

# Values Affirmation Intervention Reduces Achievement Gap between Underrepresented Minority and White Students in Introductory Biology Classes

Hannah Jordt,<sup>§</sup> Sarah L. Eddy,<sup>¶\*</sup> Riley Brazil,<sup>¶†</sup> Ignatius Lau,<sup>¶‡</sup> Chelsea Mann,<sup>¶@</sup> Sara E. Brownell,<sup>\*\*</sup> Katherine King,<sup>§</sup> and Scott Freeman<sup>§</sup>

<sup>§</sup>Department of Biology and <sup>¶</sup>School of Medicine, University of Washington, Seattle, WA 98195;

<sup>†</sup>Department of Biology, Florida International University, Miami, FL 33199; <sup>‡</sup>School of Medicine, Tufts University, Boston, MA 02111; <sup>@</sup>Stony Brook Medicine, Stony Brook, NY 11794; <sup>\*\*</sup>School of Life Sciences, Arizona State University, Tempe, AZ 85287

## ABSTRACT

Achievement gaps between underrepresented minority (URM) students and their white peers in college science, technology, engineering, and mathematics classrooms are persistent across many white-majority institutions of higher education. Attempts to reduce this phenomenon of underperformance through increasing classroom structure via active learning have been partially successful. In this study, we address the hypothesis that the achievement gap between white and URM students in an undergraduate biology course has a psychological and emotional component arising from stereotype threat. Specifically, we introduced a values affirmation exercise that counters stereotype threat by reinforcing a student's feelings of integrity and self-worth in three iterations of an intensive active-learning college biology course. On average, this exercise reduced the achievement gap between URM and white students who entered the course with the same incoming grade point average. This result suggests that achievement gaps resulting from the underperformance of URM students could be mitigated by providing students with a learning environment that removes psychological and emotional impediments of performance through short psychosocial interventions.

## INTRODUCTION

Representation of Blacks, Latino/as, Native Americans, and Hawaiian and Pacific Islanders remains low in the science, technology, engineering, and mathematics (STEM) workforce, despite their increasing proportion of the general population of the United States (National Science Foundation, National Center for Science and Engineering Statistics [NSF/NCES], 2015). This disparity persists regardless of data indicating that underrepresented minority (URM) students frequently exhibit an equal—if not higher—initial interest in majoring in STEM at the undergraduate level relative to their white peers (Anderson and Kim, 2006; Riegle-Crumb and King, 2010). Longitudinal data demonstrate that retention differences through college are partially to blame for this underrepresentation in STEM careers (Anderson and Kim, 2006). Thus, to diversify the STEM workforce, one critical step is to identify and mitigate the challenges faced by undergraduates from historically underrepresented groups.

For all students, and STEM undergraduates in particular, one of the best predictors of undergraduate retention across majors is performance in college courses (Riegle-Crumb and King, 2010; Beasley and Fischer, 2012; Westrick *et al.*, 2015). Therefore, observed disparities in performance in many STEM classrooms (Greene *et al.*, 2008; Eddy and Hogan, 2014; National Science Foundation, National Center for Science and Engineering Statistics, 2015) could help explain why URM students

Jeff Schinske, *Monitoring Editor*

Submitted December 27, 2016; Revised April 27, 2017; Accepted May 3, 2017

CBE Life Sci Educ September 1, 2017 16:ar41

DOI:10.1187/cbe.16-12-0351

<sup>†</sup>These authors contributed equally to this work and should be regarded as joint first authors.

<sup>‡</sup>These authors also contributed equally to this work.

\*Address correspondence to: Sarah L. Eddy (sarah.eddy@fiu.edu).

© 2017 H. Jordt *et al.* CBE—Life Sciences Education © 2017 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>). "ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

persist in STEM at lower rates. One strategy for increasing retention, and therefore workforce diversity, is to incorporate research-supported pedagogical strategies into the classroom that increase academic performance, such as active learning (Freeman *et al.*, 2014). When paired with out-of-class activities like preclass reading assignments and postclass review assignments, active learning reduces achievement gaps between Black and white students, first- and continuing-generation students (i.e., students with neither parent having earned a four-year college degree, or at least one parent having earned a four-year college degree, respectively), and students from lower and higher socioeconomic backgrounds (Haak *et al.*, 2011; Eddy and Hogan, 2014). These types of teaching innovations may help level the playing field by explicitly modeling strategies and skills needed to succeed in college STEM classrooms. This type of modeling could disproportionately benefit students whose prior educational experiences have not adequately prepared them for college-level work. However, even with these pedagogical interventions, achievement gaps still remain.

Classroom climate is another potential target—beyond changing curriculum—that instructors may need to consider to close achievement gaps. Classroom climate has been defined as the “intellectual, social, emotional, and physical environment in which ... students learn” (Ambrose *et al.*, 2010, p. 170). The impact of classroom climate on students has been documented across many studies. For example, feeling that the instructor cares about them reduces students’ apprehension in class and increases motivation, attitudes toward the course, and self-reported learning from the course (Ellis, 2004). In addition, believing that an instructor thinks they can improve increases the likelihood that students will incorporate instructor feedback (Cohen *et al.*, 1999) and helps students maintain their interest in the content area of the course (Good *et al.*, 2012). Thus, along with considering how learning is structured, considering the climate in a classroom may be critical for closing achievement and persistence gaps.

While there is some evidence that the use of active learning may improve classroom climate (Eddy and Hogan, 2014), the demonstrated changes have been small. In addition, changing classroom climate can be challenging, as it involves not only the instructor’s behaviors and attitudes but also those of classmates (Bright *et al.*, 1998; Holley and Steiner, 2005). Instead, it may be more efficient to bolster students against negative classroom climate or perceptions thereof to improve student performance. Fortunately, there is an easily implementable strategy termed “values affirmation” (Cohen *et al.*, 2006, 2009; Sherman *et al.*, 2013; Walton *et al.*, 2014) that is designed to bolster students who may be most likely to experience a negative classroom climate, that is, students who are often negatively stereotyped in academic settings.

A series of studies suggest students who feel at risk of upholding stereotypes or being judged based on stereotypes (termed “stereotype threat”) experience lower academic performance (Steele, 1988; Steele and Aronson, 1995; Nguyen and Ryan, 2008). Grappling with the concerns raised by stereotype threat (consciously or unconsciously) can decrease performance by decreasing working memory (Schmader and Johns, 2003) and can lead to hypervigilance (Forbes *et al.*, 2008), which may distract individuals from tasks. Stereotype threat can be especially debilitating to performance on difficult tasks

(Beilock *et al.*, 2007; Beilock, 2008; Neuville and Croizet, 2007), such as high-stakes examinations that require the entirety of a student’s mental faculties. It is not surprising, then, that stereotype threat has been shown to impact college-level course performance (Miyake *et al.*, 2010; Harackiewicz *et al.*, 2014), as high-stakes exams are often the primary contributor to course grades. In addition to short-term impacts on working memory, stereotype threat can have long-term impacts, such as students distancing themselves from a discipline with which they once identified (Fogliati and Bussey, 2013; Thoman *et al.*, 2013). This disassociation, coupled with lower performance, could contribute to a student’s decision to leave STEM.

Recent work has shown that a values-affirmation exercise can mitigate stereotype threat (Cohen *et al.*, 2006, 2009; Sherman *et al.*, 2013; Walton *et al.*, 2014). This simple exercise asks students to identify values that are important to them and write about how they incorporate these values into their lives. The positive emotions elicited as students consider how their own lives exemplify their own values reduce cortisol levels (Creswell *et al.*, 2005) and are thought to buffer students against the negative emotions caused by stereotype threat (Steele and Aronson, 1995; Cohen *et al.*, 2006). In one study among middle schoolers, administering the intervention not only increased the performance of African-American students in the term it was given but had sustained performance benefits 2 years later (Cohen *et al.*, 2006, 2009).

As the use of this intervention has spread beyond the initial researchers, however, the results have become more variable. For example, a more moderate impact was seen for a second group of African-American middle school students (Borman *et al.*, 2016), with a follow-up replication study showing no effect of the intervention (Hanselman *et al.*, 2016). In an undergraduate physics course, the intervention increased the performance of women (Miyake *et al.*, 2010), but this effect was not replicable during a following semester (Kost-Smith *et al.*, 2012). Similarly, researchers studying the impact of values affirmation on first-generation college students in introductory biology observed an effect of the intervention in one semester (Harackiewicz *et al.*, 2014) but not another (Harackiewicz *et al.*, 2016). The variability of these results suggests that more studies should be conducted in more environmental contexts before we generalize the utility of the values affirmation intervention.

Despite the variable results, there are multiple reasons why this intervention continues to be appealing to instructors. First, the intervention requires little to no (in our study) in-class instructional time and very little student effort. Each application can be administered in class (or in our case online) and takes 15 minutes at most. Therefore, if an instructor assigns it twice in a term, 30 minutes of student effort outside class could produce large effects on the performance of historically underrepresented groups. Second, the intervention is easily implemented and scalable. Students primarily write short essays in response to two questions, but the responses are not read by instructors. As a result, the intervention is feasible even in large classes.

Given the potential benefits and the ease of administering the intervention, we contribute to the growing body of research on the values affirmation intervention in specific STEM contexts by adding our novel context: one of the largest introductory biology classes in the nation (between 450 and 600 students

per section) at an R1 institution in the Pacific Northwest with predominantly white and Asian-American students. To account for potential variability of results, we deployed the intervention across 3 years and six sections of introductory biology. Specifically, we address the question: Can a values affirmation exercise increase the exam performance of students who identify as Black, Latina/o, Native American, Native Hawaiian, or Pacific Islander in introductory biology?

The setting of introductory biology is particularly important, because introductory classes often function as students' first exposure to their future profession and thus can have a larger-than-average influence on their decision to persist in STEM—making these courses an appropriate target for intervention (Cech *et al.*, 2011). In addition, relative to other STEM fields, biological science majors have the largest number of URM students (NSF/NCES, 2015). Thus, positively impacting the experience of URM students in this setting may have the largest impact on the STEM pipeline. Furthermore, URM students make up on average 12% of the students in these classes (white  $\cong$  43%, Asian  $\cong$  38%, international  $\cong$  7%, female  $\cong$  58%). Although the small number of URM students makes the statistical detection of differences difficult, this context, in which URM students are a numerical minority, is where stereotype threat may have the greatest negative effect on their performance (Thompson and Sekaquaptewa, 2002; Purdie-Vaughns *et al.*, 2008; Hanselman *et al.*, 2014). As a result, this group of students is a logical target for values affirmation. At an institutional level, we were also motivated to test the impact of values affirmation, because we had already been successful in reducing the achievement gap for historically underrepresented groups in this course due to the introduction of highly structured active-learning techniques, and we hypothesized that the remaining gap could be further reduced or eliminated by addressing psychosocial barriers to academic success.

## MATERIALS AND METHODS

### Course Context and Study Population

The study was conducted in three consecutive Fall terms of the first course of a three-quarter-long introductory biology series. A similar number of students were enrolled in the course in each term (~1100), and the class was taught as two back-to-back sections to accommodate enrollment numbers. All six sections (labeled A–F) in our study were taught by the same instructor, although in the final two Fall terms (sections C–F), a postdoctoral student worked in partnership with the main instructor, teaching 25% of the class sessions. Each term incorporated a significant amount of student-centered, active-learning techniques. These included the use of clickers, practice exams, nightly reading quizzes, and in-class group exercises. The students in these classes were predominantly sophomores intending to declare a science major; the course is required for students who intend to major in the life sciences. Information on student gender, URM status, and cumulative college grade point average (GPA) before entering the class were obtained from the office of the university registrar.

### The Intervention

Although variable results are often obtained with the values affirmation intervention, there are general suggestions in the literature for implementing the exercise in ways that would make

it more likely to work (Bradley *et al.*, 2015). We implemented our intervention with these aspects in mind, considering the content, introduction, timing, and repetition of the exercise.

For determining the effect of values affirmation on academic performance on six sections of introductory biology taught across 3 years, students in each section were randomly divided into control and treatment groups, with the instructor blind to the placement of students in each group. In the treatment group, students were given a list of 14 items they might consider valuable in their lives (independence, athletic ability, membership in a social group, etc.). After selecting two to three values that were most important to them, they wrote a brief response explaining why those values were important, summarized their top two reasons for choosing those values in writing, and answered four Likert-scale questions on the relevance of the chosen values to their lives. In the control group, students selected values from the same list, but instead indicated values that were least important to them, and answered questions on why the values would be important to someone else (for complete exercise, see the Supplemental Material). Students in both the treatment and control groups wrote positively about the values they selected, yet only in the treatment group did students evaluate these values in connection to themselves (Cohen *et al.*, 2006).

The exercise was completed online and outside class time and was designed to take ~15 minutes. No instructor-mediated introductory or follow-up discussions were provided apart from an initial email from the instructor alerting students to the assignment. As recommended (Bradley *et al.*, 2015), the intervention was framed as a standard class writing exercise, worth course points based on participation. Students completed the exercise during the first week of the quarter, as sustained benefits of values affirmation are thought to be dependent on early student success (Cohen *et al.*, 2009). Students were assigned the identical exercise again after receiving feedback from their second exam during the sixth week of the quarter, when we hypothesized that stress levels would be especially high for struggling students. The choice to implement the intervention twice in a term was based on prior research, as previous classroom studies that demonstrated positive effects of the values affirmation intervention also administered it twice (Miyake *et al.*, 2010; Harackiewicz *et al.*, 2014). The original creator of this exercise also recommended this approach as standard practice (G. Cohen, personal communication, February 22, 2010). We therefore considered a student to have fully completed the intervention only after he or she completed both rounds of the exercise.

### Study Population

Students enrolled in these classes were included in the study only if they completed all four course exams (the outcome variable), completed their assigned values affirmation exercise twice during the quarter, and consented in writing to the use of their data (University of Washington's Human Subjects Division, application #38240). A total of 2383 students satisfied these requirements (Table 1). Of this sample, 17.8% were first-year students and thus they did not have a measure of prior demonstrated college achievement (cumulative college GPA) to use as a covariate.

We did not include Asian-American students and international students in our analyses (38.7% and 5.9% of the overall sample, respectively). International students come from

**TABLE 1. Sample demographics, including the sample of students across the six sections who took all four exams, completed both dosages of the intervention, and have a measure of prior academic ability or do not have a measure of prior academic ability**

	Full sample	Sample with a measure of prior demonstrated college ability	Sample with no measure of prior demonstrated college ability
<i>N</i>	2383	1959	424
Gender			
Female	1463	1227	236
Male	920	732	188
Ethnicity/race/nationality			
Asian	922	776	146
Black	53	43	10
Hawaiian or Pacific Islander	31	29	2
Hispanic	144	122	22
International	141	88	53
White	985	837	148
Not reported	86	48	38
Median exam points earned (interquartile range)	277 (246–304)	278.5 (248–305)	270.5 (238–299)
Treatment group			
Control group	1174	970	200
Treatment group	1170	963	211

multiple cultural contexts yet are collapsed into a single category. Given how little information we had about their backgrounds, we did not feel we could make predictions about their experiences with stereotypes. Similarly, while Asian-American students are often considered overrepresented in STEM, the category comprises many groups with distinct ethnic backgrounds, some of which are underrepresented (Maramba, 2013). Disaggregated ethnic data for Asian-Americans were not available from our institution, and thus we were unable to separate this group into underrepresented and well-represented subgroups. Furthermore, combining any Asian-American students within a “majority students” category is potentially problematic as 1) Asian Americans may face “model minority” or other stereotypes in the classroom (Cheryan and Bodenhausen, 2000), and 2) prior work had established an achievement gap between Asian-American and white students in these classes (Eddy *et al.*, 2014). Thus, our analysis specifically focused on the intervention’s effects on URM students, who are historically the lowest-performing American students, and white students, who are historically the highest-performing American students in these classes, for a total of 1031 students across all three terms.

### Statistical Analyses

**Outcome Variable.** Students took four noncumulative exams worth 100 points each. In each year and section, these exams covered the same topics, although the individual questions differed. The exam questions in this course are open response, mostly short answer, and, on average, exam items are at the level of application and analysis rather than comprehension and recall (Freeman *et al.*, 2011). Across the years in our study, average exam performance was in the low 70s.

We chose to focus on total exam points rather than course grade, because stereotype threat is predicted to be induced in moments of high stress (Beilock *et al.*, 2007; Beilock, 2008; Neville and Croizet, 2007). Thus, high-stakes exams have a greater

potential for inducing stereotype threat relative to lower-stakes course assignments, and exam grades are more likely to be affected by the values affirmation exercise (Beilock, 2008). In addition, exams make up at minimum 55% of a student’s final grade in these classes, and in previous studies, exam grades have been shown to explain most of the variation in student course performance ( $R^2 = 0.89$  in one study; Freeman *et al.*, 2011).

**Covariates.** Normal variability in exam performance among students has the potential to mask small to moderate impacts of a treatment. Yet effects that might be considered small by statisticians (e.g., a 3% change in grade) may be educationally significant to students. To increase our chance of seeing even small effects of the values affirmation intervention, we included covariates in our analyses: course section, student gender, and cumulative college GPA at the start of the term, and several interaction terms.

Section, a categorical variable with six levels (each compared with section B, the reference level), was included to help account for differences in exams among the six sections and account for any among-year variation in the student population and course or any section-specific experiences that could impact performance, such as exam difficulty across sections. In addition, previous values affirmation studies saw variation in the efficacy of intervention in replication studies (Kost-Smith *et al.*, 2012; Hanselman *et al.*, 2016), so we included a three-way interaction term (treatment  $\times$  URM status  $\times$  section) to account for that potential variation.

Gender was included in the analysis, because historically these classes have shown an achievement gap between males and females (Eddy *et al.*, 2014). Although gender is not a binary, at the institution where this research occurred it is collected as a binary variable by the registrar, so in our analysis it is a categorical variable with two levels. In addition to a main effect of gender, the relationship of gender to exam performance

may vary by URM status, so we also included a gender  $\times$  URM interaction. Finally, gender has the potential to impact a student's experience of the treatment (cf. Miyake *et al.*, 2010), so we included a gender  $\times$  URM status  $\times$  treatment interaction to account for this potential variation.

Finally, cumulative college GPA at the beginning of the term a student was enrolled in introductory biology was included because it is highly predictive of academic performance in this course (Freeman *et al.*, 2011; Haak *et al.*, 2011; Eddy *et al.*, 2014). Although controlling for a measure of student ability is common in stereotype threat studies (Steele and Aronson, 1995; Miyake *et al.*, 2010; Harackiewicz *et al.*, 2014), controlling for cumulative college GPA introduces challenges regarding both interpretation and implementation. From a practical standpoint, use of this control required us to remove students in these classes who did not have a measure of cumulative GPA, reducing our sample by ~18%. In addition, some suggest that including any measure of student ability as a covariate may complicate the interpretation of the models (Yzerbyt *et al.*, 2004; Wicherts, 2005). One concern is that the student ability covariate may have a different relationship to the outcome variable for students who are and are not under stereotype threat; that is, in our case, stereotype threat may have impacted the cumulative GPA of URM and not white students (Wicherts, 2005). Thus, combining these two groups in one model could make interpretation challenging. To address this concern, we added an interaction between cumulative college GPA and URM status, which would reveal whether the relationship between cumulative college GPA varies by URM status. A second concern is that the measure of student ability may be correlated with URM status and thus could confound the results (i.e., one cannot know whether it is the effect of the treatment on lower cumulative GPA or URM status that drives the interaction; Yzerbyt *et al.*, 2004). To control for this concern, we included the interaction of cumulative GPA and treatment.

It should be noted that controlling for these variables has specific implications when interpreting the effects of the values affirmation intervention. For example, any change in performance of URM students in the treatment relative to white students is the average change for a URM student in the same section, of the same gender, and with the same entering GPA as the white student.

**Model Selection and Regressions.** We employed linear regression models to assess the relationship between the treatment students receive and their exam performance. We chose regressions because they allowed us to include covariates to account for other influences on exam performance in addition to the treatment (Theobald and Freeman, 2014).

The initial hypothetical regression model was as follows:

$$\text{Total exam points} \sim \text{cumulative college GPA} + \text{URM status} + \text{treatment} + \text{section} + \text{gender} + (\text{URM status} \times \text{treatment}) + (\text{section} \times \text{treatment}) + (\text{URM status} \times \text{section}) + (\text{cumulative college GPA} \times \text{treatment}) + (\text{URM status} \times \text{cumulative college GPA}) + (\text{URM status} \times \text{gender}) + (\text{gender} \times \text{treatment}) + (\text{gender} \times \text{treatment} \times \text{URM}) + (\text{URM status} \times \text{treatment} \times \text{section})$$

We used stepwise backward model selection to subtract individual terms from the model until we had a reduced model that

was best supported. We used Akaike information criterion (AIC) values, which estimate the quality of a given model relative to the other models, to determine the best-supported model. Specifically, AIC assesses the goodness of fit of a model to the data while simultaneously including a penalty for each additional term included in the model. The preferred model is the one with the lowest AIC corrected for small sample size value. If two models have an equivalent AIC value ( $\Delta\text{AIC} \leq 2$ ), then the model with the fewest terms is chosen. Any terms included in the preferred model are considered important for explaining the data, even if they do not pass the  $p = 0.05$  threshold (Burnham and Anderson, 2002).

Once the preferred model was identified, we ran a post hoc test on the URM status  $\times$  treatment term to determine whether the achievement gap between URM and white students in the treatment remained significant.

Finally, we ran a second set of analyses without cumulative GPA as a covariate using the full data set. We did not expect the analysis without cumulative GPA to be significant, given the large variation present in exam scores, but if the results were qualitatively similar to the analyses with GPA, they would lend support to the claim that challenges of including cumulative GPA did not meaningfully impair our results.

This resulted in a second hypothetical regression model:

$$\text{Total exam points} \sim \text{URM status} + \text{treatment} + \text{section} + \text{URM status} \times \text{treatment} + (\text{section} \times \text{treatment}) + (\text{URM status} \times \text{section}) + (\text{URM status} \times \text{treatment} \times \text{section}) + (\text{URM status} \times \text{gender}) + (\text{gender} \times \text{treatment}) + (\text{gender} \times \text{treatment} \times \text{URM})$$

We followed the same stepwise backward model-selection methods to find the best-supported model.

Analyses were implemented in R (R Core Team, 2016). Model selection via stepwise model selection with AIC was carried out through the package “stat” (R Core Team, 2016). Post hoc analyses were implemented in R using the package “phia” (de Rosario-Martinez, 2015).

## RESULTS

### Analyses with Cumulative GPA as a Covariate

In our analysis with cumulative college GPA as a covariate, model selection (Table 2) indicated the preferred model to explain exam performance was as follows:

$$\text{Cumulative exam performance} = \beta_0 + \beta_1(\text{cumulative college GPA}) + \beta_2(\text{URM}) + \beta_3(\text{treatment}) + \beta_4(\text{section}) + \beta_5(\text{URM} \times \text{treatment}) + \beta_6(\text{gender}) + \beta_7(\text{gender} \times \text{treatment})$$

These model-selection results suggest several things about our data. First, we did not see evidence that the relationship of cumulative GPA to total exam points varies by URM status ( $-\text{URM status} \times \text{cumulative GPA}$ ; Table 2), which was one of our primary concerns about including cumulative college GPA as a covariate. Second, the cumulative GPA  $\times$  treatment interaction was not selected for inclusion in the final model ( $-\text{treatment} \times \text{cumulative college GPA}$ ; Table 2), implying the impact of treatment was not driven by students with lower cumulative GPAs regardless of URM status. Third, we did not see evidence that the impact of the intervention varied largely from section to section ( $-\text{URM status} \times \text{treatment} \times \text{section}$ ; Table 2), as other research teams have observed upon replicating studies at their own institutions

**TABLE 2. Model-selection table for the analyses with cumulative college GPA identifying the preferred model<sup>a</sup>**

Initial model and terms dropped <i>Outcome: total exam points earned</i>	Analyses with cumulative college GPA				
	<i>df</i>	<i>Deviance</i>	<i>Residual df</i>	<i>Residual deviance</i>	<i>AIC</i>
<i>Initial model: cumulative college GPA + URM status + treatment + section + gender + (URM status × cumulative college GPA) + (URM status × section) + (URM status × treatment) + (URM status × gender) + (treatment × cumulative college GPA) + (treatment × section) + (gender × treatment) + (section × treatment × URM status) + (gender × treatment × URM status)</i>			1006	915,723.0	7104.2
– (Section × treatment × URM status)	5	3757.7	1011	919,480.7	7098.4
– (URM status × section)	5	734.9	1016	920,215.6	7089.2
– (Treatment × section)	5	3085.31	1021	923,300.9	7082.7
– (Treatment × URM status × gender)	1	650.79	1022	923,951.7	7081.4
– (URM status × gender)	1	356.0	1023	924,307.7	7079.8
– (URM status × cumulative college GPA × treatment)	1	1136.5	1024	925,444.2	7079.1
– (URM status × cumulative college GPA)	1	668.0	1025	926,112.2	7077.9
– (Treatment × cumulative college GPA)	1	656.6	1026	926,768.8	7076.6
<i>Final model: cumulative college GPA + URM status + treatment + section + gender + (URM status × treatment) + (gender × treatment)</i>					

<sup>a</sup>For each comparison, the term subtracted from the model is listed in the first column. As this is a cumulative table, any terms above the current row were already removed from the model before the current row was tested. Terms were removed if the AIC of the reduced model was 2 or less than the AIC value of the fuller model or if the models had equivalent AIC values ( $\Delta\text{AIC} < 2$ ). If removing the term increased the AIC by more than 2, the term was retained in the model.

(Kost-Smith et al., 2012; Harackiewicz et al., 2016). Finally, we did not see that the relationship between gender and treatment varied by race (gender × treatment × URM; Table 2), indicating that the impacts of gender on how students experienced the treatment were consistent for URM and white students.

This final model also suggests that controlling for cumulative college GPA and section increases the fit of the model. Finally, the final model suggests that the values affirmation intervention impacted the achievement gap between URM and white students and that student gender mediated this effect. The  $R^2$  of the final preferred model was 0.445.

In the control condition (and conditioned on the final model's covariates), white males were predicted to perform the highest. White women performed only marginally and not significantly lower (0.6% fewer exam points than white males;  $\beta_{\text{gender}} = -2.5 \pm 2.76$ ,  $p = 0.363$ ). URM males earned 4% fewer of the possible exam points ( $\beta_{\text{race}} = -16.01 \pm 3.58$ ,  $p < 0.00001$ ; Table 3) than white males. URM women performed only marginally lower than URM males (earning 4.6% fewer of the possible exam points than white males).

In the treatment group, we observed a reduced achievement gap between white and URM students in the same section and with the same entering cumulative GPA (see Figure 1). Although white males still received a boost from the treatment ( $\beta_{\text{treatment}} = 6.45 \pm 3.29$ ,  $p = 0.05$ ), URM students received a disproportionate boost ( $\beta_{\text{treatment} \times \text{race}} = 10.29 \pm 4.75$ ,  $p = 0.031$ ; Table 3). Specifically, male URM students in the treatment earned an additional 4.2% of the possible exam points relative to male URM students in the control, while white male students in the treatment earned only an additional 1.6% of exam points. URM women in the treatment received a more moderate boost than male URM students, as there was a significant interaction between gender and treatment (2.2%;  $\beta_{\text{treatment} \times \text{gender}} = -8.10 \pm 3.90$ ,  $p = 0.038$ ; Table 3). This was still a larger boost than white males were predicted to experience and thus slightly reduced the achievement gap between these two groups. There was roughly no difference in performance of white women in the control and treatment groups.

Post hoc analysis of the interaction demonstrated that there was still a small but significant achievement gap between white and URM students with equivalent prior GPAs in the values affirmation treatment (mean difference = 10.3,  $F(1, 1026) = 4.7$ ,  $p = 0.03$ ). Thus, averaged across all three terms, the values affirmation reduced but did not eliminate the exam achievement gap between white and URM students with the same college GPA in these introductory biology classrooms.

#### Analysis without Cumulative College GPA as a Covariate

In our analysis without cumulative college GPA as a covariate, model selection (Table 4) indicated that the best model to explain exam performance was as follows:

$$\text{Cumulative exam performance} = \beta_0 + \beta_1(\text{URM}) + \beta_2(\text{treatment}) + \beta_3(\text{section}) + \beta_4(\text{gender}) + \beta_5(\text{URM} \times \text{treatment}).$$

The preferred model for the analysis without cumulative college GPA is qualitatively similar in many ways to the model with GPA. Including section as a control still increases the fit of the model, as does including a URM × treatment interaction, implying that the treatment impacts the performance of URM and white students differently. The major differences are that the gender × treatment interaction does not increase the fit of the model to the data and that the total variance in exam scores explained by the model is much lower ( $R^2 = 0.174$ ).

As we expected, without a control for student ability, both the treatment ( $\beta = 3.7 \pm 2.370$ ,  $p = 0.117$ ) and treatment × URM status ( $\beta = 8.3 \pm 5.38$ ,  $p = 0.121$ ) were not significant. However, the treatment × URM status regression coefficient is qualitatively similar to the regression coefficient in the model with cumulative college GPA. This again implies that including this control did not substantially change the pattern of our results; it just accounted for more variance and allowed us to discern more patterns specifically driven by the values affirmation treatment. In addition, the fact that model-selection procedure retained treatment and treatment × URM status in the preferred model implies that they remain important for

**TABLE 3.** Regression coefficients for preferred models for analysis with cumulative college GPA as a covariate and without cumulative college GPA as a covariate<sup>a</sup>

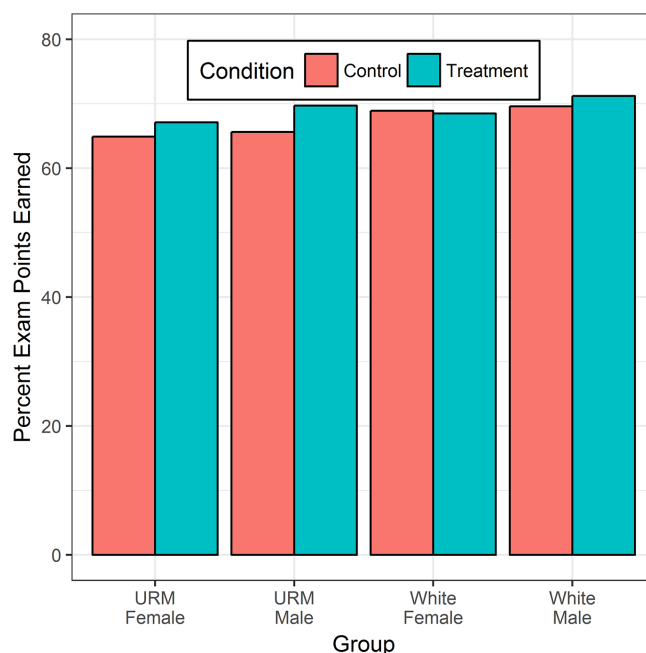
	Preferred model for students with cumulative college GPA covariate	Preferred model without cumulative college GPA covariate
	$\beta \pm SE$ ( <i>p</i> value)	$\beta \pm SE$ ( <i>p</i> value)
<b>Coefficients</b>		
Intercept	278.3 $\pm$ 3.48 ( $<0.001$ )	275.6 $\pm$ 3.51 ( $<0.001$ )
Cumulative college GPA at start of course	52.81 $\pm$ 2.42 ( $<0.001$ )	NA
Racial group (ref: white)		
URM	-16.01 $\pm$ 3.58 ( $<0.001$ )	-25.6 $\pm$ 4.07 ( $<0.001$ )
Gender (ref: male)		
Female	-2.52 $\pm$ 2.76 (0.363)	-5.4 $\pm$ 2.19 (0.0133)
Treatment group (ref: control)		
Treatment	6.45 $\pm$ 3.29 (0.0501)	3.7 $\pm$ 2.37 (0.123)
Gender $\times$ treatment group (ref: male $\times$ control)		
Female $\times$ treatment	-8.10 $\pm$ 3.90 (0.038)	NA
Race $\times$ treatment group (ref: white $\times$ control)		
URM $\times$ treatment	10.29 $\pm$ 4.75 (0.031)	8.2 $\pm$ 5.37 (0.126)
Section (ref: section B)		
Section A	30.8 $\pm$ 3.60 ( $<0.001$ )	37.8 $\pm$ 3.98 ( $<0.001$ )
Section C	-1.7 $\pm$ 3.62 (0.639)	6.3 $\pm$ 4.00 (0.116)
Section D	-9.4 $\pm$ 3.83 (0.014)	-4.5 $\pm$ 4.09 (0.273)
Section E	-0.1 $\pm$ 3.52 (0.989)	6.9 $\pm$ 3.79 (0.071)
Section F	-8.5 $\pm$ 3.58 (0.017)	-1.8 $\pm$ 3.87 (0.633)
<i>R</i> <sup>2</sup>	0.445	0.174

<sup>a</sup>The outcome variable is exam performance. For categorical variables, the reference level (ref.) is in parentheses, indicating the binary comparison that was made (e.g., section B compared with A and section B compared with section C).

explaining the outcome variable even in this model without cumulative GPA.

## DISCUSSION

The use of a values affirmation intervention led to a reduction, but not elimination, of the achievement gap between URM and white students with equivalent college GPAs in three terms of introductory biology. URM males saw the strongest



**FIGURE 1.** Predicted student exam scores for different student groups assuming all students had the average cumulative college GPA and were in the reference section. Based on preferred model including cumulative GPA.

effect of the treatment, URM women the second strongest, and white men the third strongest. White females were not affected by the intervention.

Previous work has shown that the magnitude of achievement gaps can change depending on course structure and classroom climate. Different instructors and/or instructional strategies can impact achievement gaps between white students and historically underrepresented groups in college STEM courses (Kreutzer and Boudreaux, 2012; Eddy and Hogan, 2014). Before our study, several active-learning strategies were introduced into the introductory biology courses at this university in an attempt to decrease failure rates (Freeman *et al.*, 2011). While this was accomplished successfully, it had the added benefit of decreasing the achievement gap between educationally and socioeconomically disadvantaged and nondisadvantaged students (Haak *et al.*, 2011). In the current study, we incorporated values affirmation in this same intensely active-learning environment with the hope of reducing achievement gaps still further. On the basis of our belief that gaps in classroom achievement derived from inequitable instructional practices were already mostly reduced due to the intense use of active learning, we hypothesized that the remaining achievement gap between white and URM students was largely due to psychosocial threats in the classroom. Specifically, we focused on the potential emotional or psychosocial threat of being stereotyped. We found that values affirmation benefited URM students and white males, yet disproportionately increased the exam performance of URM students, resulting in yet another reduction of the URM–white achievement gap. Taken together, the results reported earlier in Haak *et al.* (2011) and in the current study suggest that one possible recipe for minimizing achievement gaps between URM and white students in undergraduate biology courses may be 1) presenting content in

**TABLE 4. Model-selection table for the analyses without cumulative college GPA identifying the preferred model<sup>a</sup>**

Initial model and terms dropped <i>Outcome: total exam points earned</i>	Analyses without cumulative college GPA				
	<i>df</i>	<i>Deviance</i>	<i>Residual df</i>	<i>Residual deviance</i>	<i>AIC</i>
<i>Initial model: URM status + treatment + section + gender + (URM status × section) + (treatment × section) + (URM status × treatment) + (gender × treatment) + (URM status × gender) + (section × treatment × URM status) + (gender × treatment × URM status)</i>			1192	1,623,611	8832.1
– (Section × treatment × URM status)	5	10,616.7	1197	1,634,228	8830.1
– (Treatment × section)	5	1469.8	1202	1,635,698	8821.2
– (URM status × section)	5	5558.5	1207	1,641,256	8815.3
– (Treatment × URM status × gender)	1	751.4	1208	1,642,008	8813.9
– (URM status × gender)	1	1389.2	1209	1,643,397	8813.0
– (Treatment × gender)	1	2386.7	1210	1,645,784	8812.7
<i>Preferred model: URM status + treatment + section + gender + (URM status × treatment)</i>					

<sup>a</sup>For each comparison, the term subtracted from the model is listed in the first column. As this is a cumulative table, any terms above the current row were already been removed from the model before the current row was tested. Terms were removed if the AIC of the reduced model was 2 or less than the AIC value of the fuller model or if the models had equivalent AIC values ( $\Delta\text{AIC} < 2$ ). If removing the term increased the AIC by more than 2, the term was retained in the model.

a way that benefits everyone, but also disproportionately benefits underrepresented groups (such as the use of active learning); and 2) employing values affirmation or other techniques to bolster students against a negative classroom climate.

The magnitude of the effect of the values affirmation exercise for the average male URM student was a 4.2% increase in exam performance. This is almost half of an SD in raw exam points earned by students in these classes. Female URM students saw more moderate gains from the intervention: a 2.2% increase in exam scores relative to female URM students in the control condition. Learning is a complex task, and thus any intervention intended to impact student performance tends to have moderate to small effect sizes. For example, across more than 200 studies of undergraduate STEM courses, changing from a traditional lecture to an active-learning classroom increased exam scores on average by 6% for all students, or about half an SD (Freeman *et al.*, 2014). Our intervention, which required roughly only 30 minutes of student effort and very little instructor input over the course of an academic term, increased the performance of male URM students half to nearly as much as converting an entire course to active learning. The ease of distributing and completing the exercise makes this intervention a promising tool in addressing the URM achievement gap in undergraduate STEM classrooms.

Prior work with values affirmation in STEM settings other than biology indicates that it can reduce the achievement gap between men and women by raising female achievement. However, in our study, we found the opposite for white students: the intervention increased the achievement of white men but not white women. Stereotype threat can stem from many different sources (Shapiro, 2011) and can be experienced by members of any group who feel there are comparatively negative traits associated with their group. For example, researchers induced stereotype threat in white male undergraduates completing a difficult math test by telling them that their performance would be compared with Asian students (Aronson *et al.*, 1999). Because our classrooms are on average 58% female and 38% Asian, it is possible that white males in biology experience stereotype threat in relation to one or both groups. However, evidence from prior studies in this setting does not support this

hypothesis, as white males do not behave or perform as predicted for groups under stereotype threat. For example, male students in these classrooms participate at higher rates in class and report greater comfort with this participation than females; white males, in particular, outperform all other groups on in-class exams; and peers in the class perceive males to be more knowledgeable about biology than females (Eddy *et al.*, 2014, 2015; Grunspan *et al.*, 2016). These results suggest that stereotype threat is an unlikely explanation for our observed results. Instead, it may be that the values affirmation intervention, which does not specifically reference stereotype threat in any way in the writing prompt, has some additional value for students beyond stereotype threat reduction. White males may be benefiting from this alternative value. Regardless, URM students in our study disproportionately benefit from values affirmation, leading to a narrowing of the achievement gap between URM and white students.

The only other study to test the effects of values affirmation in a college biology classroom saw an impact on first-generation college students. However, our university registrar did not collect this demographic information until the final year of our study, and thus we did not explore the effects of our exercise on this group. Future studies exploring this dynamic would be informative.

Although values affirmation interventions are associated with stereotype threat reduction, the target of the intervention is not specifically stereotype threat. Thus, one could argue the impact of the values affirmation on exam scores was due to the alleviation of some other psychological process impacting student performance. We are unable to rule out this possibility, because we did not measure the degree to which individuals either 1) felt they experienced stereotype threat or 2) endorsed views aligned with common academic stereotypes. Furthermore, these measurements might have allowed us to better understand why certain students and not others were impacted by the intervention. However, obtaining a measure of stereotype threat is not common practice in classroom studies that have seen an impact of values affirmation (Cohen *et al.*, 2006, 2009; Harackiewicz *et al.*, 2014; but cf. Miyake *et al.*, 2010), and we were therefore wary of introducing such a component into our experimental design. Above all, we wanted to avoid



signaling to students that they were taking part in a study designed to reduce stereotype threat, as being aware of the intention of values affirmation has been shown to reduce its effectiveness (Sherman *et al.*, 2009).

Although we show an impact of values affirmation on URM students' achievement in biology and for white males, these results were obtained in one particular context. Prior studies suggest the effects of values affirmation interventions are sensitive to the environment in which they are used (Kost-Smith *et al.*, 2012; Cohen and Garcia, 2014; Hanselman *et al.*, 2014) or may be affected by the size of the achievement gap they attempt to address (Hanselman *et al.*, 2014). The immediate social climate a student experiences can vary widely across institutions, classrooms, and years, and interventions that help students cope in these climates may be variably useful. Thus, this study is more a demonstration of potential benefit than a guarantee that biology instructors will see similar impacts of values affirmation in their classrooms.

In addition, college achievement is greatly impacted by past academic preparation, which is highly variable among college students. We show that the gap between URM and white students is reduced only after controlling for this variation and that failing to do so swamps out any signal regarding the psychological benefit of values affirmation. The intervention cannot change a student's preparation, but it can support an environment wherein students' performance aligns more with their abilities and ensures that they are less likely to underperform at a disproportionate rate to their white peers with equal incoming GPAs.

## CONCLUSION

A URM student's decision to remain in STEM is impacted by his or her achievement and sense of belonging in the discipline (Hausmann *et al.*, 2007; Chemers *et al.*, 2011). By diminishing psychological threats in an active-learning classroom, we may be able to reduce barriers to achievement and empower a student's sense of self-value to encourage retention in STEM. Our study suggests that, at least in some cases, these benefits can be achieved with minimal effort by students and instructors through the use of a short, evidence-based psychosocial intervention.

## ACKNOWLEDGMENTS

We are grateful to Geoffrey Cohen for sharing intervention materials and generous advice throughout this project, John Parks for logistical support, and David Haak for sharing code used in our analyses. We also thank the University of Washington Biology Education Group for support and advice over the course of this project. A special thanks to Elli Theobald, Jake Cooper, Alison Crowe, Chris Runyon, Melissa Akins, Lisa Corwin, Erin Dolan, Yoi Tibbetts, and Jelte Wicherts for detailed comments that improved the paper.

## REFERENCES

- Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: 7 research-based principles for smart teaching*. San Francisco: Jossey Bass.
- Anderson, E., & Kim, D. (2006). *Increasing the success of minority students in science and technology*. Washington, DC: American Council on Education.

- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29–46.
- Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education, 15*(4), 427–448.
- Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science, 17*(5), 339–343.
- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology General, 136*(2), 256–276.
- 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.
- Bradley, D., Crawford, E., & Dahill-Brown, S. E. (2015). Fidelity of implementation in a large-scale, randomized, field trial: Identifying the critical components of values affirmation. *Society for Research on Educational Effectiveness; Spring (Conference Abstract Template)*.
- Bright, C. M., Duefield, C. A., & Stone, V. E. (1998). Perceived barriers and biases in the medical education experience by gender and race. *Journal of the National Medical Association, 90*(11), 681–88.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical 628 information-theoretical approach*. New York: Springer.
- Cech, E., Rubineau, B., Silbey, S., & Seron, C. (2011). Professional role confidence and gendered persistence in engineering. *American Sociological Review, 76*(5), 641–666.
- Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Australian Journal of Social Issues, 67*(3), 469–491.
- Cheryan, S., & Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: The psychological hazards of "model minority" status. *Psychological Science, 11*(5), 399–402.
- Cohen, G. L., & Garcia, J. (2014). Educational theory, practice, and policy and the wisdom of social psychology. *Policy Insights from the Behavioral and Brain Sciences, 1*(1), 13–20.
- Cohen, G. L., Garcia, J., Apfel, N., & Maste, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*(5791), 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*(5925), 400–403.
- Cohen, G. L., Steele, C., & Ross, L. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin, 25*(10), 1302–1318.
- 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*(11), 846–851.
- de Rosario-Martinez, H. (2015). *Phia: Post-hoc interaction analysis. R package version 0.2–1 [software]*, Available from <http://CRAN.R-project.org/package=phia> (retrieved 5 January 2015)
- Eddy, S. L., Brownell, S. E., Thummaphan, P., Lan, M.-C., & Wenderoth, M. P. (2015). Caution, student experience may vary: Social identities impact a student's experience in peer discussions. *CBE—Life Sciences Education, 14*(4), ar45.
- Eddy, S. L., Brownell, S. E., & Wenderoth, M. P. (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. *CBE—Life Sciences Education, 13*(3), 478–492.
- Eddy, S. L., & Hogan, K. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE—Life Sciences Education, 13*(3), 453–68.
- Ellis, K. (2004). The impact of perceived teacher confirmation on receiver apprehension, motivation, and learning. *Common Education, 539*(1), 1–20.
- Fogliati, V. J., & Bussey, K. (2013). Stereotype threat reduces motivation to improve: Effects of stereotype threat feedback on women's intentions to improve mathematical ability. *Psychology of Women Quarterly, 37*(3), 310–324.
- Forbes, C. E., Schmdar, T., & Allen, J. J. B. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of

- minorities under threat. *Social Cognitive and Affective Neuroscience*, 3(3), 253–261.
- Freeman, S., Eddy, S., McDonough, M., Smith, M. K., Okoroafo, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the USA*, 111(23), 8410–8415.
- Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE—Life Science Education*, 10(2), 175–186.
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, 102(4), 700–717.
- Greene, T. G., Marti, C. N., & McClennney, K. (2008). The effort-outcome gap: Differences for African American and Hispanic community college students in student engagement and academic achievement. *Journal of Higher Education*, 79(5), 513–539.
- Grunspan, D. Z., Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, A. J., & Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PLoS One*, 11(2), e0148405.
- Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, 332(6034), 1213–1216.
- Hanselman, P., Bruch, S. K., Gamoran, A., & Borman, G. D. (2014). Threat in context: School moderation of the impact of social identity threat on racial/ethnic achievement gaps. *Sociology of Education*, 87(2), 106–124.
- Hanselman, P., Rozek, C. S., Grigg, J., & Borman, G. D. (2016). New evidence on self-affirmation effects and theorized sources of heterogeneity from large-scale replications. *Journal of Educational Psychology*, 109(3), 405–424.
- Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Giffen, C. J., Blair, S. S., Rouse, D. I., & Hyde, J. S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology*, 106(2), 375–389.
- Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology*, 111(5), 745–765.
- Hausmann, L. R. M., Schofield, J. W., & Woods, R. L. (2007). Sense of belonging as a predictor of intentions to persist among African American and White first-year college students. *Research in Higher Education*, 48(7), 803–839.
- Holley, L. C., & Steiner, S. (2005). Safe space: Student perspectives on classroom environment. *Journal of Social Work Education*, 41(1), 49–64.
- Kost-Smith, L. E., Pollock, S. J., Finkelstein, N. D., Cohen, G. L., Ito, T. A., & Miyake, A. (2012). Replicating a self-affirmation intervention to address gender differences: Successes and challenges. In *2011 Physics education research conference* (Vol. 1413, pp. 231–234).
- Kreutzer, K., & Boudreaux, A. (2012). Preliminary investigation of instructor effects on gender gap in introductory physics. *Physical Review ST Physics Education Research*, 8, 010120.
- Maramba, D. C. (2013). Creating successful pathways for Asian Americans and Pacific Islander community college students (AAPIs) in STEM. In Palmer, R. T., & Wook, J. L. (Eds.), *Community colleges and STEM: Examining underrepresented racial and ethnic minorities* (pp. 156–171). New York: Routledge.
- 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(6008), 1234–1237.
- National Science Foundation, National Center for Science and Engineering Statistics. (2015). *Women, Minorities, and Persons with Disabilities in Science and Engineering (Special Report)*, Arlington, VA: National Science Foundation.
- Neuville, E., & Croizet, J. (2007). Can salience of gender identity impair math performance among 7–8 year old girls? The moderating role of task difficulty. *European Journal of Psychology of Education*, 22(3), 307–316.
- Nguyen, H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314–1334.
- Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Dittmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology*, 94(4), 615–630.
- R Core Team. (2016). *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing. Available from www.R-project.org (accessed 1 September 2012).
- Riegle-Crumb, C., & King, B. (2010). Questioning a White male advantage in STEM: Examining disparities in college major by gender and race/ethnicity. *Education Research*, 39(9), 656–664.
- Schmader, R., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440–452.
- Shapiro, J. (2011). Types of threats: From stereotype threat to stereotype threats. In Inzlicht, M., & Schmader, T. (Eds.), *Stereotype threat: Theory, process, and application* (pp. 71–88). New York: Oxford University Press.
- Sherman, D. K., Cohen, G. L., Nelson, L. D., Nussbaum, A. D., Bunyan, D. P., & Garcia, J. (2009). Affirmed yet unaware: Exploring the role of awareness in the process of self-affirmation. *Journal of Personality and Social Psychology*, 97(5), 745–764.
- Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., ... Cohen, G. L. (2013). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology*, 104(4), 591–618.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In Berkowitz, L. (Ed.), *Advances in experimental social psychology* (pp. 261–302). New York: Academic.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
- Theobald, R., & Freeman, S. (2014). Is it the intervention or the students? Using linear regression to control for student characteristics in undergraduate STEM education research. *CBE—Life Sciences Education*, 13(1), 41–48.
- Thoman, D. B., Smith, J. L., Brown, E. R., Chase, J., & Lee, J. K. (2013). Beyond performance: A motivational experiences model of stereotype threat. *Educational Psychology Review*, 25(2), 211–243.
- Thompson, M., & Sekaquaptewa, D. (2002). When being different is detrimental: Solo status and the performance of women and racial minorities. *Analyses of Social Issues and Public Policy*, 2(1), 183–203.
- Walton, G., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2014). Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, 107(2), 468–485.
- Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT® scores, high school grades, and SES. *Educational Assessment*, 20(1), 23–45.
- Wicherts, J. M. (2005). Stereotype threat research and the assumptions underlying analysis of covariance. *American Psychologist*, 60(3), 267–269.
- Zyerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, 40(3), 424–431.