



Supporting Online Material for

A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students

Gregory M. Walton* and Geoffrey L. Cohen

*To whom correspondence should be addressed. E-mail: gwalton@stanford.edu

Published 18 March 2011, *Science* **331**, 1447 (2011)
DOI: 10.1126/science.1198364

This PDF file includes:

Materials and Methods
Tables S1 to S3
References

Supporting Online Material

Materials and Methods

Overview of Procedure

Preintervention Survey (Second Semester of Students' Freshman Year)

Several weeks before being recruited to participate in the main study, students were asked to complete a survey by email or online containing preintervention measures. To prevent contamination and to avoid signaling our interest in race in the main study, this survey and the main study were administered by different experimenters and the two were not linked for participants. The survey assessed (1) academic identification (2 items, e.g., “How important is academic success to you?” $I=not\ at\ all\ important, 7=essential\ to\ who\ I\ am; \alpha=0.90$) (S1), (2) self-reported achievement behavior (6 items, e.g., “On the average day, how many hours do you spend studying outside of class?”; $\alpha=0.68$) (S1), (3) sensitivity to race-based rejection (12 scenarios, $\alpha=0.89$) (S2), (4) stigma consciousness adopted for race (10 items, e.g., “Most people have a problem viewing members of my racial group as equals” $I=strongly\ disagree, 7=strongly\ agree; \alpha=0.82$) (S3), (5) perceptions of racial prejudice (3 items, e.g., “If two [school name] students turn in an essay of generally adequate quality for a class—one a Black student, one a White student—who if anyone is more likely to get a better grade?” $I=Black\ student\ more\ likely\ to\ get\ better\ grade, 7=White\ student\ more\ likely\ to\ get\ better\ grade; \alpha=0.70$), (6) stereotype threat (4 items, e.g., “At [school name] I worry that people will draw conclusions about my racial group based on my performance” $I=strongly\ disagree, 7=strongly\ agree; \alpha=0.95$) (S4), and (7) racial identification (2 items adopted from the Race Centrality subscale of the Multidimensional Inventory of Black Identity, e.g., “My racial identity is an important part of who I am” $I=strongly\ disagree, 7=strongly\ agree; \alpha=0.88$) (S5).

Delivery of Intervention (Second Semester of Students' Freshman Year)

The intervention was delivered individually to students in a psychology laboratory. Students were told that the study investigated “students’ college experiences and attitudes” and that the purpose of the study was twofold: (1) “to better understand your personal experiences and attitudes here at [school name]” and (2) “to help us provide incoming [school name] students next year and in the years to come with more accurate expectations about what college is like.” Students in both conditions were told that the researchers had previously conducted a survey of upper-year students and that they would be asked for their help in interpreting the results of this survey. The survey, students were told, had been conducted with a random sample of upper-year students, “so the views...are representative of the upperclassmen student body...as a whole.” In addition, students were told, the “results were consistent across different demographic groups...[including] class year, race, [and] gender.” In both conditions students were then given survey results to review. The results included a “quantitative summary” and 9 “illustrative...free-response reports” attributed to ethnically diverse upper-year students (5 European Americans, 2 African Americans, 1 Asian American, and 1 Hispanic American).

In the treatment condition the survey results emphasized that upper-year students of all ethnicities worried about their social belonging at first in college but that these concerns dissipated with time and that eventually almost all students came to feel at home. For instance, the “quantitative summary” indicated that most upper-year students had worried during their first year about “whether other students would accept them” and had “felt intimidated” by professors but that, over time, most students came to feel “‘confident’ or ‘certain’ that most other students

and professors “accept them.” The 9 quotations, based on an actual survey we conducted, reinforced this theme. Examples include:

“When I first got to [school name], I worried that I was different from other [students at] [school name]. Everyone else seemed so certain that they were right for [school name], I wasn’t sure I fit in. Sometime after my first year, I came to realize that many people come to [school name] uncertain whether they fit in or not. Now it seems ironic – everybody feels they are different freshman year from everybody else, when really in at least some ways we are all pretty similar. Since I realized that, my experience at [school name] has been almost one-hundred percent positive.”

- Participant #17, [dormitory] senior, African American female

“I didn’t go to a very good high school, and I worried that my high school courses had not prepared me well for college. Honestly, when I got here, I thought professors were scary. I thought they were critical and hard in their grading, and I worried a lot about how they and other students would evaluate me. I was nervous about speaking in class and I didn’t like other people to read my papers. Around my sophomore year I felt more comfortable – I began to enjoy my classes more and I found some close friends who I trusted. I also became more comfortable speaking in class, and sometimes I asked my friends to edit my papers for me. And I saw that even when professors are critical, or their grading harsh, it didn’t mean they looked down on me or that I didn’t belong. It was just their way of motivating high-achieving [school name] students.”

- Participant #19, [dormitory] junior, White male

“The most difficult transition from high school to [school name] was coming from a situation in which I knew every student for the past seven years into a new situation in which I did not know one student before I arrived. Freshman year even though I met large numbers of people, I didn’t have a small group of close friends. I had to work to find lab partners and people to be in study groups with. I was pretty homesick, and I had to remind myself that making close friends takes time. Since then in classes, clubs, and social activities, I have met people some of whom are now just as close as my friends in high school were.”

- Participant #84, [dormitory] junior, Asian American male

These materials were based on an actual survey of upper-year students but were purified to clarify the treatment message. Past research provides a discussion of ethical issues raised by this procedure (*S1*).

Three randomized control conditions were used. In each, similar normative information was presented but this information was irrelevant to issues of belonging. In Cohort 1, the information indicated that students’ social and political attitudes became more sophisticated over time in college (*S1*). In Cohort 2, for half of control participants it indicated that students became more familiar with the physical environment in college over time (i.e., rather than the social environment); for the other half it indicated that students experienced academic difficulties at first in college (rather than social difficulties) but that these lessened with time (*S6*). In few

analyses did the two control conditions in Cohort 2 differ for either racial group (i.e., in no more than would be expected by chance alone, including no primary analysis), so these conditions were combined.

After reading the survey results, students were asked to engage in a series of activities designed to help them internalize the treatment message (S7, S8). First, students were asked, “to write an essay about why you think people’s experience in college changes in the way the Junior/Senior survey describes.” They were urged to illustrate their essay “with examples from your own experience” and were invited to look back on the survey as they worked. In addition, students were told, “we plan to take excerpts of what people write here and show them to students coming to [school name] next year or in subsequent years, so they know what their experience is likely to be like.” Students were further told, “I am sure the students who read about your experiences will appreciate the effort that you put in.” Students were given approximately 30 minutes to write their essays.

After doing so, students were told that, in addition to distributing written essays, the researchers hoped to create videos “based on the essays written by students in this study” to be shown to future students. “[W]e would like to videotape you reading your essay to the camera... [W]e think it is important to do this because we believe it will be particularly effective for [future] students if they feel as if an older student is speaking directly to them about their experiences. As you probably know, it can be difficult to come into a new situation not knowing what to expect and you, as an older student who has just gone through the same experience, are in a great position to help these freshmen out.” Most students agreed to deliver the speech (80.43% did) and this did not vary by race or condition [$\chi^2(1, N=92)s<1$]. Students who agreed were given time to edit their essay. They were then taken to another room where a large video-camera had been set up and were filmed delivering their speech.

This completed the delivery of the intervention. In total, the intervention lasted approximately 1 hour. At the end of the laboratory session, students reported demographic information (e.g., race/ethnicity, gender) and their SAT-Math and -Verbal scores.

Overview of Dependent Measures

Daily Diary Surveys (Second Semester of Students’ Freshman Year)

In the seven days immediately following the intervention, participants completed surveys by email or online each afternoon and each evening assessing their current sense of social belonging on campus and the degree of adversity they had experienced each day. These measures are described in a latter section.

End-of-College Survey (Second Semester of Students’ Senior Year)

At the end of their college tenure, students completed an end-of-college survey either in-person or by email. The primary experimenter who administered this survey differed from both the experimenter who had conducted the preintervention survey and the experimenter who had delivered the intervention and administered the daily diary surveys. Students were told that the survey followed-up on a study they had taken part in during their freshman year but were given no additional information about the study. The measures assessed are described below.

College Academic Records

After completing the end-of-college survey, students were asked to authorize the release of their complete college academic records, which were used to calculate their grade-point-average (GPA). These data were obtained from the university registrar.

Study Setting

The study took place at a selective university in which African American students were a numeric minority. The selectivity of the school is reflected in the high test scores of study participants. The SAT-Math and SAT-Verbal scores of African American study participants averaged 687 and 713, respectively; those of European Americans averaged 757 and 743, respectively. African American students represented between 5% and 15% of the total undergraduate student body at the time of the study.

Although this student body is selective, large gaps in academic achievement between European Americans and African Americans arise even among students at elite institutions (*S9*, *S10*) and large gaps in health exist between these groups among even people who are high in socioeconomic status (*S11*). Such gaps have attracted the attention of much scholarship (*S9*, *S12*, *S13*). In some ways, their existence even among the “vanguard” of a minority group is especially troubling, and thus an important focus of intervention.

Participants, Recruitment, and Retention

Participant Sample

A total of 49 African American students (34 female) and 43 European American students (24 female) participated and were randomly assigned to treatment or control condition. Cohort 1 included 37 students (18 African Americans). Cohort 2 included 55 students (31 African Americans). The campus-wide control group included all African American and European American students in the same class years as study participants but who had not taken part in the study. This group contained 194 African American students (104 female) and 1362 European American students (603 female).

Sampling Procedure in Cohorts 1 and 2

Students in Cohort 1 were recruited using a convenience sampling procedure. Students in Cohort 2 were recruited using random-sampling. Thirty-six African American and 31 European American students were randomly selected and targeted for recruitment from a list of all first-year African American and European American students. Of the targeted students, 82.09% participated (86.11% of African Americans, 77.42% of European Americans). There was no difference in preintervention GPA between study participants and students in the campus-wide control group within either racial group in either Cohort 1 [$t_s < 1.20$, $P_s > 0.20$] or Cohort 2 [$t_s < 1$]. The sample in Cohort 2 in particular is thus likely to be representative of same-race students campus-wide.

Tests of Random Assignment to Condition

A series of analyses tested the success of random assignment to condition. The results suggest that random assignment succeeded.

Specifically, analyses tested for differences between treatment and control conditions overall and for each racial group on all preintervention measures. No difference was significant (Table S1). Additionally, multivariate analyses of variance on (1) all preintervention academic measures (academic identification, achievement behavior, SAT-Math score, SAT-Verbal score,

and preintervention GPA), (2) all preintervention race-related measures (sensitivity to race-based rejection, stigma consciousness, perceptions of racial prejudice, stereotype threat, and racial identification), and (3) all 10 preintervention measures yielded no effect of condition and no race x condition interaction [all $F_s < 1.65$, $P_s > 0.16$]. In addition, men and women, and African Americans and European Americans, were equally likely to be assigned to the treatment and control conditions [$\chi^2(1, N=92) < 1$].

We also redid all these analyses focusing on the final sample (78.26% of the original sample). All measures but one remained nonsignificant [$t_s < 1.40$, $P_s > 0.15$; $F_s < 1.65$, $P_s > 0.16$]. The exception was academic identification for which there was a difference favoring the control condition overall [$t(70) = -2.00$, $P = 0.050$]. This effect was confined to European Americans [$t(68) = -2.01$, $P = 0.049$] with no difference for African Americans [$t < 1$]. Given the number of comparisons made, this effect would be expected by chance alone.

Retention Rates

Retention rates were adequate. At the end of their college careers, 78.26% of participating students responded to our communications and completed the end-of-college survey. Analyses of end-of-college survey measures are thus based on the 72 students who completed the survey (38 African Americans, 34 European Americans). Some measures were completed only by participants in Cohort 2; analyses of these measures are based on 43 students (23 African Americans).

There was no evidence of differential attrition. Survey completion rates did not vary by student race, student gender, or experimental condition [$\chi^2(1, N=92) < 1$] or by the race by condition interaction [$\Delta\chi^2(1, N=92) < 1$]. In addition, students who completed the survey did not differ on any preintervention measure from students who did not complete the survey (Table S2). Multivariate analyses of variance across (1) preintervention academic measures (academic identification, achievement behavior, SAT-Math score, SAT-Verbal score), (2) preintervention race-related measures (sensitivity to race-based rejection, stigma consciousness, perceptions of racial prejudice, stereotype threat, and racial identification), and (3) all preintervention measures yielded no effect of survey-completion and no race x survey-completion interaction [$F_s < 1$]. The results indicate that the sample of students who completed the study did not differ systematically from the original sample on any baseline-assessed measure.

After completing the end-of-college survey, students were asked to authorize the release of their complete academic records from college. Of students who completed the end-of-college survey, 97.22% agreed to release these records. Authorization rates did not vary by student race, student gender, or experimental condition overall [$\chi^2(1, N=92) < 1$] by condition for African Americans [$\chi^2(1, N=49) < 1$] or by the race by condition interaction [$\Delta\chi^2(1, N=92) < 1$]. Students who authorized the release of these records did not differ on any preintervention measure from students who did not [$t_s < 1.40$, $P_s > 0.17$]. Analyses of academic performance are thus based on 70 students (37 African Americans, 33 European Americans).

Analyses of Academic Performance

Calculation of Ethnic Differences in GPA

As European Americans' postintervention GPA did not differ by condition, European Americans were combined across conditions to calculate differences in GPA between European Americans and African Americans in each condition.

Analyses of Cumulative Postintervention GPA

Primary analyses of cumulative postintervention GPA involved multiple regressions with dummy-coded participant race (European American vs. African American), dummy-coded experimental condition (randomized control vs. social-belonging treatment), and the interaction term predicting (1) change in GPA—mean postintervention GPA minus preintervention GPA—with student gender controlled and (2) mean postintervention GPA with preintervention GPA and student gender controlled. As described in the main text, both analyses yielded a significant condition effect for African Americans [$B=0.30$, $t(65)=2.54$, $P=0.014$ and $B=0.24$, $t(64)=2.65$, $P=0.010$, respectively] and no effect for European Americans [$ts < 1$] [race x condition interactions: $B=-0.43$, $t(65)=-2.41$, $P=0.019$ and $B=-0.31$, $t(64)=-2.27$, $P=0.027$, respectively].

Grade data including the campus-wide sample was analyzed in a multiple regression with race, dummy-coded condition, and their interaction predicting residual postintervention GPA adjusted for preintervention GPA and student gender. Residual GPA was used rather than raw postintervention GPA with preintervention GPA and student gender controlled because an agreement with university officials that secured the release of the campus-wide data limits the form in which these data can be reported.

The results of this analysis are reported in the main text and depicted in Fig. 2A. As reported there, treated African Americans had higher residual postintervention GPA scores than did African Americans in both the randomized control group ($B=0.24$, $t(1620)=2.62$, $P=0.009$) and the campus-wide sample [$B=0.28$, $t(1620)=3.97$, $P=0.00008$]. The latter two groups did not differ [$t < 1$]. For European Americans, no pair-wise comparison was significant [$ts < 1$]. To calculate the race x condition interaction, we conducted a multiple regression on residual postintervention GPA using effect coding (treatment=2, randomized control=-1, campus-wide control=-1). The race x condition interaction was significant [$t(1622)=-3.08$, $P=0.002$].

Analysis of change in GPA from pre- to postintervention (i.e., difference scores) with student gender controlled yielded virtually identical results. Treated African Americans showed more improvement in GPA than did African Americans in both the randomized control group ($B=0.29$, $t(1619)=2.66$, $P=0.008$) and the campus-wide sample [$B=0.26$, $t(1619)=3.22$, $P=0.001$]. The latter two groups did not differ [$t < 1$]. For European Americans, no pair-wise comparison was significant [$ts \leq 1$]. Multiple regression using effect coding found that the race x condition interaction was significant [$t(1621)=-2.91$, $P=0.004$].

Finally, we also tested the condition effect in a multiple regression with raw postintervention GPA as the outcome and preintervention GPA and student gender included as covariates. Replicating the above results, treated African Americans had higher postintervention GPA scores than did African Americans in both the randomized control condition [$B=0.24$, $t(1625)=2.56$, $P=0.011$] and the campus-wide sample [$B=0.27$, $t(1625)=3.91$, $P=0.0001$]. The latter two groups did not differ [$t < 1$]. For European American no pair-wise comparison was significant [$ts < 1$]. Multiple regression using effect coding found that the race x condition interaction was significant [$t(1620)=-2.99$, $P=0.003$].

Analysis of the Percentage of Students in the Top and Bottom 25% of their College Class

Figs. 2B and 2C report the percentage of students in each racial group and condition whose GPA from sophomore-through-senior year fell in the top and bottom 25% of their college class in terms of (1) residual postintervention GPA adjusted for student gender and preintervention GPA (Fig. 2B) and (2) raw postintervention GPA (Fig. 2C). Chi-square analyses tested the condition effect within each racial group along three comparisons: (1) the campus-

wide control condition vs. the randomized control condition vs. the treatment condition; (2) the campus-wide and randomized control conditions (combined) vs. the treatment condition; and (3) the randomized control condition vs. the treatment condition. The third analysis provides a fully experimental test of the treatment effect but has the least statistical power. In no analysis did the campus-wide and randomized control conditions differ for either European Americans [$\chi^2(1, N=1382) < 1$] or African Americans [$\chi^2(1, N=213) < 1.65, P > 0.20$].

For European American students, for all four outcomes all three analyses were nonsignificant [$\chi^2(2, N=1395) < 1, \chi^2(1, N=1395) < 1, \text{ and } \chi^2(1, N=33) < 1$, respectively].

For African American students, for the percentage of students in the top 25% of the class in residual postintervention GPA, the condition difference was significant (1) comparing across all three conditions [$\chi^2(2, N=231) = 12.69, P = 0.002$], (2) comparing the two control conditions to the treatment condition [$\chi^2(1, N=231) = 12.69, P = 0.0004$], and (3) comparing the randomized control condition to the treatment condition [$\chi^2(1, N=37) = 4.94, P = 0.026$]. For the percentage in the bottom 25% of the class in residual postintervention GPA, all three analyses were significant [$\chi^2(2, N=231) = 8.71, P = 0.013, \chi^2(1, N=231) = 7.10, P = 0.008, \text{ and } \chi^2(1, N=37) = 8.29, P = 0.004$, respectively].

For the percentage in the top 25% of the class in raw postintervention GPA, the first and second analyses were significant [$\chi^2(2, N=231) = 8.97, P = 0.011 \text{ and } \chi^2(1, N=231) = 8.96, P = 0.003$, respectively]; the third was a trend [$\chi^2(1, N=37) = 2.27, P = 0.13$]. For the percentage in the bottom 25% of the class in raw postintervention GPA, the first and second analyses were marginally significant [$\chi^2(2, N=231) = 4.73, P = 0.094 \text{ and } \chi^2(1, N=231) = 3.25, P = 0.071$, respectively]; the third was nonsignificant [$\chi^2(1, N=37) < 1$].

Daily Diary Measures and Mediation of the Belonging-Treatment Effect on 3-Year GPA

Daily Diary Measures

In the week after the intervention, participants completed two surveys by email or online each day, one each afternoon and one each evening. First, both surveys assessed students' current sense of social belonging on campus (*SI*). This was a composite of three scales: students' (a) current sense of social fit on campus (17-items, e.g., "Right now, I feel like people at [school name] like me", "Right now, I feel like I belong at [school name]"; *1=strongly disagree, 7=strongly agree*), (b) current level of self-efficacy (2-items, e.g., "Right now, I feel confident that I have the ability to do well at [school name]"; *1=strongly disagree, 7=strongly agree*), and (c) current sense of potential to succeed in college (1-item, "Right now, how much potential, compared with other [school name] students, do you feel you have to succeed at [school name]"; *10%=more potential than 10% of students, 90%=more potential than 90% of students*). Each measure was standardized and then averaged.

Second, the evening survey assessed the level of adversity students reported having experienced that day (*SI*). Participants (a) evaluated how positive or negative their day was overall (*1=very negative, 10=very positive*) and (b) listed positive and negative events they had experienced that day and rated the positivity or negativity of each event on the same 10-point scale. As anticipated, most of the adversity students described involved social events (e.g., feeling excluded by peers) or events that could carry a social meaning (e.g., struggling with an assignment) (*SI*). For each day and participant, we calculated an overall measure of the positivity of the events listed that day, standardized both this measure and the overall day assessment measure, averaged the two, and reversed-scored the measure so that higher values reflect greater levels of adversity experienced by a given student on a given day. As predicted, there was no

effect of race and no effect of condition for either racial group on the mean level of adversity students experienced over the week [$t < 1$]. Instead, as described below, what varied by condition was students' response to adversity.

Calculation of Within-Participant Association Between Daily Adversity and Daily Sense of Social Belonging

For each participant, we calculated the correlation between the level of adversity they reported having experienced each day and their composite sense of social belonging reported (a) that evening and (b) the subsequent afternoon. We transformed each correlation into a Fisher's Z -score. The two Fisher's Z -scores correlated [$R = 0.30$, $P = 0.012$], so we averaged them into a composite *contingency index*, which represents the degree to which daily experiences of adversity predicted students' subsequent daily sense of social belonging. This procedure allowed us to calculate the key conceptual variable in two ways—the first indexes the concurrent relationship between retrospectively estimated adversity (i.e., experienced earlier that day) and currently reported belonging; the second, a longitudinal relationship between adversity reported on a given day and belonging reported the next afternoon. Averaging these measures improves the reliability of the assessment. (As reported below, there was a significant condition effect on this composite measure for African Americans; this effect was also significant for each association separately [$t > 2.15$, $P < 0.035$].)

Effect of Condition on Contingency of Social Belonging

We conducted a regression on the composite contingency index. Given our interest in this index as a potential mediator of the condition effect on long-term GPA, this analysis, the correlation reported above, and the analyses reported below are restricted to participants for whom long-term grade data were available. Following the procedures described in the next section for identifying covariates, covariates were SAT-score and preintervention level of self-reported achievement behavior and perceptions of racial prejudice. The total number of daily adversity assessments participants completed was also included as it proved significant. Following analysis, the mean Fisher's Z -scores were transformed back into correlations to index the average within-cell correlation between daily adversity and students' daily sense of social belonging.

The analysis yielded a race \times condition interaction [$B = -0.49$, $t(59) = -2.04$, $P = 0.046$]. For African Americans, the condition effect was significant [treatment: mean within-subjects $R = 0.01$, Fisher's Z -score = 0.01; control: mean within-subjects $R = -0.45$, Fisher's Z -score = -0.49] [$B = 0.49$, $t(59) = 2.99$, $P = 0.004$]. For European Americans, it was nonsignificant [treatment: mean within-subjects R and Fisher's Z -score = -0.09; control: mean within-subjects R and Fisher's Z -score = -0.09] [$t < 1$].

Test of Mediation

Next we examined the relationship between contingency (adjusted for the aforementioned covariates) and change in GPA from preintervention (fall term, first-year) to postintervention (sophomore-through-senior year) adjusted for gender. The difference score was used rather than residual postintervention GPA controlling for preintervention GPA and gender to avoid colinearity between the hypothesized mediator and the baseline performance measure; however, analysis using the later variable yields similar results. Less contingency—a weaker inverse association between adversity and belonging—was unrelated to change in grades for European

American students [$R=-0.20$, $P=0.28$]. But it predicted better grades for African American students [$R=0.51$, $P=0.001$]. The race difference in this association was significant [$t(63)=-2.65$, $P=0.010$]. The results suggest that without intervention, African Americans contended with two detrimental processes that European American students did not. Their sense of academic belonging was more environmentally contingent, rising and falling with day-to-day experience. Moreover, this contingency posed a greater liability for African Americans, predicting worse performance.

Finally, we tested for moderated mediation (*SI4*). We regressed residual change in GPA adjusted for gender on race, condition, the race x condition interaction, contingency, and the race x contingency interaction. In this analysis, the race x condition interaction was no longer significant [$t(61)=-1.38$, $P=0.17$, $B=-0.25$, $SE=0.18$] but the race x contingency interaction was significant [$t(61)=2.38$, $P=0.020$, $B=0.26$, $SE=0.11$]. A supplementary analysis restricted to African Americans tested the simultaneous effects of condition and contingency on residual change in GPA. The condition effect was reduced to nonsignificance [$t(33)=1.01$, $P=0.32$] while contingency remained significant [$t(33)=2.79$, $P=0.009$]. The reduction in the significance of the condition effect was significant [asymmetric distribution of products test 95% confidence interval: 0.23-0.07, $P<0.05$]. These analyses suggest that the 3-year gain in GPA for African American students caused by the treatment was mediated by a reduction in the extent to which African American students' sense of social belonging on campus fluctuated with daily events.

This mediating process proved precise. Correlational analyses found that neither the mean level of adversity African Americans experienced in the first week postintervention nor the mean level of belonging they reported that week predicted improvement in GPA [$R_s=0.03$ and 0.19 , respectively, $P_s>0.25$]. In a multiple regression including all three variables, only a weaker inverse association between adversity and belonging predicted improvement in GPA [$\beta=0.51$, $P=0.002$; mean level of adversity: $\beta=0.10$, $P=0.51$; mean level of belonging: $\beta=0.19$, $P=0.23$].

End-of-College Survey: Measures, Covariates, and Supplementary Analyses

Overview, Covariates, and Measures Assessed in Each Cohort.

The end-of-college survey assessed psychological constructs, health, well-being, and recall of and beliefs about the intervention. Measures are described below. Following standard procedures (*SI5*) and past practice (*SI*), in each analysis only covariates that were predictive ($P_s \leq .15$) were retained. The tested covariates were: (1) the 7 measures assessed in the preintervention survey (described above), (2) student gender, and (3) SAT-score. Interactions between student gender and student race and experimental condition were also tested and included where significant (the 3-way interaction was not included in analyses of measures assessed only in Cohort 2, given the smaller sample size). Preintervention GPA was also tested as a covariate but was never predictive. Measures were analyzed in regression except where interactions involving student gender and either student race or condition were significant; here analyses of covariance (ANCOVAs) were used. Planned contrasts tested the effect of condition within each racial group. European American students did not differ by condition on any outcome [$t_s < 1.35$, $P_s > 0.18$]. Therefore, cross-race analyses compare African Americans in each condition to European Americans in both conditions (+2, -1, -1 contrasts).

Measures were assessed in the order listed below. Students in Cohort 2 completed all measures. Students in Cohort 1 completed the following measures: (a) accessibility of negative racial stereotypes, (b) accessibility of neutral race-related constructs, (c) accessibility of self-

doubt, (d) number of recent doctor visits, and (e) recall of the intervention. For all measures assessed in both cohorts, analyses combined data from both cohorts.

Cognitive Accessibility of Negative Racial Stereotypes and of Self-Doubt

A word-stem completion task was used to assess the accessibility of key constructs (*SI6*). Participants added letters to 40 (Cohort 1) or 46 (Cohort 2) word stems to form English-language words. Some word stems could be completed to form words related to either target constructs or irrelevant constructs (e.g., R A __ could be *race* or *rain*). There were three categories of target words (*SI*): (1) words related to negative stereotypes about African Americans (i.e., *anger/angry, bias, caste, class, dumb, lazy, police, poor, riot, sold, token, welfare*), (2) words related to race but not negative racial stereotypes (i.e., *black, color, dialect, race, soul, white* and, in Cohort 2, *brother*), and (3) words related to self-doubt (i.e., *dumb, guilt, inferior, loser, shame, shamed, weak* and, in Cohort 2, *hard*). A few words could be completed more than once; in such cases, each use of the word was counted under the assumption that repeated use reflects greater accessibility.

First, we examined the accessibility of negative racial stereotypes. We summed the number of negative race-related words completed. Covariates were student gender and the gender x condition interaction. The treatment effect for African Americans is reported in the main text [$t(66)=-2.01, P=0.049$] (Fig. 3B). The race x condition interaction was not significant [$F<1$]. Cross-race comparisons found that control-condition African Americans exhibited greater accessibility of negative racial stereotypes than European Americans [$t(66)=3.68, P=0.0005$]. This difference was eliminated for African Americans in the treatment condition [$t(66)=1.33, P=0.19$].

Second, we examined the total number of neutral race-related words completed. Covariates were student gender and preintervention racial identification and stereotype threat. There was no effect of condition for either racial group and no race x condition interaction [$t_s<1$]. There was only an overall effect of participant race [with condition effect-coded: $B=1.05, t(65)=3.01, P=0.004$]. African Americans exhibited greater accessibility of race than European Americans.

Third, we examined the total number of self-doubt-related words completed. Covariates were gender, the gender x condition interaction, the race x gender interaction, and the race x gender x condition interaction. The treatment effect for African Americans is reported in the main text [$t(64)=-2.64, P=0.010$] (Fig. 3C). The race x condition interaction was significant [$F(1,64)=6.38, P=0.014$]. Control-condition African Americans exhibited greater accessibility of self-doubt than European Americans [$t(64)=4.31, P=0.00006$]. This difference was eliminated for African Americans in the treatment condition [$t(66)=1.24, P=0.22$].

Self-Reported Uncertainty About Social Belonging in College

Students' self-reported level of uncertainty about their social belonging in college was assessed using a 3-item scale (*SI*): (1) "Sometimes I feel that I belong at [school name], and sometimes I feel that I don't belong at [school name]," (2) "When something good happens, I feel that I really belong at [school name]," and (3) "When something bad happens, I feel that maybe I don't belong at [school name]." All items used the same response scale ($1=strongly disagree, 7=strongly agree$). Factor analysis found that Item 2 loaded weakly on the common factor [Item 1: 0.94; Item 2: 0.46; Item 3: 0.85]. Removing this item also increased the scale reliability [from $\alpha=0.66$ to 0.82]. Therefore, the 2-item version (the mean of items 1 and 3) was

used. Covariates were preintervention levels of self-reported achievement behavior, stigma consciousness, and racial identification.

The treatment effect for African Americans is reported in the main text [$t(36)=-2.01$, $P=0.052$] (Fig. 3A). The race x condition interaction was not significant [$t(36)=1.15$, $P=0.26$]. Cross-race comparisons found that control-condition African Americans reported greater belonging uncertainty than European Americans [$t(36)=2.10$, $P=0.043$], replicating past survey research (S1). This difference was eliminated for African Americans in the treatment condition [$t<1$].

Subjective Happiness

Subjective happiness was assessed using the Subjective Happiness Scale (S17), a widely-used 4-item measure: (1) “In general, I consider myself:” ($1=not\ a\ very\ happy\ person$, $7=a\ very\ happy\ person$); (2) “Compared to most of my peers, I consider myself:” ($1=less\ happy$, $7=more\ happy$); (3) “Some people are generally very happy. They enjoy life regardless of what is going on, getting the most out of everything. To what extent does this characterization describe you?” ($1=not\ at\ all$, $7=a\ great\ deal$); and (4) “Some people are generally not very happy. Although they are not depressed, they never seem as happy as they might be. To what extent does this characterization describe you?” ($1=not\ at\ all$, $7=a\ great\ deal$) (reverse coded). The scale was reliable [$\alpha=0.89$]. Covariates were SAT-score and preintervention levels of sensitivity to race-based rejection, stigma consciousness, and perceptions of racial prejudice.

The treatment effect for African Americans is reported in the main text [$B=1.08$, $t(35)=2.61$, $P=0.013$] (Fig. 4C). The race x condition interaction was marginally significant [$t(35)=-1.90$, $P=0.065$]. Control-condition African Americans reported less happiness than European Americans [$t(35)=-2.42$, $P=0.021$]. This difference was eliminated for African Americans in the treatment condition [$t<1$].

Self-Reported Health

Self-reported general health. Self-reported health was assessed using the 5-item General Health component of the MOS Short-Form Health Survey (SF-36) (S18, S19)—(1) “In general, would you say your health is?” ($1=excellent$, $5=poor$); (2) “I seem to get sick a little easier than other people”; (3) “I am as healthy as anybody I know”; (4) “I expect my health to get worse”; and (5) “My health is excellent” (for items 2-5, $1=definitely\ true$, $5=definitely\ false$). Following standard procedures, we reverse scored items 1, 3, and 5 so that higher values reflect better health and recoded each item onto a 100-point scale. The scale was reliable [$\alpha=0.84$] so the items were averaged. Covariates were SAT-score, gender, the race x gender interaction, and preintervention levels of academic identification, self-reported achievement behavior, stigma consciousness, and stereotype threat.

The treatment effect for African Americans is reported in the main text [$t(32)=2.48$, $P=0.019$] (Fig. 4A). The race x condition interaction was marginally significant [$F(1,32)=3.76$, $P=0.061$]. Control-condition African Americans reported worse health than European Americans [$t(32)=-2.42$, $P=0.021$]. This difference was eliminated for African Americans in the treatment condition [$t<1$].

Number of recent doctor visits. We assessed the number of times students reported having visited the doctor recently using a 1-item measure: “How many times have you been to the doctor in the past 3 months?” (Cohort 1) or “During the past month, how many times did you go to the doctor?” (Cohort 2). Students were urged to estimate if they did not know. Scores from

Cohort 1 were divided by 3 to index a 1-month period. Covariates were preintervention levels of sensitivity to race-based rejection and stereotype threat.

The treatment effect for African Americans is reported in the main text [$t(63)=-2.23$, $P=0.030$] (Fig. 4B). The race x condition interaction was not significant [$t(63)=1.15$, $P=0.25$]. Control-condition African Americans did not differ from European Americans [$t<1$]. Treatment-condition African Americans reported fewer doctor visits than European Americans [$t(63)=-2.16$, $P=0.035$].

Because the outcome was skewed [$Z=6.17$, $P<0.001$], we also analyzed it using a nonparametric procedure. We tested the condition effect within each racial group on the mean rank in the residual number of doctor visits students reported in the previous month controlling for the aforementioned covariates (i.e., a Mann-Whitney test). The analysis replicated the basic effect. The condition difference was significant for African Americans [$Z=2.04$, $P=0.041$] and nonsignificant for European Americans ($Z<1$). We also tested whether condition affected the percentage of participants who reported having gone to the doctor at all in the previous month (Cohort 2) or 3 months (Cohort 1). Fewer African Americans reported having gone to the doctor in the treatment condition (27.78% had) than in the control condition (60.00% had) [$\chi^2(1, N=38)=3.98$, $P=0.046$]. This percentage did not vary for European Americans (57.14% and 55.00%, respectively) [$\chi^2(1, N=34)<1$].

Recall of and Beliefs About the Intervention

Recall of the intervention was assessed using a cued recall procedure. Students were reminded of the general procedure of the study. They read, “In the spring of your freshman year, you participated in a psychology study. In this study, you came to a laboratory and read the results of an ‘Upperclassmen Survey’ describing how students’ experiences at [school name] had changed over time. You then wrote an essay about how your experience so far at [school name] illustrated the process described in the Upperclassmen Survey... Finally, after participating in the laboratory you completed questionnaires twice a day for a week over email describing your daily experiences.”

First, students were asked, “Do you remember participating in the study described above?” They indicated “Yes” or “No.”

Second, students were asked, “What was the most memorable and important information, if any, that you got out of reading the Upperclassmen Survey?” They were given space to respond. Two trained coders blind to participants’ race independently coded whether each response accurately described the information conveyed in the student’s condition or did not. Reliability was high [Cohen’s Kappa=0.92] and discrepancies were resolved in discussion.

Third, students were asked, “The Upperclassmen Survey described how people’s experience at [school name] changed over time. What kind of change did it describe? (If you do not remember exactly, please provide your best guess.)” They were given space to respond. Two trained coders blind to participants’ race independently coded whether each response described students’ social experience and/or sense of belonging in college as improving over time or did not. Reliability was adequate [Cohen’s Kappa=0.72] and discrepancies were resolved in discussion.

Fourth, students were asked, “Did you experience the same kind of change as the Upperclassmen Survey described or not? (If you do not remember exactly what change the survey described, please respond using your best guess.)” They indicated if they “experienced the same kind of change” or “did not experience the same kind of change” and then were asked

“How or how not?”. Two trained coders blind to participants’ race independently coded whether each written response described students’ own social experience and sense of belonging in college as improving over time or did not. Reliability was high [Cohen’s Kappa=0.86] and discrepancies were resolved in discussion.

Fifth, students were asked, “Did reading the Upperclassmen Survey and participating in that study affect your experience in college in any way?” They indicated “Yes” or “No.” If they responded, “Yes,” students were given space to describe how participating in the study had affected them. A few students who left the item blank but who wrote simply that they did not remember the study were coded as responding “no.”

Supplementary Table 1

Tests of success of random assignment to condition. Data and analyses combined across cohorts (for all measures but preintervention GPA: $N_s=89-92$; for preintervention GPA: $N=70$). For (A), independent samples t-tests tested for differences between the treatment (I) and control (O) conditions among all participants. For (B), (C), and (D), multiple regression analyses tested for (B) condition differences among African American students, (C) condition differences among European American students, and (D) race x condition interactions.

<u>Preintervention Measure</u>	A: Condition difference, all participants		B: Condition difference, African American students		C: Condition difference, European American students		D: Race x condition interaction	
	<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>
Academic identification	-1.62	0.11	-1.42	0.16	-0.15	-0.99	0.24	0.81
Self-reported achievement behavior	1.63	0.11	0.76	0.45	1.58	0.12	0.65	0.52
Sensitivity to race-based rejection	0.16	0.87	-0.08	0.93	-0.33	0.74	-0.19	0.85
Stigma consciousness	0.40	0.69	0.81	0.42	-0.81	0.42	-1.15	0.25
Perceptions of racial prejudice	1.66	0.10	1.32	0.19	0.83	0.41	-0.28	0.78
Stereotype threat	0.44	0.66	-0.05	0.96	-0.03	0.98	0.01	0.99
Racial identification	-0.36	0.72	-0.30	0.76	-1.08	0.28	-0.61	0.55
SAT-Math score	-1.29	0.20	-0.72	0.47	-0.82	0.42	-0.10	0.92
SAT-Verbal score	-0.32	0.75	0.86	0.39	-1.10	0.27	-1.39	0.17
SAT score (combined)	-0.99	0.33	0.20	0.84	-1.29	0.20	-1.08	0.28
Preintervention GPA	0.31	0.75	-0.93	0.36	0.84	0.40	1.25	0.22

Supplementary Table 2

Comparison of participants who completed the end-of-college survey ($N_s=69-72$) and participants who did not complete the end-of-college survey ($N_s=18-20$) on preintervention measures. Data and analyses combined across cohorts. For **(A)**, independent samples t-tests tested for differences between completers (I) and noncompleters (O) among all participants. For **(B)**, **(C)**, and **(D)**, multiple regression analyses tested for **(B)** differences between completers and noncompleters among African American students, **(C)** differences between completers and noncompleters among European American students, and **(D)** race x survey-completion interactions.

<u>Preintervention Measure</u>	A: Difference between end-of-college survey completers and noncompleters, all participants		B: Difference by survey completion, African American students		C: Difference by survey completion, European American students		D: Race x survey-completion interaction	
	<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>
Academic identification	1.42	0.16	0.82	0.42	1.24	0.22	0.36	0.72
Self-reported achievement behavior	-0.27	0.79	-0.45	0.65	0.09	0.93	0.37	0.71
Sensitivity to race-based rejection	-0.96	0.34	-1.23	0.22	-0.30	0.76	0.60	0.55
Stigma consciousness	-0.35	0.73	-0.63	0.53	0.14	0.89	0.53	0.60
Perceptions of racial prejudice	1.36	0.18	0.93	0.36	0.99	0.32	0.08	0.93
Stereotype threat	-0.20	0.84	-0.20	0.84	-0.14	0.89	0.03	0.97
Racial identification	-0.47	0.64	-0.94	0.35	0.22	0.82	0.80	0.43
SAT-Math score	0.20	0.84	0.37	0.71	0.15	0.88	-0.15	0.88
SAT-Verbal score	0.16	0.88	-0.70	0.49	1.03	0.30	1.23	0.22
SAT score (combined)	0.51	0.61	-0.05	0.96	0.78	0.44	0.61	0.55

Supplementary Table 3

Recall of and beliefs about the intervention 3-years postintervention. Data and analyses combined across cohorts. For **(A)** the condition effect was nonsignificant for both European Americans and African Americans [$\chi^2(1, N_s=34 \text{ and } 38) < 1$]. For **(B)** the condition effect was marginal for European Americans [$\chi^2(1, N=34)=3.04, P=0.081$] and nonsignificant for African Americans [$\chi^2(1, N=38)=1.37, P=0.24$]. For **(C)** and **(D)** the condition effect was trending for European Americans [$\chi^2(1, N=34)=3.28, P=0.070$ and $\chi^2(1, N=34)=1.20, P=0.27$, respectively] and significant for African Americans [$\chi^2(1, N=38)=3.79, P=0.052$]. For **(E)**, the condition effect was nonsignificant for both groups [$\chi^2(1, N=34)=2.14, P=0.14$ and $\chi^2(1, N=37) < 1$, respectively].

		A: Percent (%) who reported remembering the study	B: Percent (%) who accurately recalled the most “memorable and important information” they learned in the study	C: Percent (%) who guessed that the survey described students’ social experience in college as improving over time	D: Percent (%) who spontaneously described their own social experience and sense of belonging in college as improving over time	E: Percent (%) who reported that participating in the study had affected their college experience in any way
European American Students	Randomized Control Group	75.00%	0.00%	15.00%	25.00%	5.00%
	Social-Belonging Treatment	85.71%	14.26%	42.86%	42.86%	21.43%
African American Students	Randomized Control Group	85.00%	5.00%	20.00%	20.00%	15.79%
	Social-Belonging Treatment	72.22%	16.67%	50.00%	50.00%	16.67%

Supporting References

- S1. G. M. Walton, G. L. Cohen, A question of belonging: Race, social fit, and achievement. *J. Pers. Soc. Psychol.* **92**, 82-96 (2007).
- S2. R. Mendoza-Denton, G. Downey, V. J. Purdie, A. Davis, J. Pietrzak, Sensitivity to status-based rejection: Implications for African American students' college experience. *J. Pers. Soc. Psychol.* **83**, 896-918 (2002).
- S3. E. Pinel, Stigma consciousness: The psychological legacy of social stereotypes. *J. Pers. Soc. Psychol.* **76**, 114-128 (1999).
- S4. G. L. Cohen, J. Garcia, "I am us": Negative stereotypes as collective threats. *J. Pers. Soc. Psychol.* **89**, 566-582 (2005).
- S5. R. M. Sellers, S. A. J. Rowley, T. M. Chavous, J. N. Shelton, M. A. Smith, Multidimensional inventory of black identity: A preliminary investigation of reliability and construct validity. *J. Pers. Soc. Psychol.* **73**, 805-815 (1997).
- S6. T. D. Wilson, M. Damiani, N. Shelton in *Improving Academic Achievement: Impact of Psychological Factors on Education*, J. Aronson Ed. (Academic Press, Oxford, England, 2002), pp. 91-110.
- S7. J. Aronson, C. B. Fried, C. Good, Reducing the effect of stereotype threat on African American college students by shaping theories of intelligence. *J. Exp. Soc. Psychol.* **38**, 113-125 (2002).
- S8. J. Cooper, R. Fazio in *Advances in Experimental Social Psychology*, L. Berkowitz, Ed. (Academic Press, New York, 1984), vol. 17.
- S9. W. G. Bowen, D. Bok, *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions* (Princeton University Press: Princeton, NJ, 1998).
- S10. C. M. Steele, S. J. Spencer, J. Aronson in *Advances in Experimental Social Psychology*, M. P. Zanna Ed. (Academic Press, San Diego, CA, 2002), pp. 379-440.
- S11. D. R. Williams, Race, SES, and health: The added effects of racism and discrimination. *Ann. N.Y. Acad. Sci.* **896**, 173-188 (1999).
- S12. D. S. Massey, C. Z. Charles, G. Lundy, M. J. Fischer, *The Source of the River: The Social Origins of Freshmen at America's Selective Colleges and Universities* (Princeton University Press: Princeton, NJ, 2002).
- S13. C. M. Steele, A threat in the air: How stereotypes shape intellectual identity and performance. *Am. Psychol.* **52**, 613-629 (1997).
- S14. D. Muller, C. M. Judd, V. Y. Yzerbyt, When moderation is mediated and mediation is moderated. *J. Pers. Soc. Psychol.* **89**, 852-863 (2005).
- S15. R. Darlington, *How many covariates to use in randomized experiments?* (<http://www.psych.cornell.edu/darlington/covarnum.htm>)
- S16. C. M. Steele, J. Aronson, Stereotype threat and the intellectual test performance of African Americans. *J. Pers. Soc. Psychol.* **69**, 797-811 (1995).
- S17. S. Lyubomirsky, H. S. Lepper, A measure of subjective happiness: Preliminary reliability and construct validation. *Soc. Indic. Res.* **46**, 137-155 (1999).
- S18. C. A. McHorney, J. E. Ware, J. F. R. Lu, C. D. Sherbourne, The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med. Care* **32**, 40-66 (1994).
- S19. L. Stewart, J. E. Ware Eds., *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach* (Duke University Press, Durham, NC, 1992).