Published in final edited form as:
J Educ Psychol. 2014 May 1; 106(2): 375–389. doi:10.1037/a0034679.

Closing the Social Class Achievement Gap for First-Generation Students in Undergraduate Biology

Judith M. Harackiewicz,
Department of Psychology, University of Wisconsin – Madison

Elizabeth A. Canning,
Department of Psychology, University of Wisconsin – Madison

Yoi Tibbetts,
Department of Psychology, University of Wisconsin – Madison

Cynthia J. Giffen,
Department of Zoology, University of Wisconsin – Madison

Seth S. Blair,
Department of Zoology, University of Wisconsin – Madison

Douglas I. Rouse, and
Department of Plant Pathology, University of Wisconsin – Madison

Janet S. Hyde
Department of Psychology, University of Wisconsin – Madison

Abstract

Many students start college intending to pursue a career in the biosciences, but too many abandon this goal because they struggle in introductory biology. Interventions have been developed to close achievement gaps for underrepresented minority students and women, but no prior research has attempted to close the gap for first-generation students, a population that accounts for nearly a fifth of college students. We report a values affirmation intervention conducted with 798 U.S. students (154 first-generation) in an introductory biology course for majors. For first-generation students, values affirmation significantly improved final course grades and retention in the second course in the biology sequence, as well as overall GPA for the semester. This brief intervention narrowed the achievement gap between first-generation and continuing generation students for course grades by 50% and increased retention in a critical gateway course by 20%. Our results suggest that educators can expand the pipeline for first-generation students to continue studying in the biosciences with psychological interventions.

Many students start college intending to pursue a career in the biomedical sciences, but too many abandon this goal because they struggle in introductory biology courses. Underrepresented minority (URM) students are particularly likely to struggle in

Correspondence concerning this article should be addressed to Judith Harackiewicz, Department of Psychology, University of Wisconsin – Madison, 1202 W Johnson St, Madison, WI 53706. jmharack@wisc.edu.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

mathematics and science courses, and there have been many attempts to address these achievement gaps (Aronson & Dee, 2012; Haak, HilleRisLambers, Pitre, & Freeman, 2011). Gender gaps also occur in mathematics and some sciences, especially physics, and interventions have addressed these gaps as well (Miyake et al., 2010). Missing from these achievement-gap research efforts, however, is attention to another at-risk group: first-generation college students. First-generation (FG) college students are those for whom neither parent received a 4-year college degree, and they comprise roughly 15–20% of students in American universities (Bowen, Kurzweil, & Tobin, 2005; Saenz, Hurtado, Barrera, Wolf, & Yeung, 2007). These students tend to perform more poorly and have higher dropout rates than continuing-generation (CG) students -- those with at least one parent with a 4-year degree (Sirin, 2005). This performance discrepancy has been referred to as the social-class achievement gap, because parental education is considered to be a proxy for social class or socio-economic status (SES) (Pascarella & Terenzini, 1991; Jackman & Jackman, 1983; Snibbe & Markus, 2005). In other words, FG students are more likely to come from working class backgrounds as compared to the middle- and upper-class backgrounds of CG students, and they may face significant economic and social barriers in college.

A number of economic and social factors contribute to the social class achievement gap in college performance, including poverty (Reardon, 2011), quality of high school (Terenzini, Springer, Yaeger, Pascarelli, & Nora, 1996), rigor of high school preparation (Warburton, Bugarin, & Nunez, 2001), and parenting practices (Guryan, Hearst, & Kearney, 2008; Horvat, Weininger, & Lareau, 2003; Lareau, 2003; Ramey & Ramey, 2010). However, the achievement gap may also reflect psychological factors, to the extent that FG students experience the college environment as threatening, due to stereotypes about their group, or a mismatch of cultural values (Croizet & Claire, 1998; Johnson, Richeson, & Finkel, 2011; Stephens, Fryberg, Markus, Johnson, & Covarrubias, 2012). Here we report on a social-psychological intervention designed to address the social-class achievement gap and promote retention in an introductory biology sequence for FG students.

Theoretical Framework

The theoretical framework for this research involves a novel integration of the stereotype threat model with cultural mismatch theory. The values affirmation (VA) intervention pioneered by Cohen and colleagues (Cohen, Garcia, Apfel, & Master, 2006) was designed to close achievement gaps by buffering students against the possibility of confirming stereotypes about their group, known as “stereotype threat” (Steele, 1997). Steele argued that individuals experience apprehension when confronted with personally relevant stereotypes that threaten their social identity or self-esteem, and that this apprehension impairs performance on challenging academic tasks. Numerous laboratory experiments have shown that minority group members (or women in math and science contexts) perform more poorly when told that a test is diagnostic of ability, or when stereotypes about their group are made salient, relative to non-evaluative, non-diagnostic, controls (Aronson & Inzlicht, 2004; Aronson et al., 1999; Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002). These results have been replicated in more than 300 laboratory and field studies, ranging from studies of minority students in middle school to white athletes in college, women in undergraduate physics classes, and elderly participants performing cognitive tasks (see Walton & Spencer, 2009, for meta-analytic review). A few studies have examined stereotype threat and social class, and the results suggest that low SES college students demonstrate stereotype threat effects, performing more poorly when tested in evaluative contexts that make SES salient (Croizet & Claire, 1998; Croizet & Dutrevis, 2004; Harrison, Stevens, Monty. & Coakley, 2006; Spencer & Castano, 2007). These results suggest that FG students may be vulnerable to the debilitating effects of stereotype threat.

To combat threats to the self, Steele and Liu (1983) developed a technique to promote self-integrity and self-worth via a writing intervention called self-affirmation or values affirmation (VA), and Steele and Aronson (1995) were the first to apply this technique to the problem of stereotype threat. The VA intervention involves students writing about their most important values, which can help them cope with identity threat (Fein & Spencer, 1997; Sherman, Nelson, & Steele, 2000). When individuals affirm their core personal values in a threatening environment, they can reestablish a perception of personal integrity and worth, which bolsters them against challenges and reduces stress (see McQueen & Klein, 2006; Sherman & Cohen, 2006 for review). For example, Creswell et al (2005) found that a VA intervention reduced physiological measures of stress for participants giving a presentation to a judgmental audience, and Sherman, Bunyan, Creswell, and Jaremka (2009) found that a VA intervention reduced stress for students preparing for important exams.

In one of the earliest laboratory studies of gap closing with the VA intervention, Martens, Johns, Greenberg, & Schimel (2006) first documented that women showed stereotype threat effects on a challenging math test, performing more poorly than males when the test was presented as highly diagnostic of math ability. Under these same diagnostic conditions, however, women who had completed a VA intervention prior to the math test performed significantly better, relative to both males and women in the control condition. In sum, values affirmation alleviates social identity threat by making alternative sources of self-integrity salient, thereby relieving evaluative pressure and stress, allowing students to perform better on challenging tasks.

Based on these laboratory studies, researchers have developed VA interventions that have been tested in randomized field studies in middle-school and college classes. The results have been striking. Cohen et al. (2006) showed that a VA intervention implemented in middle-school classes significantly increased course grades for African-American students, but did not affect grades for European American students. The achievement gap for African-American students, relative to European American students, was reduced by 40%. Moreover, follow-up research showed that the VA intervention had long-lasting effect on students' overall GPA across all academic classes, persisting over a two-year period with supplemental VA exercises (Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009). More recently, Sherman et al. (2013) tested the VA intervention in a middle school where 45% of students were Latino American and 47% were White, and found that the intervention significantly improved grades for Latino students, but did not affect grades for White students, thereby partially closing the achievement gap for Latino students. These effects persisted over three years (the length of the study): in control conditions, grades for Latino students decreased sharply over time, whereas grades for Latino students in VA conditions remained stable, suggesting that the VA intervention changed the trajectory of academic performance for Latino students.

Finally, one study addressed the gender achievement gap in physics, and tested the VA intervention in a double-blind randomized study in a college physics class (Miyake et al., 2010). Results indicated that men outperformed women in control conditions, but that the VA intervention improved exam scores as well as scores on a standardized test of conceptual mastery of physics concepts for women. The gender gap in physics performance was reduced by 61%, suggesting that this brief intervention was effective in helping women perform better. Considered together, the results of these three randomized field studies suggest that the VA intervention can be a powerful tool for educators.

Although the VA technique was first used to address concerns about stereotype threat for minority students and women (Steele & Aronson, 1995), we hypothesized that the VA intervention might also prove effective for FG students, either by addressing the stereotype threat that FG students experience in academic contexts (Croizet & Millett, 2012) or by addressing cultural identity threat issues. In particular, Stephens and colleagues recently proposed cultural mismatch theory (Stephens, Fryberg, Markus, Johnson, & Covarrubias, 2012; Stephens, Markus, & Fryberg, 2012), in which they argued that FG students face an unseen disadvantage due to a cultural mismatch between the independent norms of the American university system and their own interdependent motives for attending college. They provided empirical support for three claims: 1) American university culture reflects pervasive middle-class norms of independence, 2) students are disadvantaged when there is a mismatch between their personal norms and university culture, such that students who endorsed interdependent motives for attending college obtained lower grades in their freshman year, and 3) the cultural mismatch experienced by FG students causes them to experience college settings as relatively unfamiliar, uncomfortable, and difficult, leading to a reduced sense of "fit" or belonging (Smart-Richman & Leary, 2009; Walton & Cohen, 2007) and poorer performance (Goldrick-Rab, Carter, & Wagner, 2007; Ostrove & Long, 2007; Trent, Orr, Ranis, & Holdaway, 2007). More recently, Stephens, Townsend, Markus, and Phillips (2012) have also shown that FG students experienced more stress (indexed by cortisol levels) and negative emotions when a task was framed with independent versus interdependent cultural norms. These results highlight the discomfort that FG students can experience when their personal motives are inconsistent with university norms, offering further support for cultural mismatch theory (Stephens, Fryberg, et al., 2012).

Although the VA technique has proven to be effective in promoting performance for stereotyped groups in middle school and college physics classes, it has not been tested with FG students. Both the stereotype threat model and cultural mismatch theory lead to the prediction that the VA intervention should be effective for FG students. When FG students write about their most important values, they may bolster themselves against perceived identity threats (Cohen, Purdie-Vaughns, & Garcia, 2012), whether those threats are due to stereotypes about their group (Croizet & Claire, 1998) or a mismatch between personal and institutional norms (Stephens, Fryberg, et al., 2012). Indeed, recent research suggests that VA interventions promote a sense of social belonging or academic fit (Cook, Purdie-Vaughns, Garcia, & Cohen, 2012; Good, Rattan, & Dweck, 2012; Schnabel, Purdie-

Vaughns, Cook, Garcia, & Cohen, 2013) and this may be particularly effective for FG students who endorse more interdependent motives for attending college and who may experience a lower sense of academic belonging (Stephens et al., 2012). In other words, both theoretical models suggest that FG students experience stress and uncertainty in college courses. Focusing on important values may help FG students cope with this uncertainty and stress and promote more effective performance in classes.

Scaling up: Can the VA Intervention Work in Undergraduate Introductory Biology?

Calls for reform in science education have been extensive (Bybee & Fuchs, 2006; Mervis, 2013; Schulz, 2009), including reform in undergraduate biology education (Brewer & Smith, 2011; Momsen, Long, Wyse, & Ebert-May, 2010). At the undergraduate level, reforms typically must be implemented in the challenging context of large-enrollment introductory courses. However, the VA intervention has not been tested on the scale of a large introductory course with multiple instructors and numerous discussion and laboratory sections. In such classes, each student deals with many instructional personnel on a daily basis. Previous implementations of the VA intervention were tested in several middle-school classes, but each class was taught by a single teacher (Cohen et al, 2006; Sherman et al., 2013). In the only college study, the intervention was implemented in two sections of an introductory physics course taught by the same lecturer (Miyake et al., 2010).

It may be significantly more challenging to implement the VA intervention in a larger course where there are multiple sections of large lectures, many instructional staff, and several discussion and laboratory sections, all of which contribute to an impersonal context in which it may be difficult to administer a personal writing exercise. Yet these are precisely the conditions under which thousands of students take these courses every year, and it is important to test the VA intervention in this context. Cohen and colleagues (2012) have argued that the success of this intervention depends on students perceiving the writing exercise as a course assignment coming from the instructor. The exercise needs to be presented in class, so that students see it as an integral part of the course (as opposed to something that researchers bring to students, for example). In complex, large science classes, however, there may be many instructors, and it is not clear whether the intervention can work across multiple lecturers in the same course. In addition, Cohen and colleagues posit that implementation efficacy in college courses depends on the confidentiality of the students’ writing, so that they can write about their core values without worrying about their instructors seeing or judging their writing. Thus the writing exercise must be part of the course, but not graded or evaluated, which creates a tricky balance in an undergraduate science class where every assignment factors into a complex grading structure. These important requirements may be easier to satisfy in the context of a small class taught by a single instructor, and more difficult on a larger scale (Yeager & Walton, 2011).

The Current Study

In the study reported here, we tested the VA intervention in a double-blind randomized experiment in an introductory biology sequence at a large public Midwestern university. We addressed Cohen and colleagues' (2012) recommendations about course-connectedness and confidentiality to implement the VA intervention on a large scale by working closely with course administrators. We hypothesized that this VA intervention would prove effective in closing the social class achievement gap and in promoting FG students' decisions to continue in biology.

Method

Overview of Course

The study was conducted in the two-semester sequence of introductory biology for biology majors and pre-medical students, a course typically taken in the sophomore year. Without this course, students cannot go on to any of 34 undergraduate biomedical majors (e.g., bacteriology, biochemistry, neuroscience, nursing, zoology); it is the critical gateway course for pre-medical preparation and further study in the biological sciences. We conducted the experiment in the first course of this sequence. This 15-week course covered 3 units: cellular biology, genetics, and evolutionary biology; students attended three 50-minute lectures per week, in one of three lecture sections, each taught by two or three different lecturers. The content covered was comparable across the three lecture sections. Exams were specific to each lecture section but grading standards were consistent across sections and students were graded on the same scale.

Course structure—In addition to lectures, students also attended a weekly 3-hour laboratory section, led by a graduate teaching assistant. There were 40 laboratory sections with about 20 students in each section, taught by a total of 14 teaching assistants. Students also attended a weekly 50-minute recitation. There were 40 recitation sections and they were led by a different set of 8 graduate teaching assistants. This complex course structure exposed students to many instructional staff (multiple lecturers and two different TAs) as well as a course coordinator (an academic staff member) for each lecture section. All communications about course requirements and grading came from the course coordinator.

Participants

Of the 804 undergraduate students who received a final grade at the end of the semester, 798 had agreed to participate in this research and gave consent for access to their academic records.[1] The final sample comprised 320 men and 478 women, with 644 continuing-generation and 154 first-generation students. In this sample, 7.6% of students were underrepresented minorities (URM: African American, Hispanic, or Native American), and 92.4% were White or non-targeted ethnic minority students (80% White, 12.4% Asian or Asian American), hereafter referred to as majority students. Of the 154 FG students, 3.2% were African American, 7.1% were Hispanic, and 1.3% were Native American. Of the 644 continuing-generation students, 2.6% were African American, 3.1% were Hispanic, and 1% were Native American, suggesting that URM students were somewhat overrepresented in

[1]Twenty-four students dropped the course over the semester, and 12 of these students were FG, 2 were Black (1 FG), and 3 were Hispanic (2 FG). The fact that 50% of the dropouts were FG students, whereas only 15% of the sample was FG, and that 20% were URM whereas only 7.6% of the sample was URM, highlights the importance of addressing motivation and retention issues for these students.

the first-generation group, as expected, $\chi^2(1, N=798)= 4.421$, $p=.04$. However, URM students were not as heavily represented among FG students as on other campuses (for example, Stephens, Fryberg, et al. (2012) found that 50% of their sample of FG students were African American or Hispanic). Indeed, 91% of the FG students in this study were majority. This ethnic distribution is advantageous because it permits a test of FG effects disentangled from URM status.

The Intervention

Students were blocked on generational and URM status, gender, and lecture section, and randomly assigned to condition within lab sections, in a double-blind design. Students completed either a VA writing exercise or a control writing exercise in their laboratory sections. These writing assignments were delivered early in the semester (week 3), with a second administration shortly before the second exam (week 8). Each student completed either two values-affirmation writing exercises or two control writing exercises of similar format and length. There were 325 continuing-generation and 77 first-generation students in the affirmation condition and 319 CG and 77 FG students in the control condition.

The Friday before each writing exercise was to be administered in laboratory sections, the course coordinators included the following text in the weekly newsletter emailed to all students: "This week there will be a special writing exercise in the first ten or fifteen minutes of lab. This is designed to give you additional practice in both critical thinking and writing, which are essential parts of any career in biology (or anything else for that matter). There is no need to study for this. This in-class writing will be about something you know well. We want you to feel comfortable with this type of practice, so we've asked an independent group to administer the writing exercise so that it can be confidential. They will let us know if you have completed the exercise so that you can receive credit for it, but your TAs, professors, and coordinators will not see your work. This is one of two such exercises which are required in lab." These instructions ensured that students knew that the assignment came from their professors and was required for class, but that they would be writing in confidence, and that the content of their work would not be evaluated.

Laboratory sections were led by graduate student TAs who were naïve to the purpose of the study. Study personnel arrived at the beginning of the laboratory period and distributed personally addressed manila envelopes that contained the writing assignment (which had been assigned in advance, based on the randomized blocked design). Although there were two versions of the writing assignment (values affirmation and control), the envelopes and formatting of the two exercises were similar. Students in each experimental condition received a three-page packet. The first page listed 12 values: *being good at art; creativity; relationships with family and friends; government or politics; independence; learning and gaining knowledge; athletic ability; belonging to a social group (such as your community, racial group, or school club); music; career; spiritual or religious values*; and *sense of humor*. The values and procedures were similar to those developed and validated in past

[2]Two students did not answer the concern about background question. Although significantly correlated with Confidence (r(794) = −.39, p<.05) factor analyses indicated that this item did not load on the Confidence factor (including it in the Confidence scale reduced alpha from .89 to .21) and it was therefore treated as a separate variable.

research (Cohen et al., 2006; McQueen & Klein, 2006; Miyake et al., 2010; Sherman & Cohen, 2006). Cohen et al. (2006) note that it is important that the list include a broad set of values, but not include values directly related to academic performance (grades, evaluation, etc.). Students in the affirmation condition were instructed to circle the two or three values *most* important to them, whereas students in the control condition were instructed to circle the two or three *least* important values. The second page instructed students to describe in a few sentences either why the selected values were important to *them* (VA condition) or why they might be important to *someone else* (control condition). Students were told to focus on their thoughts and feelings, and not to worry about spelling and grammar or how well written their essay was. The final page reinforced the manipulation by asking students to look again at the values they had selected earlier. They were then asked to list either the top two reasons why these values were important to them (VA condition) or the top two reasons why someone else might pick these values as important (control condition). To encourage further reflection about the values, the third page ended by asking students to indicate their agreement with several items using numerical scales (e.g. *In general, I try to live up to these values* in the affirmation condition vs. *In general, some people try to live up to these values* in the control condition). Students put the writing exercise back in the manila envelope when they were done, ensuring that TAs and study personnel remained blind to condition.

A second administration of the writing exercise was delivered in the same manner shortly before the second midterm exam (week 8). The writing exercise was similar to the first writing exercise, except in the second exercise 4 values were added to the list: *curiosity, school spirit, nature and environment, and online social networking and/or gaming*, to make the assignment seem slightly different.

If students were not present during their laboratory section, they were given the opportunity to complete the writing assignment online via an emailed link. 797 students completed the first writing exercise (795 in laboratory sections) and 793 students completed the second writing exercise (790 in laboratory sections). Several steps were taken to ensure that all instructional personnel associated with the course were unaware of students' condition assignment. All but senior research personnel were blind to the study's purpose and hypotheses. Both writing exercises occurred without the course instructors or coordinators present. Laboratory TAs were informed by course coordinators that their students would be completing a confidential writing assignment, and remained blind to experimental condition.

Baseline Measures

Baseline measures were obtained in the second week of the course. Students were asked to complete a survey about their attitudes toward biology in their laboratory section (students absent during this week were sent a link via email to fill out the survey online). All items were answered on a 7-point scale ranging from *not at all true* to *very true*. Confidence about Performance was measured with 3 items, alpha = .89 ("I am confident that I will do well in Introductory Biology", "I expect to get a good grade in this course", "I am confident that I can obtain a final grade of B or better in this course"). Concern about Background was measured with 1 item ("I am not sure I have the right background for this course").[2]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Students were asked to report their age, gender, race/ethnicity, and year in school. Students were also asked to indicate the highest level of education that their mother (or guardian) and father (or guardian) completed: grade school, high school, technical school, some college, bachelor's degree, or graduate degree. We identified students as FG college students if neither of their parents or guardians earned a bachelor's degree.

We obtained other baseline data from students' academic records, specifically their ACT and/or SAT scores, their cumulative credits at the university, and their GPA for the prior spring semester.

Outcome Measures

Questionnaire measures were obtained in the fourteenth week of the 15-week course. Students were asked to complete a brief survey in their assigned laboratory section (students absent during this week were sent a link via email to fill out the survey online). Confidence about Performance and Concern about Background were measured with items comparable to those used at baseline, adjusted to reflect timing of measurement.

Grades—Course coordinators provided final course grades at the end of the semester, and students' grades for that semester across all courses were obtained from university records. We constructed a GPA for the other courses taken by students that semester, excluding the five credits of biology, so that Biology Grade and Semester GPA could be analyzed separately. Grades at this university are calculated on a 4.0 scale (A=4.0, AB=3.5, B=3.0, BC=2.5, C=2.0, D=1.0, F=0).

Continuation in the second semester—We tracked students for two months after completing the first course, to see whether they enrolled in the second course in the biology sequence in the following spring semester (when 75–80% of students typically continue).

Results

Preliminary Analyses

Randomization check—We used a 2 (Control vs. VA) × 2 (FG vs. CG) analysis of variance model to test whether there were differences on baseline measures of age, cumulative credits, confidence about performance in the class, or concern about background. These analyses allowed us to test whether randomization was effective, whether FG students differed from CG students at baseline, and whether there was a significant interaction between experimental condition and generational status on any baseline measure. We also analyzed two measures of prior performance, although we did not have complete data from all students: prior spring GPA (not available for the 104 first-year students and 4 upperclass students) and ACT scores (not available for 87 students). There were no significant differences between conditions or significant interactions of condition with generational status for any baseline variable, indicating that randomization was successful. We did find two significant main effects for generational status: on age, $t(796) = 5.17$, $p<.001$, indicating that FG students (M=19.70, SD=2.08) were slightly older than CG students (M=19.16, SD=.82), and on ACT, $t(707) = 6.86$, $p<.001$, indicating that CG students (M=28.85, SD=2.63) had higher scores than FG students (*M*=27.11, *SD*=3.02). No other differences were significant.

Distribution by school year—We also tested whether the distribution of students' year in school varied as a function of condition and found that it did not. Table 1 presents means, standard deviations and/or frequencies for all baseline measures as a function of treatment condition; no differences were significant, suggesting that randomization was successful.

Primary Analyses

The three primary outcome measures were the final grade in the biology class, semester GPA (for all courses taken that semester, excluding biology), and continuation in the second course in the introductory biology sequence. Because students were randomly assigned to condition within lecture and lab sections (and blocked on gender, FG and URM status), the data were analyzed using multiple regression models, testing treatment effects at the student level, controlling for lecture section. We used OLS regression for the two continuous outcome measures (Biology course grade and Semester GPA) and logistic regression for the dichotomous outcome measure (continuation in the second biology course). The same regression model was tested for each outcome measure, controlling for lecture section. We tested all interactions between treatment condition (control, −1, VA intervention +1), generational status (CG −1, FG +1), gender (males +1, females −1) and lecture section (2 orthogonal codes to control for differences between the 3 sections), and then trimmed all interaction terms that were not significant in any model. The final model reported for all three outcome measures includes 8 terms: the main effects of treatment condition, generational status, gender, and lecture section (two terms), as well as 3 two-way interactions (one between condition and generational status, and two between generational status and lecture condition).[3] Table 2 presents the means, standard deviations and intercorrelations for all measures.

Biology course grade—A significant main effect of generational status indicated that FG students obtained lower grades than CG students in the biology class, $t(789)= 5.59$, $p<.01$, $\beta = -.20$). However, this main effect was qualified, as predicted, by a significant interaction between treatment condition and generational status, $t(789)= 2.17$, $p=.03$ ($\beta = 0.10$), indicating that the VA intervention reduced the achievement gap in course grades. As reported in Table 3 and shown in Figure 1, CG students outperformed FG students in the control condition (the achievement gap), but this difference was considerably smaller in the VA condition. The effect size for the achievement gap was moderate in the control condition (Cohen's d=.39; $t(789)= 5.52$, $p<.001$) but much smaller in the VA condition, (d=.18; $t(789)= 2.49$, $p<.05$). The treatment effect for FG students was significant, $t(789)= 2.10$, $p<.05$, reflecting an average grade difference of .24, or approximately a quarter of a grade point, resulting in a 50% reduction in the social class achievement gap.

[3]We found very few effects of lecture section. A significant main effect of lecture section on both biology course grades and semester GPA, $t(789)= 3.11$, $p<.01$ ($\beta = .14$ for Biology grade), and $t(789)= 3.01$, $p<.01$ ($\beta = .13$ for Semester GPA), indicated that students in Lecture 1 obtained higher grades than students in the other two lectures. There was also a nearly significant generational status × lecture interaction on Biology grade, $t(789)= 1.97$, $p=.05$ ($\beta = -.09$), suggesting that there was a larger achievement gap in Lecture 2. No other interactions with lecture section or gender were significant on any of the three measures, indicating that treatment effects did not vary as a function of lecture section or gender.

To explore the nature of the treatment condition by generational status interaction in greater detail, we analyzed the distribution of grades, and found that FG students were more likely to earn B's in the VA condition (62.3%) than in the control condition (41.6%), whereas FG students were more likely to earn C's in the control condition (35.1%) than in the VA condition (18.2%), $\chi 2(1, N=121)=7.25$, $p=.01$, (Figure 2).[4] In contrast, the distribution of grades for CG students did not differ as a function of experimental condition, $\chi 2(1, N=491)= .59$, $p=0.46$. This pattern of results suggests that the intervention was most effective in moving FG students from C's to B's, which could be critically important for retention in a field.

Semester GPA

An analysis of students' semester GPA (excluding the biology course grade), indicated a significant main effect of generational status, $t(789)= 4.79$, $p<.001$ ($\beta = -.17$), indicating that FG students obtained lower grades than CG students in all other courses taken that semester. However, this effect was qualified by the predicted condition × generational status interaction, $t(789)= 2.38$, $p=.02$ ($\beta = .11$), indicating that the VA intervention improved overall semester GPA for FG students, relative to the control condition. The treatment effect for FG students was significant, $t(789)= 2.36$, $p=.02$, reflecting an average grade difference of .24, or approximately a quarter of a grade point, resulting in a 50% reduction in the social class achievement gap. The main effect of gender was also significant, $t(789)= 2.63$, $p=.01$ ($\beta = -.09$), indicating that males had lower grades than females in other courses taken that semester.

Continuation in second semester—We tested the same model used for biology grade and semester GPA, but used binary logistic regression to examine whether students enrolled in the second course in the biology sequence. We found a significant main effect of condition, Wald= 4.68, $p=.03$ (B= .24); however, this was qualified by a significant condition × generational status interaction, Wald= 8.41, $p<.01$ (B= .33), indicating that the VA intervention promoted continued enrollment for FG students relative to those in the control condition. As shown in Figure 1, in the control condition, CG students were more likely to enroll in the second course (77.7%) than FG students (66.2%), but in the VA condition, FG students (85.7%) were more likely to enroll than CG students (74.8%), representing a 20% increase in enrollment for FG students, $\chi 2(1, N=154)=8.00$, $p<.01$. In contrast, CG students' enrollment did not differ according to experimental condition, $\chi 2(1, N=644)=.786$, $p=.41$.[5]

[4]In these analyses, we included BCs in the B category and ABs in the A category.

[5]We also analyzed the primary outcome variables controlling for previous performance, using a composite measure of prior achievement. We z-scored prior spring GPA, as well as ACT and SAT scores for all participants for whom we had these measures. If we had prior GPA, we used that z-score; if not, we used the ACT z-score, and if we didn't have ACT, we used the z-score for the SAT. This allowed us to estimate a "baseline performance" measure for all but 24 students (for whom we lacked any information about prior performance). The remaining missing data were handled through multiple imputation (Rubin, 1987). There was no significant difference between conditions or a significant interaction of condition with generational on this composite measure. We controlled for this baseline performance variable in our three primary analyses, and found that the significant interaction of condition and generational status remained significant for course grade $t(788)= 2.54$, $p=.01$ ($\beta = .09$), semester GPA $t(788)= 2.78$, $p=.01$ ($\beta = .10$), and continuation Wald= 8.80, $p<.01$ (B= .34). These analyses indicate that our central finding – a positive treatment effect for FG students on performance and retention in the class – remained significant, controlling for prior performance.

Mediation analysis—We used Preacher and Hayes' (2004) bootstrapping procedure to test whether final grade in the first course mediated the treatment effect on enrollment in the second course, and found evidence for partial mediation such that the VA intervention improved continued enrollment for FG students by increasing their grades in the first course (see Table 4). Results based on 5000 bootstrapped samples indicate that the total effect (TE) of the condition × generational status interaction was significant (TE=.32, SE = .11, p<.01) as was the direct effect (DE=.29, SE=.12, p=.01), suggesting partial mediation. Final grade mediated the relationship between the interaction and enrollment, 95% CI [.0024, .1353]. The fact that zero falls outside this interval indicates significant mediation, p =.04.

Supplementary Analysis

Our primary analyses focused on generational status, but of course many first-generation students are also from URM groups, and in any study, it is important to consider whether treatment effects are due to students' FG or URM status. Given the small number of URM students in our sample, we can be confident that our intervention was differentially effective because of the generational status of students, and not their URM status, but we examined this possibility in more detail in supplementary analyses.

Specifically, we examined treatment effects as a function of both generational and URM status, although the small number of URM students in the sample limited statistical power to detect treatment effects for URM students. In these supplementary analyses, we categorized students into one of three groups: Majority CG (n=601), Majority FG (n=136) or URM (n=61). In other words, students who were both FG and URM were categorized as URM for these analyses, and students who were both CG and URM were also categorized as URM. This classification system allowed us to evaluate treatment effects for URM students as well as for Majority FG students, in comparison to Majority CG students, using dummy codes to test for Majority FG and URM effects, with regression analysis, with course grade as the outcome measure.

We found significant main effects for both Majority FG and URM status, indicating that both groups of students performed more poorly in the biology class than Majority CG students, t(791) = 4.78, p<.001 (β= −.24 for Majority FG students), and t(791) = 3.37, p<.01 (β= −.17 for URM students). The intervention effect for Majority FG students was significant, t(791) = 2.00, p<.05 (β=.10), whereas the intervention effect was not significant for URM students, t(791) =.11, p=0.91 (β=.01), indicating that the intervention improved performance for FG students, but did not affect performance for URM students (see Figure 3). We found a similar pattern of effects on Semester GPA and Continuation: the intervention effect was significant for Majority FG students on each measure, t(791) =2.10, p=.04 (β= .11 for Semester GPA), and Wald =8.01, p=.01 (B=1.37 for Continuation), and the intervention effect for URM students was not significant.[6] These results help to

[6]We also examined course grades in a 2×2x2 model in which we tested experimental condition by generational status by URM status, and found that the condition by generational status interaction was significant, t(785)= 2.03, p=.04 (β = .09), and that the condition by URM interaction was not significant, p=.51. Moreover, the 3-way interaction between condition, FG status, and URM status was not significant, p=.70. These results indicate that the treatment effect reported earlier remained significant with URM status controlled, and that treatment effects did not differ as a function of URM status.

disentangle FG effects from URM effects, and suggest that the VA intervention was effective for FG students because of their generational status, and not their URM status.

Process Analyses

We examined the values that students chose in the values affirmation condition, to test whether FG and CG chose different values to write about. Table 5 shows the percentage of students who chose each value; FG and CG students did not differ significantly in their choice of values to write about.

We examined whether confidence about performance or concern about their background changed as a function of treatment condition or generational status. Although FG and CG students did not differ on either measure at the outset of the class, they did by the end of the semester. FG students were less confident about their performance in the class, $t(788)= 2.90$, $p<.01$ ($\beta = -.09$), and more likely to believe that they did not have the right background for the class, $t(786)= 1.98$, $p=.05$ ($\beta = .07$), compared to CG students, across experimental conditions.

The VA intervention did not influence confidence about performance at the end of the semester for FG students, $t(788)= 1.13$, $p=.26$ ($\beta = .05$), but we did find a significant condition by generational status interaction showing that the VA intervention reduced concerns about background for FG students by the end of the semester $t(786)= 2.00$, $p=.05$ ($\beta = -.08$). Because baseline levels of concern about background were controlled in this analysis, this effect represents the *change* in concern over time. Figure 4 shows concern about background scores at the beginning and end of the semester. FG students in the control condition reported greater concern about their background (M=3.08, SD=1.92) than CG students (M=2.57, SD=1.53) at the end of the semester. This gap was completely closed in the intervention condition (M=2.74, SD=1.74 for FG students; M=2.74, SD=1.59 for CG students).[8] For FG students, this increase in concern may occur as they gain a more realistic understanding of the difficulty level of university-level biology and receive feedback about their performance, but the VA intervention seemed to offset these mounting concerns.

We also found a significant effect of gender showing that women reported lower levels of confidence than men, both at the outset and end of the semester, $F(1,796)$=44.52 and 38.31,respectively, $p<.01$ (although there were no gender differences in course grades), but did not find an effect of gender on Concern about Background at either timepoint.[7] For confidence, for males at baseline M = 6.00 (SD=.81), and at outcome, M = 5.32 (SD = 1.19). For females, at baseline M = 5.57 (SD=.95), and at outcome = 4.87 (SD = 1.33), and d = .49 and .36 for the gender difference, at baseline and end of semester, respectively, reflecting a moderate effect size for the gender difference. This gender difference for confidence contrasts with the absence of a gender difference in course grade and a significant gender

[8]We tested whether concern about background mediated the treatment effect on final grades and enrollment in the second course, and although it was a significant predictor of final grade, $t(788)= 7.98$, $p<.001$ ($\beta = -.27$) and enrollment, Wald =9.81, p<.01 (B=−.16), it was not a significant mediator of either treatment effect, 95% CI [−.0027, .0366].

[7]We also examined whether URM students differed from Majority students in terms of confidence about performance or concern about background, at baseline or by the end of the class, but found no significant differences (p ≥20 for all four tests). As noted earlier, the small number of URM students in this sample limited the power to detect effects.

difference favoring females for overall semester grades. For measures such as confidence about performance, debate centers on interpretation of the gender difference. Do females lack confidence or are males unrealistically over-confident? Either direction could carry costs. A lack of confidence might lead to a student not taking on challenging academic tasks. What is less recognized is that unrealistic over-confidence may lead to negative behaviors such as not studying adequately for an exam. Research indicates that females tend to underestimate themselves by about as much as males overestimate themselves (Cole et al., 1999).

Survey Study: Characterizing the Experience of FG Students in Biology

In addition to concern about their background for the course, FG students may worry about “fitting in” more generally, and may also experience discrepancies between their motives for attending college and University norms (Stephens et al., 2012). In a separate survey study, conducted in a section of the same biology course in a different semester, we administered a series of questionnaires to 772 students at the end of the semester (318 males, 454 females; 613 CG, 159 FG students; 53 URM, 719 Majority). We administered Stephens et al.’s (2012) measure of motives for attending college, Sherman, Bunyan, Creswell, and Jaremka’s (2009) measure of academic and social concerns, Walton and Cohen’s (2007) measures of belonging uncertainty and level of belonging, as well as the same measures of confidence about performance and concern about background reported earlier, to explore differences between FG and CG students in greater detail.

We administered a shorter version of Stephens and colleagues’ (Stephens et al., 2012) scale assessing students’ motives for attending college, in which students were asked to indicate which of 10 items characterized their reasons for completing their college degree (checking as many as were relevant). Half the items referred to independent motives that reflect typical American University values (e.g. becoming an independent thinker, exploring new interests) whereas the other half referred to interdependent motives commonly associated with working-class values (e.g. giving back to my community, helping the family). We constructed measures of independent and interdependent motives, $r(770) = .31$, $p<.001$ (see Table 6 for percentage of interdependent and independent items endorsed by FG and CG students).

The Sherman et al. (2009) measure consisted of four items referring to academic and social concerns at college (“In college, I sometimes worry that people will dislike me”, “In college, I worry that people will think I’m unintelligent if I do poorly,” “I am usually confident that others will have a good impression of my ability,” “In college, I often get nervous and worried when I talk to people”). The Walton and Cohen (2007, 2011) measures included two items to measure belonging uncertainty, adapted for this university (“When something bad happens, I feel that maybe I don’t belong at University X,” “Sometimes I feel that I belong at University X, and sometimes I feel that I don’t belong at University X”), and a single item to measure level of belonging (“I belong at University X”).

In addition to the 4 items used to assess confidence about performance and concern about background in the intervention study, we added 3 new items to assess students’ sense of belonging in the course and the field of biology (“There were times that I felt I didn’t belong in this class," "I felt like an outsider in this class," "I don't know if I really belong in the field of biology"). We conducted correlational and factor analyses to explore the interrelations of our single-item measure of concern about background with perceived confidence, measures of course-specific belonging, and Walton and Cohen's (2007) single-item measure of belonging. Based on principal axis extraction, we identified two factors with eigenvalues of 4.05 and 1.18, explaining 65% of the variance. After oblique rotation, the first factor had 3 items with high loadings (>.30) that measured confidence about performance. The second factor had 5 items with high loadings that measured level of belonging. Based on these findings, we constructed two scales, a 3-item Confidence about Performance scale (alpha = .89), identical to the measure used in the intervention study, and a new 5-item Academic Belonging scale (alpha = .78), which included our original measure of Concern about Background (reversed), Walton and Cohen's (2007) single belonging item, and the 3 new items (all reversed). These two scales were correlated, $r(769)=.56$. These results suggest that the concern about background measure tested in the intervention study is strongly related to feelings of belonging (in the course, at the university, and in the field). However, these results also suggest that feelings of academic belonging in a class are strongly related to confidence about performance in the class.

Table 7 shows the intercorrelations of all the survey measures. We tested whether FG and CG students differed on these measures, and also tested for effects of gender and URM status in a 2 (FG vs. CG) × 2 (URM vs. majority) × 2 (male vs. female) regression model. No interactions were significant. We found that FG students were significantly higher than CG students in interdependent motives, $t(768)= 3.12$, $p=.002$ ($\beta = .11$), and significantly lower in independent motives, $t(768)= 2.87$, $p=.004$ ($\beta = -.10$). They were also significantly higher than CG students in academic concerns, $t(767)= 2.64$, $p=.009$ ($\beta = .09$) and belonging uncertainty, $t(767)= 3.60$, $p<.001$ ($\beta = .13$), and lower in academic belonging, $t(768)= 2.12$, $p=.035$ ($\beta = -.08$). They did not differ from CG students in perceived confidence, $p>.25$. The fact that FG students did not differ from CG students in confidence, as they did in the intervention study, suggests that this effect is not robust. However, FG students did differ from CG students on the academic belonging measure. Given that this scale included the concern about background measure that showed an FG difference in the intervention study, our results suggest that the psychological experience of academic concerns may be more important to understanding the experience of FG students. We found one significant effect for URM status and one marginal effect: URM students reported higher levels of interdependent motives than majority students, $t(768)= 3.05$, $p=.002$ ($\beta = .11$) and marginally lower levels of confidence, $t(767)= 1.68$, $p=.09$ ($\beta = -.06$). We also found three significant effects of gender: women reported lower levels of confidence than men, $t(767)= 2.04$, $p=.04$ ($\beta = .07$) and higher levels of belonging uncertainty, $t(767)= 2.50$, $p=.013$ ($\beta = -.09$) and academic concerns, $t(767)= 3.50$, $p<.001$ ($\beta = -.12$). Table 7 shows the means for CG and FG students. Considered together, these findings suggest that FG students experienced a mismatch between their motives and university norms, lower levels of perceived belonging in the course, higher levels of academic and social concerns, as well as higher levels of uncertainty about belonging in the course. All of these factors may contribute to the social class achievement gap observed here, and may also help account for the effectiveness of the VA intervention for FG students in this biology class.

Discussion

If we wish to expand the pipeline of students entering the biomedical sciences, it will be critically important to promote performance and retention for FG students in introductory courses that act as a gateway to successive courses and careers in biology. Introductory biology, as taught at large public universities, can be an overwhelming and impersonal experience for students, and the VA intervention played a critical role in helping FG students meet the challenges of this environment. This study is the first, to our knowledge, to address the achievement gap for any group of underrepresented students planning on entering the biomedical sciences. Nationwide, FG students represent a large pool of potential scientists. To provide the most equitable opportunities for these individuals, and to maximize discovery of talent for the nation, it is crucial that FG students have a positive experience in their first biology course. For example, a student who receives a C gets a very different message from one who receives a B; the C may indicate that you can't make it in the field, whereas the B signals that you can. This study demonstrates that a values affirmation intervention can narrow the social class achievement gap, improve the success rate for FG students in biology class (as well as college classes more generally), and keep them on track for progress in biology courses, even in the context of a large, impersonal course.

The fact that we observed intervention effects on grades in other classes taken in the same semester, and on continuation into the second semester of the course highlights the power of the VA intervention to influence performance and course enrollment decisions in an ongoing manner (Cohen et al., 2012). The biology class studied here was a five-credit class for majors and pre-medical students, and was probably the most important course of the semester for most students. A positive experience in this critical course could influence FG students' academic performance more generally, and color their experience at the university, with far-reaching effects. Cohen et al. (2012) have argued that early performance outcomes can be carried forward through recursive cycles, and start a positive chain reaction. For example, Cohen et al. (2006) found VA effects on grades in the course in which the affirmation was completed, but follow-up analyses (Cohen et al., 2009) showed long-lasting and more general effects, specifically, on students' grade-point-averages (across classes) and perceived belonging over two years, among initially low-performing African American students (Cook et al., 2012; Sherman et al. 2013). Our results contribute to the growing body of work documenting the far-reaching effects of VA interventions.

There are many disadvantaged, stereotyped, and underrepresented students who struggle in college, and it is important to recognize that they struggle for different reasons. FG students face a unique set of financial and cultural challenges that place them at risk in college, particularly in introductory science courses where they may not have the same background and preparation as CG students (Bowen et al., 2005; Terenzini, Springer, Yeager, Pascarella, & Nora, 1996). These students have different motives for attending college (Stephens, Fryberg, et al., 2012) and they expressed concerns about their preparation for introductory biology in our study.

We found that FG students endorsed more interdependent reasons for attending college, and fewer independent reasons, relative to CG students, replicating Stephens, Fryberg et al. (2012). FG students also expressed lower levels of academic belonging and higher levels of uncertainty about belonging (Walton & Cohen, 2007), as well as higher levels of academic concerns (Sherman et al., 2009), relative to CG students. Considered together, these findings suggest that FG students were at a considerable disadvantage in this course. Whether their discomfort reflects the same type of stress experienced by other stereotyped students, or, as suggested by our survey results, a unique type of discomfort attributable to cultural mismatch (or a combination of stressors), our results demonstrate that a VA intervention can help FG students in the same way that it has been shown to help African American and Hispanic students in middle school and women in physics (Cohen et al., 2006; Miyake et al., 2010; Sherman et al, 2013). This research represents a novel application of the VA method and demonstrates that it is applicable to groups that have not previously been considered.

The relatively low representation of URM students among FG students in this sample was actually advantageous insofar as it allowed us to disentangle whether it was FG or URM status that accounted for the effectiveness of the intervention. Indeed, just as we attempted to disentangle generational status from ethnic minority status in this study, researchers working on closing URM achievement gaps may wish to disentangle these effects from generational status. For example, Sherman et al. (2013) noted that almost all of their Latino participants were also low-SES students, and that the race gap in their study was largely redundant with a social class gap, suggesting that their results may have been due to social class as much as to Latino status. Given the increasing overlap of racial and SES groupings in American society (Reardon, 2011), it seems that our understanding of racial achievement gaps may be informed by consideration of social class and cultural mismatch, and that continued efforts to integrate stereotype threat and cultural mismatch theory are warranted. At the same time, it will be important to identify the specific mechanisms underlying the under-performance of different groups (whether identity threat, broadly construed, social rejection threat, stereotype threat or cultural mismatch) so that future interventions can be tailored to the specific type of identity threats students face in particular contexts (Shnabel et al., 2013).

Indeed, a limitation of our study is that we were unable to distinguish between stereotype threat and cultural mismatch mechanisms, or more critically, measure variables that might have mediated the effects of the VA intervention for FG students. The two theoretical models characterize the *source* of identity threat differently – concerns about stereotypes about one’s group versus a mismatch of personal and institutional norms, but they characterize the *experience* of identity threat similarly – uncertainty about belonging, stress, and discomfort. As discussed, our survey study provides some support for a cultural mismatch interpretation of FG students’ experience in this class, but the concerns about belonging are also consistent with the stereotype threat model (Good et al., 2012). It will be important to measure perceived threats and stereotypes about FG students in future research to further elucidate the nature of the FG experience.

With respect to mechanism, one interesting point of possible theoretical conflict is whether reflecting on important values has the potential to induce even more cultural mismatch

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(Stephens, Fryberg, et al., 2012). In other words, if students affirm important interdependent values, might they become even more aware of the mismatch between their own motives for attending college and institutional norms? Or, does the opportunity to reflect on these core values help them cope with academic difficulties and their uncertainty about belonging? Our results suggest that *all* students were most likely to choose an interdependent value to write about ("relationships with friends or family"), but that FG students were no more likely to select this value (or any other) than CG students. However, the process of reflecting on those values proved to be especially powerful for FG students in this study, and it will be important to examine this process in future research. It is not clear what the best strategy would be for addressing cultural mismatch – changing the academic context or perception of situational norms, as suggested by Stephens et al. (2012), changing personal motives to match the context, which may be difficult to accomplish, or re-affirming core values to help students cope with challenging situations. What produces identity threat for FG students and why VA interventions ameliorate those threats are two separate questions that each demand further research.

This study is also the first to test the VA intervention on a large scale in college courses, across three sections of an introductory biology course, multiple instructors, and numerous laboratory sections. We worked closely with course faculty and instructional staff to ensure that the writing exercises were administered in a way that made the assignment part of the class, while keeping the exercise non-evaluative and the content of the essays confidential. We tried to implement the intervention in the most effective way possible, but to meet the course-connectedness and confidentiality conditions established by Cohen et al. (2012), we had independent researchers collect the essays so that instructors did not see the essay itself, and this strategy for implementation is cumbersome and costly. Some streamlining will be needed if it is to be used routinely in large science courses, and future research might explore which of these conditions are essential, and which might be relaxed without losing treatment efficacy. Although these conditions were challenging to establish, our results suggest that this intervention can be effectively administered on this scale without reducing the potency of the intervention. These findings contribute to a growing base of research documenting the power of brief social-psychological interventions (Blackwell, Trzesniewski, & Dweck, 2007; Walton & Cohen, 2011; Hulleman & Harackiewicz, 2009; Harackiewicz, Rozek, Hulleman, & Hyde, 2012) and begin to address the challenge of scaling up interventions for implementation on a wider scale. These interventions are relatively simple, making them feasible to implement by non-psychology faculty with some training and attention to implementation details. Although these interventions may seem like "magic bullets" because they are so simple (Yeager & Walton, 2011), they are powerful because they focus on changing the mindset of the students (Wilson, 2011). As such, they can complement other educational interventions that focus on changing the learning environment (Brewer & Smith, 2011; Ruiz-Primo, Briggs, Iverson, Talbot, & Shepard, 2011). Progress in education will be maximized by considering both types of curriculum reform.

Acknowledgments

We thank Kerstin Krautbauer, course coordinators Jean Heitz, Kerry Martin, Brian Parks, and Carlos Peralta, and all the research assistants who helped to implement this study. We also thank Geoffrey Cohen, Tiffany Ito, Steve Kosciuk, Christopher Rozek, and Lillian Tong for their help with the project. This research was supported by the National Institutes of Health (Grant R01GM102703). This research was also supported by the Graduate School, the Office of the Provost, and the Assessment Council of the University of Wisconsin, as well as the Institute of Education Sciences, U.S. Department of Education, through Award # R305B090009 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education or the National Institutes of Health.

References

Aronson, J.; Dee, T. Stereotype threat in the real world. In: Inzlicht, M.; Schmader, T., editors. Stereotype threat: Theory, process, and application. New York: Oxford University Press; 2012.

Aronson J, Inzlicht M. The ups and downs of attributional ambiguity: Stereotype vulnerability and the academic self-knowledge of African-American students. Psychological Science. 2004; 15(12):829–836. [PubMed: 15563328]

Aronson J, Lustina MJ, Good C, Keough K, Steele CM, Brown J. When white men can't do math: Necessary and sufficient factors in stereotype threat. Journal of Experimental Social Psychology. 1999; 35:29–46.

Blackwell LS, Trzesniewski KH, Dweck CS. Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. Child Development. 2007; 78:246–263. [PubMed: 17328703]

Bowen, WG.; Kurzweil, MA.; Tobin, EM. Equity and excellence in American higher education. Charlottesville, VA: University of Virginia Press; 2005.

Brewer, CA.; Smith, D. Vision and change in undergraduate biology education: A call to action. Washington, D.C: American Association for the Advancement of Science; 2011. http://visionandchange.org/files/2011/03/Revised-Vision-and-Change-Final-Report.pdf

Bybee RW, Fuchs B. Preparing the 21st century workforce: A new reform in science and technology education. Journal of Research in Science Teaching. 2006; 43:349–352.

Cohen GL, Garcia J, Apfel N, Master A. Reducing the racial achievement gap: A social-psychological intervention. Science. 2006; 313:1307–1310. [PubMed: 16946074]

Cohen GL, Garcia J, Purdie-Vaughns V, Apfel N, Brzustoski P. Recursive processes in self-affirmation: Intervening to close the minority achievement gap. Science. 2009; 324:400–403. [PubMed: 19372432]

Cohen, GL.; Purdie-Vaughns, V.; Garcia, J. An identity threat perspective on intervention. In: Inzlicht, M.; Schmader, T., editors. Stereotype threat: Theory, Process, and Application. New York: Oxford University Press; 2012.

Cole DA, Martin J, Peeke L, Seroczynski A, Fier J. Children's over- and underestimation of academic competence: A longitudinal study of gender differences, depression, and anxiety. Child Development. 1999; 70:459–473. [PubMed: 10218266]

Cook JE, Purdie-Vaughns V, Garcia J, Cohen GL. Chronic threat and contingent belonging: Protective benefits of values affirmation on identity development. Journal of Personality and Social Psychology. 2012; 102:479–496. [PubMed: 22082058]

Creswell JD, Welch WT, Taylor SE, Sherman DK, Gruenewald TL, Mann T. Affirmation of personal values buffers neuroendocrine and psychological stress responses. Psychological Science. 2005; 16:846–851. [PubMed: 16262767]

Croizet J-C, Claire T. Extending the concept of stereotype and threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. Personality and Social Psychology Bulletin. 1998; 24:588–594.

Croizet J-C, Dutrévis M. Socioeconomic status and intelligence: Why test scores do not equal merit. Journal of Poverty. 2004; 8:91–107.

Croizet, J-C.; Millet, M. Social class and test performance: From stereotype threat to symbolic violence and vice versa. In: Inzlicht, M.; Schmader, T., editors. Stereotype threat: Theory, process, and application. New York: Oxford University Press; 2012.

Fein S, Spencer SJ. Prejudice as self-image maintenance: Affirming the self through derogating others. Journal Of Personality And Social Psychology. 1997; 73:31–44.

Goldrick-Rab S, Carter DF, Wagner RW. What higher education has to say about the transition to college. Teachers College Record. 2007; 109:2444–2481. http://www.tcrecord.org/content.asp?contentid=12565.

Guryan J, Hurst E, Kearney M. Parental education and parental time with children. Journal of Economic Perspectives. 2008; 22:23–46.

Haak DC, HilleRisLambers J, Pitre E, Freeman S. Increased structure and active learning reduce the achievement gap in introductory biology. Science. 2011; 332:1213–1216. [PubMed: 21636776]

Harackiewicz JM, Rozek CR, Hulleman CS, Hyde JS. Helping parents motivate adolescents in mathematics and science: An experimental test. Psychological Science. 2012; 43:899–906. [PubMed: 22760887]

Harrison LA, Stevens CM, Monty AN, Coakley CA. The consequences of stereotype threat on the academic performance of White and non-White lower income college students. Social Psychology of Education. 2006; 9:341–357.

Horvat E, Weininger EB, Lareau A. From social ties to social capital: Class differences in the relations between schools and parent networks. American Educational Research Journal. 2003; 40:319–351.

Hulleman CS, Harackiewicz JM. Promoting interest and performance in high school science classes. Science. 2009; 326:1410–1412. [PubMed: 19965759]

Jackman, MR.; Jackman, RW. Class awareness in the United States. Berkeley, CA: University of California Press; 1983.

Johnson SE, Richeson JA, Finkel EJ. Middle class and marginal? Socioeconomic status, stigma, and self-regulation at an elite university. Journal of Personality and Social Psychology. 2011; 100:838–852. [PubMed: 21280968]

Lareau, A. Unequal childhoods: Class, race and family life. Berkeley, CA: University of California Press; 2003.

Martens A, Johns M, Greenberg J, Schimel J. Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. Journal of Experimental Social Psychology. 2006; 42:236–243.

McQueen A, Klein WP. Experimental manipulations of self-affirmation: A systematic review. Self And Identity. 2006; 5:289–354.

Mervis J. Transformation is possible if a university really cares. Science. 2013; 340:292–296. [PubMed: 23599476]

Miyake A, Kost-Smith LE, Finkelstein ND, Pollock SJ, Cohen GL, Ito TA. Reducing the gender achievement gap in college science: A classroom study of values affirmation. Science. 2010; 330:1234–1237. [PubMed: 21109670]

Momsen JL, Long TM, Wyse SA, Ebert-May D. Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. CBE—Life Sciences Education. 2010; 9:435–440. [PubMed: 21123690]

Ostrove JM, Long SM. Social class and belonging: Implications for college adjustment. The Review of Higher Education. 2007; 30:363–389.

Pascarella, ET.; Terenzini, PT. How college affects students: Findings and insights from twenty years of research. San Francisco, CA: Jossey-Bass; 1991.

Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behavior Research Methods, Instruments, and Computers. 2004; 36:717–731.

Ramey, G.; Ramey, A. Brookings Papers on Economic Activity. Spring; 2010. The Rug Rat race; p. 129-176.

Reardon, SF. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In: Murnane, R.; Duncan, G., editors. Whither opportunity? Rising inequality and the uncertain life chances of low-income children. New York: Russell Sage Foundation Press; 2011.

Rubin, D. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.

Ruiz-Primo MA, Briggs D, Iverson H, Talbot R, Shepard LA. Impact of undergraduate science course innovations on learning. Science. 2011; 331:1269–1270. [PubMed: 21393529]

Saenz, VB.; Hurtado, S.; Barrera, D.; Wolf, D.; Yeung, F. First in my family: A profile of first-generation college students at four-year institutions since 1971. Los Angeles, CA: Higher Education Research Institute; 2007. http://www.heri.ucla.edu/PDFs/pubs/TFS/Special/Monographs/FirstInMyFamily.pdf

Schulz RM. Reforming science education. Part I. The search for a philosophy of science education. Science & Education. 2009; 18:225–249.

Sherman DK, Bunyan DP, Creswell JD, Jaremka LM. Psychological vulnerability and stress: The effects of self-affirmation on sympathetic nervous system responses to naturalistic stressors. Health Psychology. 2009; 28:554–562. [PubMed: 19751081]

Sherman, DK.; Cohen, GL. The psychology of self-defense: Self-affirmation theory. In: Zanna, MP., editor. Advances in experimental social psychology. Vol. 38. San Diego, CA: Academic Press; 2006. p. 183-242.

Sherman DK, Hartson KA, Binning K, Purdie-Vaughns V, Garcia J, Taborsky-Barba S, Tomassetti S, Nussbaum AD, Cohen G. Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. Journal of Personality and Social Psychology. 2013; 104:591–618. [PubMed: 23397969]

Sherman DK, Nelson LD, Steele CM. Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. Personality and Social Psychology Bulletin. 2000; 26:1046–1058.

Shnabel N, Purdie-Vaughns V, Cook JE, Garcia J, Cohen GL. Demystifying values-affirmation interventions: Writing about social belonging is a key to buffering against identity threat. Personality and Social Psychology Bulletin. 2013

Sirin SR. Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research. 2005; 75:417–453.

Smart-Richman L, Leary MR. Reactions to discrimination, stigmatization, ostracism, and other forms of interpersonal rejection: A multimotive model. Psychological Review. 2009; 116:365–383. [PubMed: 19348546]

Snibbe AC, Markus HR. You can't always get what you want: Educational attainment, agency, and choice. Journal of Personality and Social Psychology. 2005; 88:703–720. [PubMed: 15796669]

Spencer B, Castano E. Social class is dead. Long live social class! Stereotype threat among low socioeconomic status individuals. Social Justice Research. 2007; 20:418–432.

Steele CM. The psychology of self-affirmation: Sustaining the integrity of the self. Advances in Experimental Social Psychology. 1988; 21:261–302.

Steele CM. A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist. 1997; 52:613–629. [PubMed: 9174398]

Steele CM, Aronson J. Stereotype threat and the intellectual test performance of African Americans. Journal Of Personality And Social Psychology. 1995; 69:797–811. [PubMed: 7473032]

Steele, CM.; Spencer, SJ.; Aronson, J. Contending with group image: The psychology of stereotype and social identity threat. In: Zanna, MP., editor. Advances in experimental social psychology. Vol. Vol. 34. San Diego, CA US: Academic Press; 2002. p. 379-440.

Stephens NM, Fryberg SA, Markus HR, Johnson CS, Covarrubias R. Unseen disadvantage: how American universities' focus on independence undermines the academic performance of first-generation college students. Journal of Personality and Social Psychology. 2012; 102:1178–1197. [PubMed: 22390227]

Stephens NM, Markus H, Fryberg SA. Social class disparities in health and education: Reducing inequality by applying a sociocultural self model of behavior. Psychological Review. 2012; 119:723–744. [PubMed: 23088339]

Stephens NM, Townsend SM, Markus H, Phillips L. A cultural mismatch: Independent cultural norms produce greater increases in cortisol and more negative emotions among first-generation college students. Journal of Experimental Social Psychology. 2012; 48:1389–1393.

Terenzini PT, Springer L, Yeager PM, Pascarella ET, Nora A. First-generation college students: Characteristics, experiences, and cognitive development. Research in Higher Education. 1996; 37:1–22.

Trent W, Orr MT, Ranis S, Holdaway J. Transitions to college: Lessons from the disciplines. Teachers College Record. 2007; 109:2207–2221. http://www.tcrecord.org/content.asp?contentid=12594.

Walton, GM.; Carr, PB. Social belonging and the motivation and intellectual achievement of negatively stereotyped students. In: Inzlicht, M.; Schmader, T., editors. Stereotype Threat: Theory, Processes, and Application. New York: Oxford University Press; 2012.

Walton GM, Cohen GL. A brief social-belonging intervention improves academic and health outcomes of minority students. Science. 2011; 331:1447–1451. [PubMed: 21415354]

Walton GM, Spencer SJ. Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. Psychological Science. 2009; 20:1132–1139. [PubMed: 19656335]

Warburton, E.; Bugarin, R.; Nunez, A. Bridging the gap: Academic preparation and postsecondary success of first-generation students (Report No. NCES 2001–153). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office; 2001. http://nces.ed.gov/pubs2001/2001153.pdf

Wilson, TD. Redirect: The surprising new science of psychological change. New York: Little, Brown; 2011.

Yeager DS, Walton GM. Social-psychological interventions in education: They're not magic. Review of Educational Psychology. 2011; 81:267–301.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

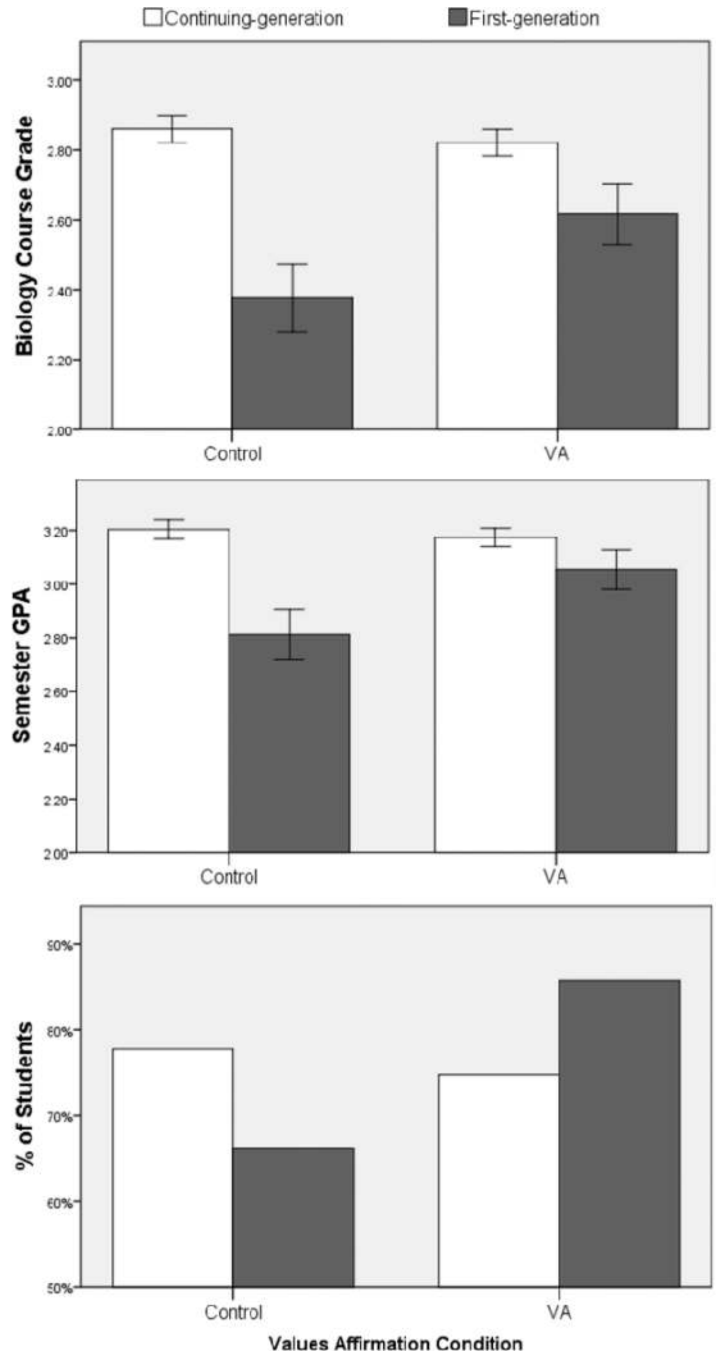

Figure 1.

Performance in the course, semester GPA, and percentage of students who enrolled in the second semester of biology as a function of generational status (CG versus FG) and treatment condition (control versus VA). Error bars represent +/− 1 standard error.

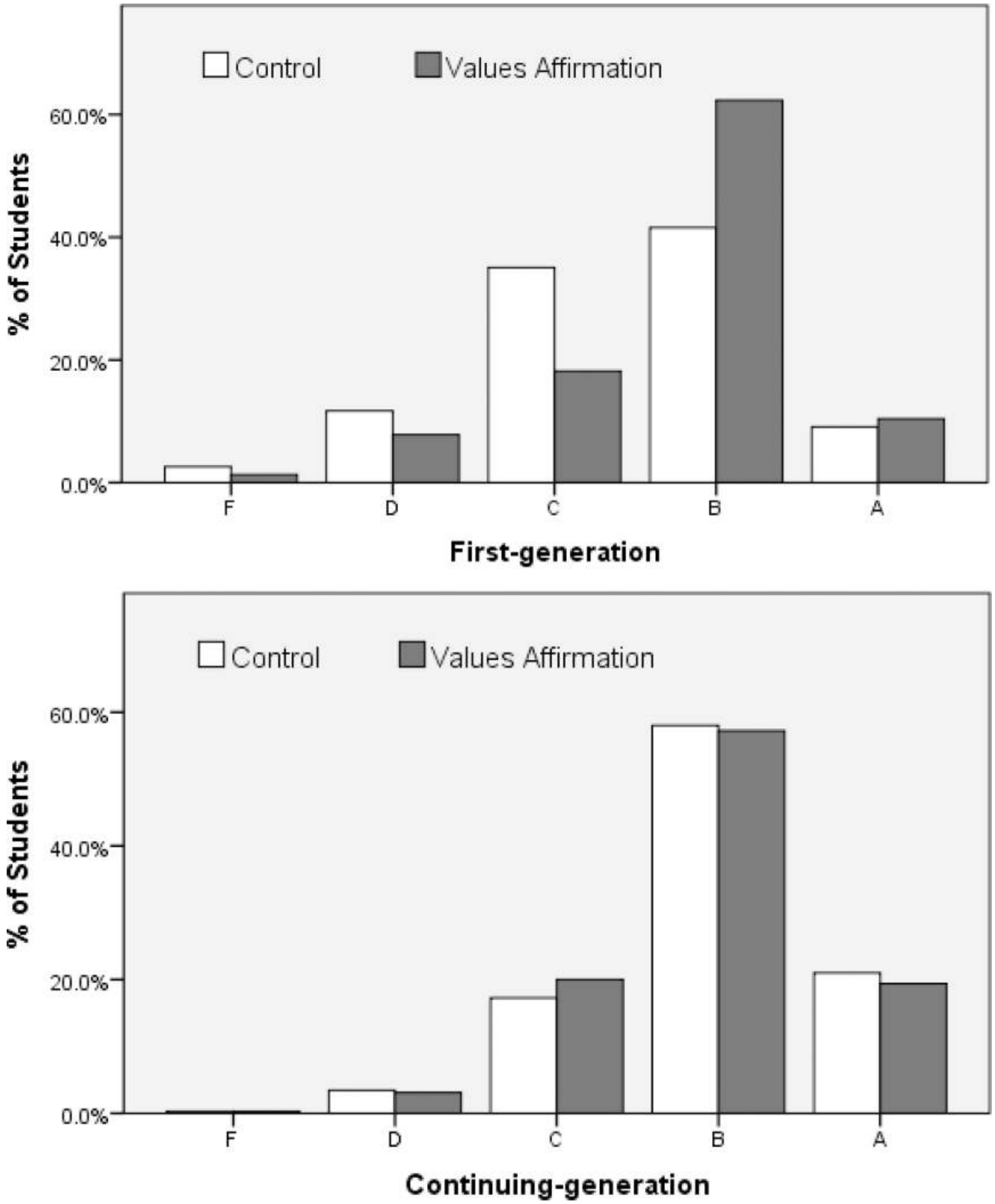

Figure 2.

Percentage of students receiving each letter grade (A, B, C, D, and F) as a function of generational status (CG versus FG) and treatment condition (VA versus control).

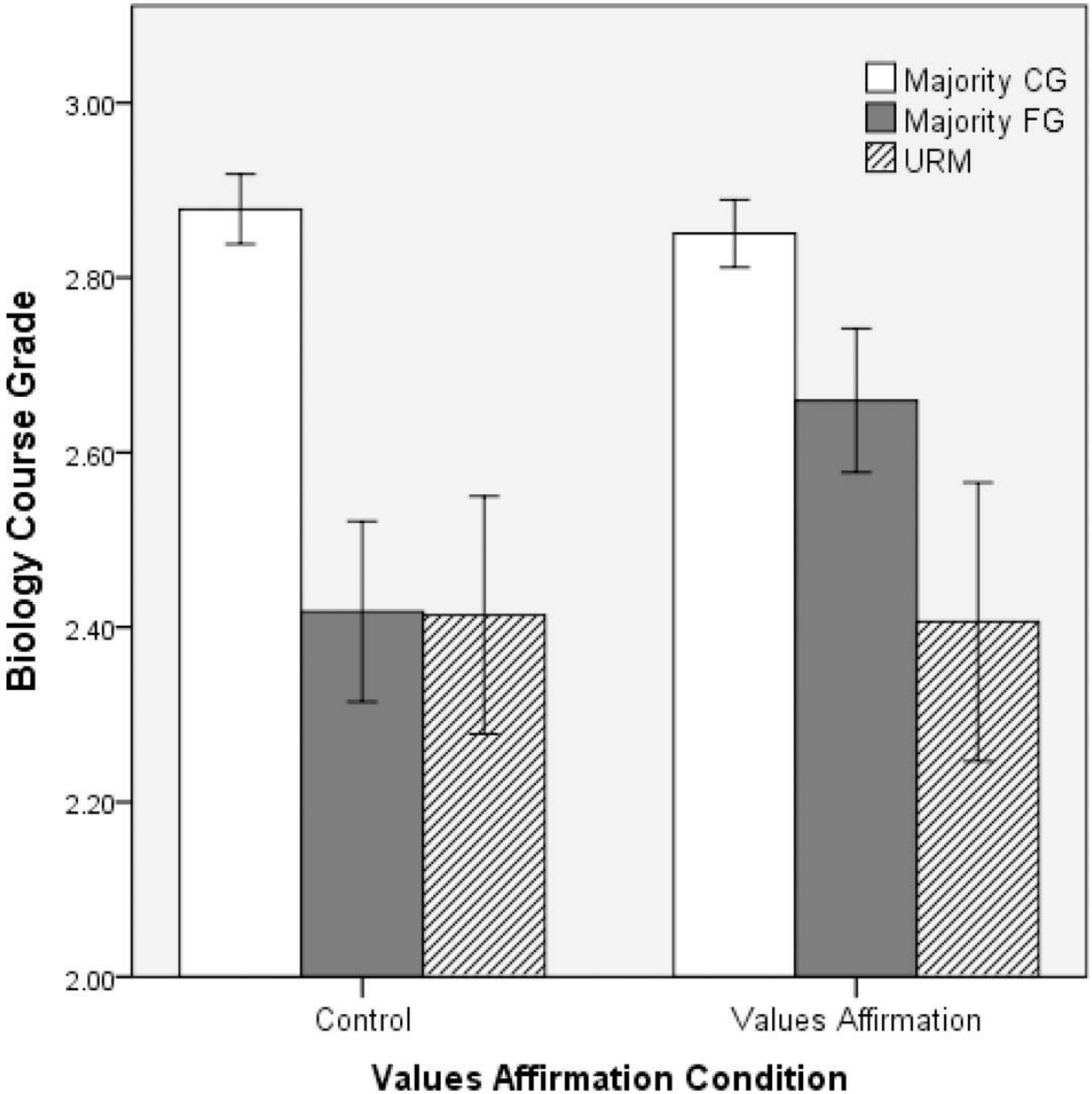

Figure 3.
Student performance in the course as a function of URM and generational status (majority continuing-generation and majority first-generation) and treatment condition (values affirmation versus control). Error bars represent +/− 1 standard error.

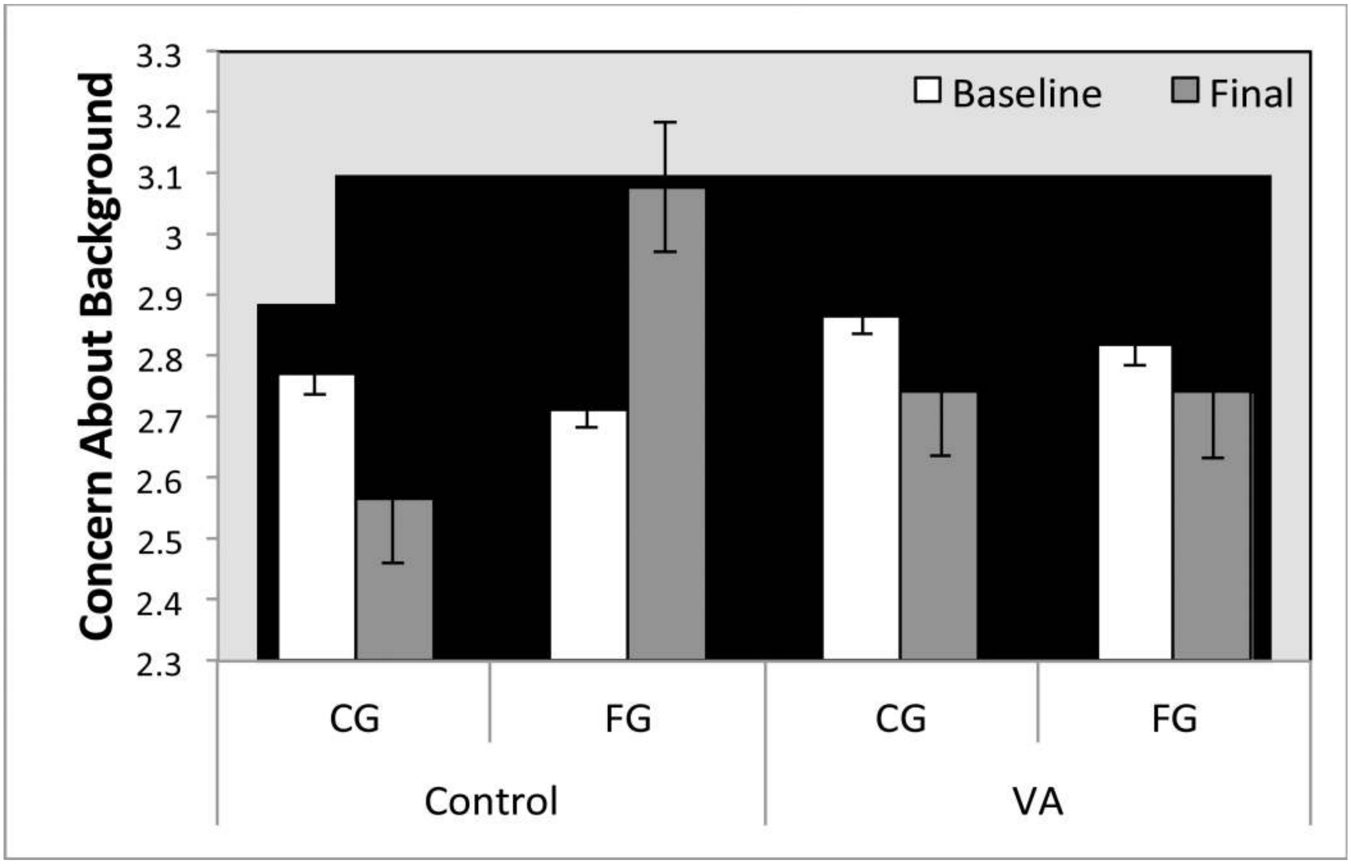

Figure 4.

Changes in concern about background over the semester. Baseline scores were collected in the 2nd week of the semester and final scores were collected in the 14th week. High scores represent higher levels of concern about background.

Table 1

Baseline comparisons between experimental conditions

	Control	VA	
Measure	***M (SD)***	***M (SD)***	**N**
Age	19.26 (1.15)	19.27 (1.23)	798
ACT	28.46 (2.77)	28.55 (2.84)	711
Prior Spring GPA	3.23 (.50)	3.20 (.50)	696
Cumulative Credits	46.43 (16.95)	46.78 (15.17)	696
Confidence about Performance	5.77 (.87)	5.71 (.97)	798
Concern about Background	2.76 (1.53)	2.86 (1.59)	796
Year in School	%	%	
Freshmen	13.1	12.9	
Sophomores	79.8	81.6	
Juniors	5.1	3.7	
Seniors	2.0	1.7	

Note: VA = Values Affirmation condition. There were no significant differences between experimental conditions on any of the baseline measures. The distribution of students' year in school (freshmen, sophomores, etc.) did not differ by condition, $\chi 2(3, N=798)=.96$, $p=.81$.

Table 2

Means, standard deviations, and intercorrelations for all variables

Measure	*M (SD)*	1	2	3	4	5	6	7	8
1. Baseline Performance	--	1							
2. Course Grade	2.77 (0.73)	.61*	1						
3. Semester Grade	3.14 (0.65)	.60*	.66*	1					
4. Continuation	76.2%	.19*	.28*	.20*	1				
5. Confidence (base)	5.74 (0.92)	.18*	.22*	.17*	.12*	1			
6. Confidence (final)	5.05 (1.29)	.28*	.54*	.29*	.19*	.43*	1		
7. Concern about Background (base)	2.81 (1.56)	−.06	−.13*	−.04	−.07	−.39*	−.30*	1	
8. Concern about Background (final)	2.71 (1.62)	−.08	−.29*	−.09	−.12*	−.26*	−.46*	.45*	1

Notes: Baseline performance is a z-scored composite measure (see Footnote 5)

Continuation is a measure reflecting the percentage of students who continued into the second semester of the course

* $p<0.01$

Table 3

Means and standard deviations for outcome variables as a function of generational status and treatment condition

	Continuing-generation		First-generation	
	Control	**VA**	**Control**	**VA**
Course Grade	2.86 (.69)	2.82 (.69)	2.38 (.85)	2.62 (.78)
Semester GPA	3.20 (.63)	3.17 (.62)	2.81 (.81)	3.05 (.64)
Continuation	77.7%	74.8%	66.2%	85.7%
N	319	325	77	77

Note: VA = Values Affirmation condition. The Continuation measure reflects the percentage of students who enrolled in the second semester of the biology class the next semester.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 4

Mediation Analysis

Independent Variable (IV)	**Mediating Variable (M)**	**Dependent Variable (DV)**	**Effect of IV on M (a path)**	**Direct Effect of M on DV (b path)**	**Total Effect of IV on DV (c path)**	**Direct Effect of IV on DV (c′ path)**
Generational Status X Condition	Course Grade	Continuation to Next Semester	.07* (.03)	.06* (.03)	.32** (.11)	.29* (.12)

Note. Numbers represent unstandardized coefficients and numbers inside parentheses represent standard errors. Regression analyses include gender and lecture codes as covariates.

* $p \leq .05$;

** $p \leq .01$;

*** $p \leq .001$.

Table 5

Values Selected by Students in the Values Affirmation Condition

Values Selected	First-generation	Continuing-generation
Relationships with Friends or Family	88.20%	84.00%
Learning and Gaining Knowledge	63.20%	57.20%
Sense of Humor	32.90%	31.10%
Career	31.60%	29.20%
Independence	25.00%	22.60%
Spiritual or Religious Values	17.10%	17.60%
Creativity	17.10%	11.90%
Athleticism	3.90%	8.20%
Music	3.90%	11.00%
Belonging to a Social Group	3.90%	6.60%
Government or Politics	1.30%	0.60%
Being good at Art	0.00%	0.60%

Note: Numbers indicates the percentage of students who selected each value to write about; Students could select two or three values.

Table 6

Percentage of Interdependent and Independent Items Endorsed by First-Generation and Continuing-Generation Students

Survey items	First-generation	Continuing-generation
Interdependent items		
Help my family out after I'm done with college *	61.0%	46.3%
Be a role model for people in my community	57.2	53.5
Show that people with my background can do well *	44.0	26.4
Give back to my community	64.8	66.2
Provide a better life for my own children	80.5	73.2
Scale Mean **	**3.08**	**2.66**
Independent items		
Expand my knowledge of the world	72.3%	79.4%
Become an independent thinker	66.0	69.5
Explore new interests *	62.9	74.9
Learn more about my interests *	60.4	69.5
Expand my understanding of the world	66.0	73.4
Scale Mean **	**3.28**	**3.67**

Note.

* $p<.05$, based on chi-square tests, χ^2 (1, N=772), comparing first-generation and continuing generation students.

** Scale means differ significantly, $p<.01$.

Table 7

Means and Intercorrelations of Measures in Survey Study

	Continuing-Generation	First-Generation						
Measure	***M (SD)***	***M (SD)***	**1**	**2**	**3**	**4**	**5**	**6**
1. Independent Motives	3.67 (1.49)	3.28 (1.66)	1.00					
2. Interdependent Motives	2.66 (1.54)	3.08 (1.46)	.31*	1.00				
3. Academic Belonging	5.78 (0.97)	5.60 (1.07)	.13*	.03	1.00			
4. Belonging Uncertainty	3.12 (1.69)	3.69 (1.83)	−.08	.04	−.45*	1.00		
5. Academic and Social Concerns	3.35 (1.15)	3.65 (1.27)	−.09	−.05	−.36*	.58*	1.00	
6. Confidence about Performance	5.37 (1.20)	5.28 (1.24)	.14*	.07	.56*	−.33*	−.28*	1.00

Note:

* $p < 0.01$