaton, 2008). Therefore, an intervention of this sort may produce different effects depending on the demographic mix of students. It has been argued, for example, that the psychological impact of stereotypes may depend importantly on whether a group has "critical mass" in an environment (Steele, 2010). Thus, there are grounds for suspecting that the degree of ethnic representation moderates the effectiveness of self-affirmation interventions.

Material and Method

Participants and Design

We conducted the affirmation intervention in two schools: one, an inner-city New York City school and a second in a wealthier district in upstate New York. In both schools, we sampled the ninth-grade class, the freshman year of high school, which represents a challenging transition for many students and contexts for experiences of stereotype threat (Steele & Aronson, 1995). The inner-city school was small and homogenous with 126 students (94% Black and Hispanic) while the upstate school had 328 students (26% Black and Hispanic).

In the original intervention, there were 119 "negatively stereotyped" students (Black and Hispanic) and 124 "positively stereotyped" students (White and Asian); because we tested a larger number of students, any null finding is unlikely due to a lack of power. We randomly assigned participants to either the experimental or control condition using student lists provided by the two schools. The intervention took place within the first 3 weeks of the semester.

Ethics Statement

This research was approved by the New York City Department of Education, The Bedford School District Superintendent's office, and The New York University Committee of Activities Involving Human Subjects. The participants and their parents provided written consent. This research was conducted during the 2009 academic year.

Materials and Procedure

Students were told an outside research team is conducting research on what students find important to them and to other

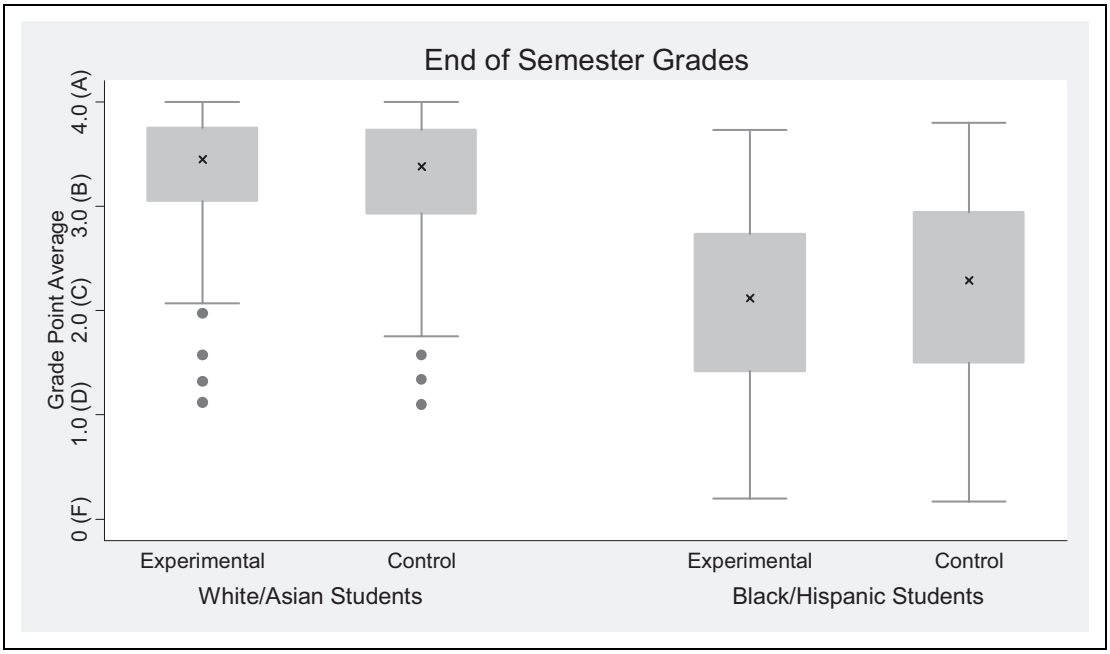

Figure 1. Effects of the intervention on student's academic achievement by ethnicity. Grade point averages White/Asian students: Experimental = 3.316 (*SD* = 0.586), control = 3.262 (*SD* = 0.603); Black Hispanic students: Experimental = 1.991 (*SD* = 1.022), control = 2.051 (*SD* = 1.073). Graph shows analysis of variance results opposed to analysis of covariance (ANCOVA), in keeping with previous concerns (Sackett, Hardison, & Cullen, 2004a, 2004b) about the use of ANCOVA results in graphs.

people. The affirmation manipulation involved handing out packets in manila envelopes to students during class. Packets were identical to one another; only a few of the instructions differed as a function of experimental condition. Each packet contained a list of 11 personal values. Students in the experimental condition circled the three values *most* important to them, while those in the control condition circle those three *least* important to them. All students then wrote about why those three values are most important to them (or, in the control condition, why they might be important to others). Students were instructed not to worry about grammar or spelling. Then, to reinforce the manipulation, students answered a 6-point Likert-type scale (*strongly disagree* to *strongly agree*) on the following four questions: (1) these values have influence my life, (2) in general I try to live up to these values, (3) these values are an important part of who I am, and (4) I care about these values.

As in the original intervention, students filled out the packets and returned them to the envelopes, which were then passed to the teacher. The students and teachers were blind to the hypothesis, experimental manipulation, and the research assistants were kept blind to student condition.

As noted earlier, the other deviation (from study 1 of Cohen et al., 2006) was as follows: In the original intervention as packets were distributed by the teachers and the intervention "was presented as a regular class assignment" (Cohen et al., 2006, p. 1308). Restrictions placed on us by the Department of Education prevented us from presenting the materials as an in-class assignment. Therefore, in our intervention, teachers then passed the packets to research assistants who then left the room. In every phase of the research, we employed the exact same materials (Cohen et al., 2006).

Participants were given consent forms for them and their parents to sign 2 weeks before the experiment. About 10.1% (33 of the 328) of the rural school did not return consent forms and were not included in the analysis, while 40.3% (54 of the 134) of the urban school did not return consent forms and were thus excluded from the analysis. Therefore, 64.5% of participants successfully returned the consent forms. Only one parent refused to allow their child to participate in the study. All participants completed the intervention tasks assigned them. No data were excluded from the analysis nor were there any additional treatment conditions of the intervention.

Results and Discussion

Did the intervention influence student grades? We obtained end of semester grades from school records and submitted them to a 2 × 2 analysis of variance (ANOVA) (Experimental Condition × Black/Hispanic vs. White/Asian Student) to test for differences in whether the intervention changed students end of semester grade point average (GPA). Black and Hispanic students in the two schools performed significantly worse (GPA: 2.02, standard deviation [*SD*] = 1.043) than White and Asian students (GPA: 3.29, *SD* = 0.593), F (1, 344) = 91.36, $p < .001$), but there was no effect of the intervention on either overall GPA or that of Black and Hispanic students specifically (both *p*s >.515). Therefore, as can be seen in Figure 1, we can

Table 1. Students' Most and Least Important Values.

Summary of Regression Model Predicting Grade in Course: Cohen et al. (2006) Experiment 1				Summary of Regression Model Predicting Grade in Course: Replication		
Variable	*B*	*SE*	*T*	*B*	*SE*	*t*
Student race	−0.44	.14	−3.22**	.07	.05	1.63
Student gender	−0.11	.12	−0.089	−.02	.04	−0.43
Teacher code[a]						
Preintervention in-class performance	1.05	.08	13.19**	.02	.005	3.67**
Previous GPA	.007	.08	0.83	.81	.03	23.48**
Experimental condition	.009	.12	0.73	−.001	.03	−0.03
Race × Gender	.03	.15	2.06*	−.1	.05	−2.18*
Gender × Condition	−0.23	.14	−1.59	.004	.04	−0.1
In-Class perf. × Teacher code[b]						
Race × Condition	0.29	.14	2.00*	−.07	.05	−1.49

Note. GPA = grade point average.
[a]The original intervention had three classrooms, this variable was dummy coded. In our analysis, we do so as well for all of the classrooms. This variable was removed in the multilevel model which directly nests students within classrooms. [b]In our analysis, we include this interaction as well for all of the classrooms. In the multilevel model, this variable was removed and instead in-class performance is included as a variable at the classroom-level nesting.
*$p < .05$. **$p < .001$.

conclude that across these two schools, the affirmation intervention did not have the hoped for effect on student grades.

Did the effectiveness of the intervention vary by school? To explore this possibility, we treat school as a variable in a 2 × 2 × 2 (Experimental Condition × Ethnic Stereotype Status × School) ANOVA. Only one difference emerged in this analysis: Students at the suburban school had marginally better end-of-semester grades ($p < .055$) than those in the urban school. No other main effects or interactions approached significance (all ps > .245). Therefore, our attempt to replicate the affirmation effect was equally ineffective in the two schools.

We also explored whether affirmations worked best for the lowest performing students, as has been found in prior research (Sherman et al., 2013). To pursue this possibility, we included the students' previous term GPA as a covariate and interaction term with experimental condition and ethnic stereotype status. This controls for prior differences in achievement and allows an even more powerful test of the effect. In this model, previous term GPA significantly predicted later term GPA ($b = .952$, 95% confidence interval [CI] [0.879, 1.024]), no other main effect or interaction approached significance (all ps > .266).

We also used a time series analysis (four academic grading periods in the semester) to test whether the intervention altered the trajectory of grades for Black and Hispanic students. No such alteration in the academic trajectories was apparent ($p > .582$), neither when controlling for previous grades ($p > .233$) nor when controlling for the interaction between previous school performance and ethnic stereotype status ($p > .519$). We can thus conclude that the intervention did not influence performance in either school for any students.

Analytical replication. In the original analysis of Cohen et al., a number of controls were added to the analysis to find the Significant Race × Experimental Condition interaction (Cohen et al., 2006); this is reproduced in Table 1. Therefore, we use the same variables to exactly replicate the model originally run (right hand of Table 1).

As can be seen in Table 1, the results still fell far short of statistical significance ($p > .137$) when replicating the model used in the original investigation.

Multilevel modeling. In addition, we tested whether nesting the students within classrooms would alter the results, even though the intervention was administered to students within the entire school (not by classroom). Testing the simple model of the intervention, race, and their interaction on end of semester grades, nesting students within classrooms provided significantly better fit to the data, likelihood ratio test $\chi^2(1) = 20.26$, $p < .001$. In addition, nesting students within classrooms within schools provided significantly better fit to the data as well, likelihood ratio test $\chi^2(1) = 3.85$, $p < .05$. Therefore, we proceeded with the multilevel modeling nesting students in both classrooms and schools.

The nested model for race, experimental condition, and their interaction showed that Black/Hispanic students performed significantly worse than White/Asian students ($b = -.865$, 95% CI [−1.119, −0.61]). There was no effect of the intervention ($b = .044$) or the interaction ($b = -.152$, both ps > .35).

To compare the replication model with the nested model, we ran an altered version of the replication model, omitting the teacher code due to the nesting within classes and running the interaction of preintervention in-class performance by teacher code as a single variable at the classroom level. These results were the same as the replication model with the exception that the Race × Gender interaction was no longer significant ($b = -.046$, $p > .3$). Again, the Intervention × Race interaction was not statistically significant ($b = -.064$, $p > .157$).

Finally, the original study excluded Hispanic and Asian students from one of their analyses, but indicated the results were

identical with Hispanic students put in the "negatively stereotyped" and Asian students put in the "positively stereotyped" groups (Cohen et al., 2006). Because more recent affirmation interventions have focused solely on Hispanic students (e.g., Sherman et al., 2013), we combined the groups in the above analyses. None of the results change when analyzing only Black versus White students (race × exp: $p > .313$ for ANOVA model; $p > .608$ for replication model; $p > .277$ for multilevel ANOVA model; and $p > .742$ for multilevel covariate model).

We attempted to replicate a brief social–psychological intervention reported in Cohen et al. (2006) that has been replicated with varying degrees of success by others. In neither of the two schools did our version of the affirmation intervention appear to boost performance as we had hoped.

Possible Reasons for the Different Results

Because we used materials supplied by the authors of the original intervention, it seems reasonable to assume that school demographics moderate the role of affirmation effects. How did the contexts differ? For one thing, in the original intervention and its follow-up (Cohen et al., 2006, 2009), the schools were very close to having equal White and Asian and Black and Hispanic student representation. Later replications have also involved schools with roughly even numbers of positively and negatively stereotyped students (Miyake et al., 2010; Sherman et al., 2013). By contrast, our schools were far less integrated, mirroring the de facto the segregation seen in modern U.S. schools (Orfield, Kucsera, & Siegel-Hawley, 2012).

Why might affirmations not have succeeded in these environments? Stereotype threat may be more likely or intense in settings where one is in a salient minority; it is likely less so if there is a critical mass of one's group mates (Steele, 2010). In a study of East German Orchestras (Allmendinger & Hackman, 1995), for example, males were the overwhelming majority and orchestra members were found to be happy, friendly with each other, and judged themselves to produce good music. As more women entered, however, tensions arose; women began to be treated poorly and feel like outsiders. As the proportion of men and women approached equal representation, the environment improved for the women. By this logic, threats should be greatest where minority representation is large enough to be noticeable yet small enough to produce feelings of being outnumbered (e.g., Inzlicht & Ben-Zeev, 2000). Because of this, affirmations should be *most* effective in schools where Black and Hispanic students feel most outnumbered, rather than in schools with roughly equal representation. Across the successful replications (e.g., Sherman et al., 2013) and this failure to replicate, the evidence in the field would suggest this is not true. While we tested school composition and believe it is the reason for the different results—as this is variable we tested—critical mass theory may not be the mechanism. There are other possibilities for the difference in outcomes aside from school composition.

Table 2. Most and Least Important Values Indicated by Ninth Graders Across Both Schools in Order of Most Chosen.

Sub-Sample	Most Important Values	Least Important Values
Whole sample	Relationships, grades, and sports	Art, politics, and religion
White/Asian	Relationships, grades, and sports	Art, politics, and religion
Black/ Hispanic	Relationships, grades, and music	Art, politics, and sports

Subject Age

The original intervention investigated 7th graders, while we investigated 9th graders. Both are times of stress and change (transition into middle school/high school). In addition, this intervention has worked in college-aged students for combating female stereotype threat in math and science (Miyake et al., 2010) so age is unlikely to be the reason for the discrepancy.

Values Chosen

Because we used the same materials as used in the original intervention rather than materials we created ourselves, it is possible the values were inappropriate for our sample. Were this the case, the affirmation materials may not have affirmed the values of the students. To explore this possibility, we examined the values rated as most and least valued by our participants (Table 2). We found no substantive differences between the values selected by the White and Asian Students and those selected by Black and Hispanic students. While the original intervention does not report which values were selected by students as most and least important, we expect our results to be very similar. In addition, the original report indicated that Black and Hispanic students favored religion more than White and Asian students (Cohen et al., 2006); we also found this pattern in our data. There were no differences in values for urban and suburban students across and within ethnicities. Overall, we do not believe which values chosen is the reason for the discrepancy.

This commonality of values between ours and the original study suggests that participants in our replication took the intervention as seriously and had the same pattern of results as students in the original intervention.

Number of Interventions

In this intervention, we aimed to replicate study 1 of Cohen et al. (2006). This initial intervention included only one affirmation intervention delivered within the first 3 weeks of the semester. Study 2 of the same paper administered two affirmation interventions over the course of the semester. It is possible that a single affirmation is insufficient to reduce the ethnic achievement gap. Yet, in the original intervention, there was no difference in the effectiveness of the once administered versus the twice-administered intervention, indicating a one-time

intervention is likely sufficient to boost performance. Moreover, later investigations have experimented with multiple readministrations of the intervention and found no benefit from these booster interventions (Sherman & Cohen, 2006; Sherman et al., 2013). There does exist the possibility that there is a School Composition × Number of Interventions interaction, where more sessions are necessary in more segregated schools. We present it as a possibility and a direction for future research.

Teacher Role in the Intervention

As noted, our intervention differed procedurally in that we could not present the affirmations as an in-class assignment, as was done in the original intervention. Rather, the affirmation was presented as information of interest to outside researchers, not a part of regular instruction. If this were the reason our students did not benefit from the affirmation intervention, it would suggest that such affirmation effects may derive their potency in part due to the students thinking their teacher cares to know about their values. Many laboratory studies, however, do not have experimenters read the affirmations and often employ different experimenters to administer stressful tasks than those who administered the affirmation conditions (e.g., Creswell et al., 2005). We cannot fully rule out this alternative hypothesis and thus recommend that future studies treat this as a variable. It is not unreasonable to assume that a process that involves recursive social processes (Yeager & Walton, 2011) would involve such perceptions. Without such direct manipulation, the possibility that this is a causal moderator remains speculative.

Another possibility is that the affirmation intervention is not moderated by ethnic composition and instead the discrepant results come from sampling error. Although this is the least interesting possibility, it would call into question the overall validity of the affirmation intervention. This work is both larger in scale and more powerful than Cohen et al. (2006). It may be that there exists substantial publication bias as the smaller studies of the original intervention and its replications show larger, significant effects and this more powerful study shows no such effect (Bakker, van Dijk, & Wicherts, 2012). While a possibility, we do not believe this is the case as the intervention is also backed by replicated laboratory studies (Sherman & Cohen, 2006).

On replications. Replications in general suffer from concerns about interpretability (e.g., Maxwell, Lau, & Howard, 2015), especially when the result does not reach statistical significance. We suspect that the reason we failed to find a self-affirmation effect is contextual rather than conceptual; that is, self-affirmation works, but not everywhere.

If one were to reject this premise and simply consider the results here as a "failed replication," we could better understand the extent by adopting the "small telescopes" evaluation of replication results (Simonsohn, 2015).

The interpretation of Figure 2 is as follows: The simple effect of the intervention on negatively stereotyped students

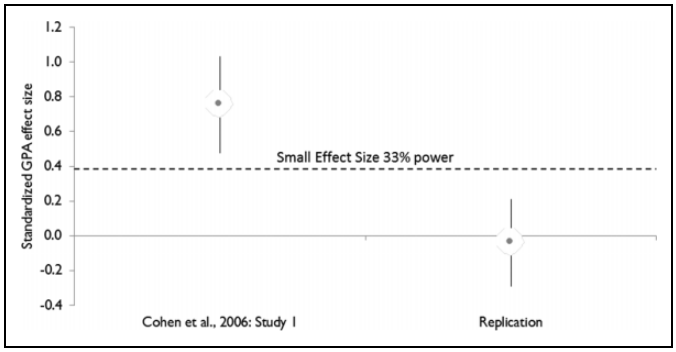

Figure 2. Results from the original intervention and the present replication of it. The markers indicate standardized grade point average differences of the intervention on Black/Hispanic students. Error bars are 95% confidence intervals. The dashed line indicates the effect size that would give the original study, with 21 negatively stereotyped students per condition, 33% power on a one-tailed test.

from the original Cohen et al. study was large ($d = .753$). This was calculated by deriving the pooled variance by dividing the GPA difference (.26) by the t value (2.44). Assuming homogeneity of variance and equal n's would mean the SD of GPA for negatively stereotyped students was .345. This means a .26 GPA difference is a standardized effect size of .753.

Since the CI of the original simple effect does not cross the 33% power line in Figure 2, we can reject the idea that the original study was severely underpowered, on this criterion (Simonsohn, 2015).

The simple effect on negatively stereotyped students in this replication is significantly smaller than both the original effect size, the 33% power line, while not being significantly different from zero. Our replication, in other words, is not an effect size that could have been detectably different from zero with the original sample. As we can also see from Figure 2, there is no overlap whatsoever in the effect sizes. As will be discussed, this cannot be seen then as a result of low statistical power that actually upholds the original results (e.g., Maxwell et al., 2015).

Power Analysis

We conducted a post hoc power analysis. Unfortunately, there was not enough information in the original Cohen analysis to calculate an effect size for the covariate-adjusted interaction. Without this effect size, we cannot calculate the power of the current study for the interaction in the replication model. In addition, it is unclear from the original or supplemental study materials whether the unstandardized GPA difference in study 1 for negatively stereotyped students in the experimental versus control condition (.26) is covariate adjusted or not. From the data in study 1, we calculated the standardized effect size of .753 on the GPA of negatively stereotyped students.

To detect this with 80% power, we would have needed 29 participants per group. Instead, with our 206 Black and Hispanic students, we had 99.959% power to detect a comparable effect size in end of quarter GPA in response to the

intervention. By contrast, the original study 1 had 66.318% power to detect a simple main effect of the intervention on Black and Hispanic students (note that the real power under the covariates model would be higher due to a decrease in error variance). Therefore, our results are not due to a lack of power, nor are the original results a product of being underpowered.

Conclusion

Overall, our results add to the growing literature, suggesting that affirmation interventions, while clearly effective sometimes, may require certain conditions to be operative. It will not be enough for the field to state that affirmation interventions only work in some contexts; the value of such a statement should be turned into the scientific pursuit of *how* (e.g., Lewin, 1931). Our results suggest two possibilities for *how* demographic composition could moderate the effects of affirmation interventions on the academic performance of Black and Hispanic students. Either affirmation interventions do not work in contexts, where the percentage of negatively stereotyped students is far from half of the student population; or there is a curvilinear relationship between the percentage of the population of negatively stereotyped students and the effect size. Were this second option the case, then our study contained true effect sizes that were too small to detect, due to the de facto segregation in the schools used. Future research should focus on whether the context moderates the effects in a stepwise or curvilinear fashion. We believe these are especially worthy avenues for future investigations into this promising intervention strategy.

Acknowledgments

The authors are indebted to Geoff Cohen for providing invaluable information about the original experiment and supplying the original materials for our experiment. The authors would like to thank Lindsay Juarez for her help administering the intervention. The authors would also like to thank Dick Nisbett, Valerie Purdie-Vaughns, Jelte Wicherts, and David Sherman for their insight and views on this experiment, the original intervention, and other replications. The authors would also like to thank Hal Pashler and Jonathan Schooler for their help and encouragement, and previous reviewers for their comments.

Authors' Note

Protzko and Aronson came up with the study design, Protzko collected the data and analyzed the results, and Protzko and Aronson wrote and revised the article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Allmendinger, J., & Hackman, J. R. (1995). The more, the better? A four-nation study of the inclusion of women in symphony orchestras. *Social Forces*, *74*, 423–460.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543–554.

Borman, G. D. (2012). *Examination of a self-affirmation intervention in St. Paul public schools*. The Senior Urban Education Research Fellowship Series Volume IX. Washington, DC: Council of Great City Schools.

Borman, G. D., Grigg, J., & Hanselman, P. *An effort to close achievement gaps at scale through self-affirmation*. Unpublished Manuscript.

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.

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

Eaton, S. (2008). *The children in Room E4: American education on trial*. Chapel Hill, NC: Algonquin Books.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365–371.

Koole, S. L., Smeets, K., van Knippenberg, A., & Dijksterhuis, A. (1999). The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology*, *77*, 111–125.

Lewin, K. (1931). The conflict between Aristotelian and Galileian modes of thought in contemporary psychology. *The Journal of General Psychology*, *5*, 141–177.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, *70*, 487.

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, *330*, 1234–1237.

Orfield, G., Kucsera, J., & Siegel-Hawley, G. (2012). *E pluribus ... separation: Deepening double segregation for more students*. Report for The Civil Rights Project.

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004a). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist*, *59*, 7–13.

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004b). On the value of correcting mischaracterizations of stereotype threat research. *American Psychologist*, *59*, 48–49.

Schmeichel, B. J., & Vohs, K. (2009). Self-affirmation and self-control: Affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology*, *96*, 770–782.

Sethi, R., & Somanathan, R. (2004). Inequality and segregation. *Journal of Political Economy*, *112*, 1296–1321.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology*, *38*, 183–242.

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.

Simonsohn, U. (2015). Small telescopes detectability and the evaluation of replication results. *Psychological Science*, *26*, 559–569.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (*Vol. 21*, pp. 261–302). New York, NY: Academic Press.

Steele, C. M. (2010). *Whistling Vivaldi and other clues to how stereotypes affect us*. New York, NY: W. W. Norton & Company.

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

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

Author Biographies

John Protzko is a postdoctoral research scholar at the University of California, Santa Barbara. He received his PhD in developmental psychology at New York University. His research sometimes focuses on experimental attempts to enhance cognitive ability.

Joshua Aronson is an associate professor of applied psychology at New York University. His research examines the social and psychological factors involved in human intellectual functioning, learning, and academic achievement.