Supporting Information Appendix

This pdf file includes:

Supporting Information Text

Materials and Methods

Study 1

Site and Student Demographics

As reported in *1*, the middle school site for Study 1 was located in a middle-class suburban school district in the Mountain West region of the United States. The affirmation manipulations were delivered across the entire school to students in the 6th, 7th, and 8th grades. The study took place at approximately the same time for all three grade-level cohorts, and thus analyses focus simultaneously on all three groups over the observation period of three years, i.e., for each grade-level cohort, through the 8th, 9th, and 10th grades, respectively, as done in the original study (*1*). "Year 1" designates the year in which the intervention was administered for each grade level, "Year 2" the grade after the intervention year, and "Year 3" the grade two years after the intervention year.

School records indicated that approximately 47% of the student body was White and 45% was Latino American; 8% belonged to another ethnic group. In both Study 1 and Study 2, we use "White" to designate students who were Caucasian but not of Latino origin. Although the school in Study 1 was located in a middle-class neighborhood, its Latino American students faced language barriers, and tended to be poor. 78% of Latino Americans at the school indicated Spanish was their primary language spoken at home, and 90% received meal assistance. Poorer Latino Americans tended to face the strongest language barriers. Among those who received free or reduced lunch, approximately 81% were children of first-generation immigrants, predominantly from Mexico.

Participants

As a result of intensive recruitment efforts, in Study 1 we were able to obtain participation from 73% of the students in the school, yielding a total of 185 students who began the study as 6th, 7th, or 8th graders. Below we refer to these as the 6th grade cohort, 7th grade cohort, and 8th grade cohort, respectively. The sample comprised students who had official grade point average (GPA) data available during the intervention year, and whose ethnicity, as confirmed via school records, was either White (N=104) or Latino American (N=81). Ninety-three participants had been randomly assigned to the affirmation condition (52 White, 41 Latino American), and 92 to the no-affirmation control condition (52 White, 40 Latino American). The foundational paper (*1*) excluded one participant whose core course GPA–the focal dependent measure in that paper–fell more than three studentized residuals from the model-predicted value. Given that the student was not an outlier for outcomes measured in subsequent years, we included this participant in the present analyses. Inclusion or exclusion of this participant had virtually no impact on the results, with statistical significance unchanged (as was also the case in *1*). No students were excluded on the basis of exercises completed. Incidentally, all students in the study completed at least two exercises.

<u>Availability of Outcome Data and Attrition</u>

In Study 1, there were four primary outcomes: 1) course difficulty; 2) enrollment in remedial clinics and 3) the AVID college preparatory program in 8th grade; and 4) enrollment in the mainstream high school. The availability of some outcome data varied as a function of students' grade level, because some institutional channels were available only in certain grades. However, as we detail below, in no case was there differential attrition of participants or differential availability of data by condition, either in the sample overall or within the focal ethnic group.

Each analysis included all available participants. Participants were treated as missing for a given analysis when it was not possible to measure that outcome. Missing data occurred for the following reasons: because the outcome was specific to a grade level the cohort had already completed or had not yet reached, or because students had left the school district at the time an outcome was measured.

For some outcomes, data were available only for two of the three grade-level cohorts (see Table S1). Additionally, for all three cohorts, there was some attrition within cohort (see Table S2), due to students leaving the district (and thus not having available transcript data). On average, attrition was 7% each year during the data collection period. By the end of Year 3, 13.5% (*N*=25) of students who were present during the intervention year had stopped attending a school in the district (e.g., due to dropping out or moving away) and thus were considered lost to attrition. Students who transferred to other schools in the district were included in analyses and thus were not counted toward attrition. Analyses of attrition and school transfers are reported below.

As would be expected given that students were randomized to condition within grade level, in no case did the differential availability of an outcome vary by condition, either overall or among minority students (We report the Fisher's Exact test in cases where the expected value for some cells is less than 5). For course difficulty, enrollment in AVID, and enrollment in remedial clinics, attrition did not vary by experimental condition either across all participants [χ^2s<0.20, *P*s>0.65] or for the focal ethnic group of Latino Americans [χ^2s<0.41, *P*s>0.50; Fisher's Exact test: *P*s>0.55].

Importantly, when data were missing, the participant was excluded from analysis, with one exception. The exception was whether students enrolled in the mainstream high school or not in Year 3. For this outcome alone, students who left the district---most often this occurred in the transition from middle to high school---were treated as 0. Such students did not appear on enrollment lists in any high school in the district at the time of data collection and could be assumed to have dropped out. They were thus combined with other students who attended alternative, non-traditional high schools that were intended for students not on a college track. Suggesting the appropriateness of our designation, both students who left the district and students who enrolled in alternative high schools had lower middle school performance than those who enrolled in the mainstream high school (see p. 20 in Detailed Results section below). As a robustness check, we report an alternative analysis that omits students who left the district, and examines treatment effects on whether minority students were enrolled in a mainstream versus alternative high school in Year 3. Such an analysis yields similarly statistically significant results (again, see SI Text p. 20), suggesting that treatment effects for minority students along this outcome were robust to treating the students who left the district as missing and excluding them from analysis.

Below we provide further detail on the availability of data for each outcome. We obtained course difficulty data in Year 3 for 160 of 185 students (86% of the original sample) across all three grade cohorts, 67 of 81 Latino Americans (83%) and 93 of 104 Whites (89%). By this year, variability in course difficulty levels was high because two of the three cohorts had by this time entered high school, where there was more differentiation in course levels. Course difficulty in Year 3 thus constitutes our key dependent variable, which we report in the main text. We also examined course levels averaged across Years 2 and 3. Regardless of outcome (i.e., course difficulty in Year 3 alone or Years 2 and 3 combined), analyses yield similar statistically significant results, as discussed in the Detailed Results section below (see SI Text p. 19). Analysis of course difficulty in Year 1 as an outcome was not feasible, because course levels in this year had been determined prior to intervention for the two younger cohorts and after only three of five doses for the 8th grade cohort. As shown in the Detailed Results section (SI Text p. 19), controlling for pre-intervention course levels (Year 1 for the 6th and 7th grade cohort and Year 0 for the 8th grade cohort) makes the results stronger.

We assessed enrollment in remedial clinics (enrolled versus not) and in the college readiness program Advancement via Individual Determination (AVID) for the 90% of students who had commenced the intervention in 6th or 7th grade and who had not left the school district by the time these clinics were offered, in 8th grade (Year 2 for the 7th grade cohort and Year 3 for the 6th grade cohort; see Table S1). These outcomes were available only for the two younger cohorts. Eighth grade remedial clinics were a new policy implemented by the school in Year 2, after the oldest cohort had already matriculated to 9th grade. The remedial clinic data encompassed 53 of the 61 Latino American students (87% of Latino Americans in the 6th and 7th grade cohorts) and 67 of the 73 White students (92% of Whites in the 6th and 7th grade cohorts). Enrollment in AVID had already been determined for the oldest cohort prior to the commencement of the intervention. In addition, AVID enrollment was assessed only for Latino Americans, as a program aimed at minority youth (only 1 White student in our sample was enrolled). AVID data were available for the same 53 Latino Americans not lost to attrition.

As with course difficulty, we analyzed high school enrollment for all students in the two older cohorts (N=130) in Year 3, two years after the commencement of the intervention. High school data for the youngest cohort (the 6th grade cohort) were unavailable because these students were still in middle school at this time (8th grade). Analyses for this outcome included 52 Latino Americans and 78 Whites (100% of those students who began the study as 7th or 8th graders).

Procedures

Both affirmation and control exercises followed procedures similar to those developed and validated in prior research (*2*-*3, 4*). The written instructions used to guide students through the exercises had been thoroughly tested to ensure that they were intelligible, age-appropriate, and self-explanatory. A pilot study in the previous year served to refine the materials so as to make them comprehensible for students of all grades and linguistic backgrounds at the school, an important step given the English literacy limitations of some students. Additional details can be found in the original report (*1*).

Through scripts and rehearsals, teachers were trained in how to administer the exercises and in how to respond to questions from students that might arise. For example, if a student asked about the purpose of the exercise, teachers were instructed to reply, “This is an exercise that the school gives to students a few times a year.” Teachers were told to describe the exercise neither as part of a research study nor as an activity intended to benefit students but as a regular classroom assignment.

Several features of the protocol ensured that teachers remained unaware of experimental condition and that students were unaware of differences between the writing exercises. The research team wrote each student’s name on the outside of an envelope containing his or her randomly assigned exercise, allowing teachers to deliver the appropriate exercises to their students without knowing students' condition assignment. The affirmation and control exercises were designed to look similar and to require equal amounts of time and effort. To further minimize the likelihood that students would discover differences in the materials, they were instructed to remain quiet and to ask questions only by raising their hand, at which time their teacher would approach them individually. The research team worked behind the scenes and outside of regular school hours to prevent students from linking the exercises to an external research project.

In both the affirmation and control conditions, the writing exercises were delivered on four to five occasions. The number of exercises offered varied by cohort. In Year 1, the three cohorts received two or four exercises. The 8th grade cohort received only two because, as part of an initial piloting of the study materials, they had received three exercises in the previous year (*1*). In Study 1, students in our sample completed between 2 and 5 exercises over the two years of the study (M=3.84 to 4.82, SD=0.50 to 0.43, across the three cohorts). Exercises completed did not differ by condition (M_{Control}=4.04, $M_{\text{Treatment}}$=4.14, $t(183)$=0.997, P=0.320; see Table S3). However, no students were excluded from analyses on the basis of number of exercises completed. Even as students varied in the number of exercises they were offered, the proportion of offered exercises completed was high. On average, students in our sample completed 96% of the exercises offered to them.

For each exercise, students in the values affirmation condition were asked to reflect on core personal values, such as relationships with friends and family, and write about why these values were important to them. By contrast, students in the control condition were asked to reflect on and write about non-affirming topics, such as a value that was unimportant to them personally and why it might be important to someone else, or a daily routine. The materials featured icons to make them more engaging and memorable. To avoid repetitiveness over multiple administrations, three different but conceptually similar writing tasks were developed (see *1* for more details). The structure of the exercises varied slightly. The first intervention in a given year omitted values that were directly relevant to the academic domain (e.g., being smart). For interventions occurring later in the school year, students either repeated this exercise with the academic domain added as a value option, wrote a more open-ended essay in which they were encouraged to generate important values on their own, or received a writing exercise tailored to a specific value they had emphasized in a previous exercise (See SI Text pp. 22-24 in the Detailed Results section for a discussion of the effect of the tailored exercises specifically).

Outcomes

Course Difficulty Scores

We computed an average course difficulty score for each student for the courses taken in the school grade two academic years after the intervention (Year 3), when students were in the 8th, 9th, or 10th grades, depending on cohort. Course difficulty scores were calculated based on the "level" of the courses in which students were enrolled at the beginning of the school year, according to their official school transcripts. As noted previously, the focal year for this outcome was Year 3 because of limited variability in course difficulty levels in Year 2.

Course difficulty was determined based on the course progressions outlined in the school district's course catalogs for middle and high school. Because both the number of course subjects with differentiated levels (e.g., English, Math) and the number of difficulty levels within each subject (e.g., remedial to advanced) varied across grades, course levels data were first standardized for each course subject in each grade. The z-scores for each course subject were then averaged for each student, yielding a single course difficulty score. The score represented each student's average course difficulty across all his or her course subjects with differentiated levels, with higher scores signifying greater difficulty. In cases where a student had incomplete course levels data (e.g., data for only four out of five course subjects), the course difficulty composite was calculated based on the average of the available course data. Additional analyses indicated that results were not sensitive to alternative treatments of incomplete data. Below we describe how course difficulty was first scored for each course in each grade (before these scores were standardized and averaged into a composite).

In 8th grade, students' course levels varied only for two course subjects, Math and English. For each of the two subjects, the relevant course levels index was first assigned a value of 1 (remedial clinics), 2 (basic level), or 3 (advanced level). While students in remedial courses for a given course subject also took part in the basic course for that subject, for this composite they were assigned a value of 1 to differentiate them from students who were only in basic courses (assigned a value of 2).

Once in high school, students had more course options. In 9th grade, the first year of high school, course levels varied in five course subjects rather than only two: Math, English, Science, Social Studies, and foreign language courses. For each subject, levels were again assigned a numerical value, with higher values representing greater difficulty. Math courses ranged from 1 (Algebra 1A) to 4 (Advanced Geometry). English courses ranged from 1 (Sheltered Language Arts) to 3 (Advanced Language Arts). Science courses ranged from 1 (Biology or Physical Science) to 2 (Advanced Biology or Advanced Physical Science). Social Studies courses ranged from 1 (U.S. History) to 3 (A.P. U.S. Government and Politics). Foreign language courses ranged from 1 (e.g., Spanish 1) to 4 (e.g., Spanish 4).

In 10th grade, students' course levels again varied for Math, English, Science, Social Studies, and foreign language courses, with higher scores again indicating greater difficulty. Tenth grade course levels differed from 9th grade course levels in the following minor ways: Math courses now had six levels of differentiation rather than only four, ranging from 1 (Algebra 1A) to 6 (Advanced Pre-Calculus); Social Studies courses had two levels of differentiation rather than three, ranging from 1 (World History) to 2 (A.P. World History); and foreign language courses had five levels of differentiation rather than only four, ranging from 1 (e.g., Spanish 1) to 5 (e.g., A.P. Spanish 5)

Enrollment in Remedial Clinics

Starting in Year 2, the middle school began offering remedial Math and English clinics for eighth graders. These clinics were designed to help struggling students keep up with their peers in the basic courses in each subject. Clinic assignment was based on a combination of factors, including previous performance and grades, standardized test scores, and teacher recommendations. Enrollment in remedial clinics was measured using a dichotomous variable equal to 1 if the student (in the 6th or 7th grade cohort) was enrolled and 0 if not. As noted above, this outcome was unavailable for the 8th grade cohort.

Enrollment in AVID

The AVID course, part of a nationwide college readiness program, aims to cultivate skills, motivation, and support systems needed to help underserved students graduate high school and go to college. To enroll in the course, offered in 8th grade, students had to participate in a selective admissions process that entailed a written application and oral interview that mimicked the procedure for applying to a job or college. Virtually all the students admitted were Latino American (93%, or 14 out of 15 students in the two youngest cohorts, the ones who began the intervention prior to 8th grade and whose AVID enrollment could thus have been affected by the intervention). Enrollment in AVID was measured using a dichotomous variable equal to 1 if the student (in the 6th or 7th grade cohort) was enrolled and 0 if not. As noted above, this outcome was unavailable for the 8th grade cohort.

Enrollment in the Mainstream High School

During the data collection window following the intervention, students in the two older cohorts could have enrolled in the mainstream high school or not. By Year 3, they were either in the 9th or 10th grade. Discussion with the guidance counselor at the school site indicated that the mainstream high school was the school where "college bound" students tended to matriculate after middle school. Thus, analyses focus on whether students (in the 7th or 8th grade cohort) were enrolled in the mainstream high school (=1) or not (=0) in Year 3. (Focusing on whether students were enrolled at the start of each cohort's 9th grade year did not change the results.) If students did not enroll in the mainstream high school, there were two possibilities: they were either no longer in the dataset (N=18), due to dropping out or enrolling in another school outside the district, or they attended one of four types of alternative high schools in the district (N=11). Alternative high schools varied in their design and purpose. Most of the students who went to an alternative high school (N=6) attended one that offered more personalized instruction to learn the same concepts taught to students in the mainstream high school. The remainder attended a school designed for those who had been adjudicated or who were progressing at a slower pace than same-aged peers (N=2), a school that taught technical skills for a specific vocation (N=1), or another traditional high school in a different geographical region of the district (N=2). Discussion with the partner middle school suggested that the latter geographical transfers typically occurred in problematic cases when students' parents wished to extricate their child from bad peer influences at the original school. Removal of these two more ambiguous cases did not change the statistical significance of the results. As noted above, this outcome was unavailable for the 6th grade cohort.

Baseline Covariates

Pre-Intervention Performance

There were two measures of pre-intervention performance, obtained from students' official school records (as reported in *1*): GPA in core courses (Math, Social Studies, Science, and English) and performance on the state achievement test, both assessed in the academic year prior to the intervention year, Year 0. Our primary analyses included as a covariate a single composite representing the mean of these two performance indicators after each had been standardized, as done in the original study (*1*). Using this composite permitted all subjects to be included in analyses, even when they had missing values for one of the indicators. In particular, seven students were missing standardized test scores. For these students the composite score was simply based on their standardized pre-intervention GPA.

As expected, minority students had significantly lower baseline performance [M=-0.74, SD=0.62] than White students [M=+0.55, SD=0.72], $t(179)=12.60$, $P<0.001$. For this reason, as noted in the Analytic Model section (SI Text p. 7), we mean-centered the pre-intervention performance covariate on 0 for each ethnicity.

Was Random Assignment Successful?

Yes. Table S3 shows the means for the key baseline covariates by condition, for the Study 1 sample as a whole. T-tests revealed no significant differences along ethnicity, gender, or baseline performance by condition [*t*s<0.3, *P*s>0.75]. Nor were there significant condition differences on any baseline variable among the focal ethnic group, Latino Americans [*t*s<0.55, *P*s>0.58].

As another check on random assignment, we regressed the pre-intervention performance composite and student gender separately on the main effects of ethnicity, condition, and their interaction. As expected due to randomization, neither the main effect of condition [*zs*/*t*s<0.30, *Ps*>0.75] nor the ethnicity × condition interaction was significant for either measure [*zs*/*t*s <0.55, *Ps*>0.55]. The main effect of ethnicity was also non-significant [*z*s/*t*s <0.80, *P*s>0.43], though this is less relevant because treatment effects were assessed withn ethnicity. Isolating the focal subgroup of Latino Americans, the simple main effect of condition was also non-significant for both baseline measures [*zs*/*t*s <0.60, *Ps*>0.57].

As a further check on random assignment, we used a supplemental technique that isolates the portion of variance in the outcome associated with baseline covariates. We estimated the predicted values of each of our primary outcomes using a multiple regression that controlled for pre-intervention performance and gender. As expected, these predicted values were associated neither with condition [*t*s<0.25, *P*s>0.80] nor with its interaction with ethnicity [*t*s<0.80, *P*s>0.40] in the full sample, or with condition in the subsample of Latino Americans [*t*s<0.70, *P*s>0.50].

Data Analyses

Linear regression was used to analyze the continuous course levels variable, logistic regression to analyze the dichotomous outcomes. The key coefficients were those associated with the ethnicity × condition interaction and the simple effect of condition for minority students.

<u>Analytic Model</u>

For each outcome, a regression analysis included the following predictors: contrast-coded student ethnicity (-1=Whites, +1=Latino Americans), contrast-coded student gender (-1=male, +1=female), students' randomly assigned intervention condition (0=control, +1=affirmation), the composite of student pre-intervention performance, and the two-way ethnicity × condition interaction. A 0/1 dummy code was used for condition so that the simple effect test of condition would yield a coefficient that corresponded to the effect of condition (i.e., a one-unit increase in the predictor) for a given ethnic group (the ethnic group defined as 0).

As in the foundational report (*1*), baseline performance was mean-centered on zero for each ethnic group, by subtracting the average score for a student's ethnic group from his or her individual score (i.e., a student with average performance for his or her ethnic group would have a score of 0 for the mean-centered variable). Consequently, the covariate-adjusted means reflect the outcome for the average-performing student within each ethnic group. Because Latino Americans had lower baseline performance than White students (see SI Text p. 6), failure to mean-center the baseline performance covariate within ethnicity would artificially inflate the outcomes for Latino Americans and deflate them for Whites.

Neither of these analytic conveniences—dummy coding condition or mean-centering the baseline performance covariate on zero within ethnicity—has any bearing on the statistical significance of any reported effect. Additionally, including terms for the gender × condition and gender x ethnicity interaction in either study did not significantly explain additional variance, except for two outcomes, as described in the Detailed Results section below (see SI Text p. 24).

Analytic Approach

For each outcome, we used multiple regression to obtain the ethnicity × condition interaction. This assessed whether the effect of condition varied by ethnic group. To compute the simple effect of condition for the focal ethnic group of interest, we used a dummy-coded ethnicity variable, in which Latino Americans (Study 1) or African Americans (Study 2) were designated as 0, Whites as +1. With this configuration, the main effect of condition in the model signifies the effect of affirmation for minority students. To compute the effect of condition for Whites, we reversed this coding (i.e., Whites as 0 and minorities as +1). Ancillary analyses that tested the condition main effect in each ethnic group separately, rather than with the full-sample simple effects approach, yield virtually identical results, as reported in the Detailed Results section below (see SI Text p. 18-19).

Study 2

Site and Student Demographics

The school site in Study 2 was located in a middle- to lower-middle-class suburban school district in the northeastern United States. The affirmation manipulation was delivered early in 7th grade to three successive cohorts of students. Its short-term effects on the African Americans in the first two cohorts were documented in a 2006 *Science* report (*2*); the third cohort was added in a 2009 *Science* report (*3*). Each cohort was followed for at least seven years after 7th grade (ranging from 7 years for cohort 3 to 9 years for cohort 1), a sufficient period of time to assess whether students had enrolled in college. School records indicated that approximately half of the student body was White, the remainder mostly African American, with a few students from other ethnic groups. The students at the school were not as poor as the Latino American students in the Mountain West school of Study 1. Still, the school site for Study 2 contained many students of lower socioeconomic status. While these data were not available for the school as a whole, 16% of students in our sample received free or reduced school lunch, with that figure rising to 28% among African American students. Approximately 50% of the full student body participated in the study. Additional background and details for this study can be found in the previous two reports (*2*-*3*).

Participants

We include all participants in the final sample of the 2009 *Science* report (*3*). These were all the participants (three cohorts of 7th graders who participated in the study in sequential chronological years) who had GPA data for each academic term of the 7th and 8th grades. Henceforth we refer to 7th grade as year 1 and 8th grade as year 2. (A few students repeated 7th grade and for these students their year 2 GPA was the GPA they earned when they repeated 7th grade). Our strategy for computing baseline performance, in which available indices were averaged into a single composite, as described in the Methods of Study 1 above (SI Text p. 6), enabled us to include one additional African American student who had a missing value for 6th grade GPA. Inclusion or exclusion of this participant had virtually no impact on the results, with statistical significance unchanged (as was the case in *2* and *3*).

As in our original two *Science* reports (*2*-*3*), analyses in the main text focused on African American students and White (non-Hispanic) students due to the relatively small number of students from other ethnicities. Our final sample consisted of 339 students. Of these, 47% were African American (*N*=158) and 53% were White (*N*=181). The sample contained approximately equal numbers of males (48%, *N*=162) and females (52%, *N*=177). As in the second foundational paper (*3)*, each of the three cohorts contributed roughly a third of the students to the final sample (ranging from *N*=103 to 120 students, or 30% to 35% of the sample). As in the original reports (*2*-*3*), using an alternative, bimodal category for ethnicity, comprised of Latino Americans and African Americans in a “negatively stereotyped” group (*N*=189) and Whites and Asian Americans in a “non-negatively stereotyped” group (*N*=195), yielded similar results. This combining of ethnicities was a reasonable analysis in this study as it yielded 45 more participants. In Study 1, the analogous addition of African Americans and Asian Americans to the sample yielded only 16 more participants. 51% of participants (*N*=172) were randomly assigned to the control condition, and 49% (*N*=167) to the affirmation condition.

<u>Availability of Outcome Data and Attrition</u>

Each analysis in the main text included all available African American and White participants (N=339), except for college selectivity, which included only those students who attended 4-year colleges (N=199). This sample (N=339) included all students for whom we had available outcome data during middle school (i.e., middle school GPA for the four semesters in 7th and 8th grade).

For the primary college outcome, enrollment in a 2-year or 4-year college, there was no attrition because we retained all students included in the 2009 foundational report (*3*) and collected data on each student's college enrollment status from public databases, based on students' names and birthdates. We obtained data related to students' college enrollment from the National Student Clearinghouse (NSC). The NSC is a comprehensive database frequently used by researchers to track outcomes related to college enrollment and persistence (*5*). More than 3,600 colleges and universities participate. The comprehensiveness of the NSC database can be assessed by its "coverage rate": enrollment in colleges in the database divided by total enrollment at all U.S. colleges (*5*). The latter are available from the Integrated Postsecondary Education Data System (IPEDS), which lists every college that accepts federal financial aid (over 6,000). As of fall 2013, when the data for Study 2 were collected, the NSC covered 95.6% of all students enrolled in 4-year institutions and 95.4% of all students enrolled in 2-year institutions in the U.S. In the Northeast Region, where Study 2 took place, the coverage rate was slightly higher; the NSC covered records for over 97% of African Americans and over 96% of non-Hispanic Whites enrolled in public 2- and 4-year colleges and private 4-year colleges (*5*). If the 7th grade participants were successfully promoted to each subsequent grade without repeating any, the first cohort would have been eligible to begin college in fall 2009 (year 7 of the study), the second in fall 2010, and the third in fall 2011 (year 9 of the study). Consistent with standard practices (*5*), students who had no record in the NSC database were deemed not to have enrolled in college.

To reduce the possibility that students' names could be matched incorrectly, we relied on a number of best practices and on the input of experts in the use of the NSC database. We submitted multiple versions of first names, included middle names whenever possible, and submitted multiple birthdates where any minor typographical errors in the official data records were suspected (e.g., a year indicated as a day of the month). Despite these precautions, measurement error can still occur, specifically with respect to for-profit colleges. These institutions are less likely to submit data to the NSC, and when they do, it is less likely to be on-time or accurate. Where possible, we also checked enrollment using official online sources such as college directories.

College selectivity scores were available for the 199 students who attended 4-year colleges (59% of the sample), as detailed below. Only 4-year colleges were included in this composite because there were no official sources of data for the selectivity of the 2-year colleges in our dataset. A t-test revealed that students in each condition were equally likely to have 4-year college selectivity data available ($M_{\text{control=}}$0.58, $M_{\text{Treatment}}$=0.60, t(337)=-0.433, P=0.665). This was also true when analyses were confined to the focal subgroup of African Americans ($M_{\text{control=}}$0.51, $M_{\text{Treatment}}$=0.62, t(156)=-1.302, P=0.195).

Procedures

The experimental procedures were the same as those in Study 1, with two exceptions (see *3* for more detail). First, the configuration of the cohorts differed because of logistical considerations tied to each school site. Whereas Study 1 involved three grade cohorts (6th, 7th, and 8th graders) enrolled simultaneously, Study 2 involved three sequential years of 7th graders (the other grades did not participate). Second, the administration of the manipulation was spread over two years for all cohorts rather than one.

In Year 1, all students received several exercises (between three and five depending on cohort). The number of exercises offered varied from cohort to cohort because of practical constraints (*3*). Within cohort, treatment dosage varied across years. In year 2, half the treatment group (*N*=84, 25% of the entire sample) was randomly assigned to complete additional doses of affirmation (between two and four, depending on cohort). The dosage in Year 2 did not provide additional benefit, as discussed in the Detailed Results section (see SI Text p. 22). In both years, students not completing affirmation exercises at any given administration instead completed control exercises.

Beyond practical constraints and planned differences in treatment dosage, students were sometimes absent on days in which the treatment and control exercises were administered. As a result of all these factors, students in our sample completed between 1 and 8 exercises over the two years of the study (M= 6.54 to 7.68, SD=0.82 to 0.89, across the three cohorts). Despite this variation, exercises completed did not differ by condition either across both years (M_{Control}=6.98, $M_{\text{Treatment}}$=7.02, $t(337)$=0.442, P=0.659) or in year 1 (M_{Control}=3.73, $M_{\text{Treatment}}$=3.80, $t(337)$=0.861, P=0.390). See Table S4. However, as in Study 1, no students were excluded from analyses on the basis of exercises completed. On average, students completed 95% of the exercises assigned to them.

Outcomes

College Enrollment (Ordinal)

To obtain an overall picture of how affirmation affected college enrollment, we first created a trichotomous/ordinal outcome to represent students' college enrollment status, defined as whether they ever enrolled in college at any time before or including the point of submission of their name and birthdate to the NSC in December 2013, 10 years after the 7th grade students in cohort 1 took part in the affirmation intervention. A value of -1 corresponded to no college, 0 to enrollment in a 2-year college, and +1 to enrollment in a 4-year college. If a student enrolled in a 2-year college and eventually went to a 4-year college, he or she was classified as having enrolled in a 4-year college.

College Enrollment (Dichotomous)

To permit more detailed analyses, we decomposed the ordinal outcome into two dichotomous outcomes: enrollment in any college (=1) versus no college (=0), and enrollment in a 4-year college (=1) versus a 2-year college or no college (=0). In the Detailed Results Section (see SI Text p. 20), we also directly compare attendees of 4-year colleges (=1) with attendees of 2-year colleges (=0), excluding those who did not attend college.

Selectivity of 4-Year Colleges Attended

We constructed a selectivity composite from four indicators linked to the colleges that students attended. The composite included all major selectivity indicators that were officially reported and publicly available, from either the Integrated Postsecondary Education Data System or Barron's Profiles of American Colleges, near the time of participants' eligibility for college enrollment. The indicators were (1) the acceptance rate of each college (reverse-coded), (2) the individual section means (e.g., Math, Verbal, Writing) of the SAT and (3) ACT for students at the 25th and 75th percentile of the college's entering class, and (4) each college's rating in Barron's. We chose Barron's because it is often used by researchers as a measure of college selectivity. It provides a system that ranks schools in a single inclusive list, whereas other popular ranking systems such as the U.S. News and World Report use several distinct lists (e.g., there are separate rankings for liberal arts colleges, regional universities, etc.). The college's Barron's category was assigned a numerical value from 0 to 7, using the following metric: 0 (*less competitive*), 1 (*competitive*), 2 (*competitive* +), 3 (*very competitive*), 4 (*very competitive* +), 5 (*highly competitive*), 6 (*highly competitive* +), 7 (*most competitive*). Each of the four indices was an average of a given college's annual score over the three most recent years that data were available near the time when the participants began to enter college (2008 to 2010). For the 57 students (21% of college goers) who enrolled in more than one four-year college, a cross-college average was computed for the same three years.

A principal components analysis of the four indices of college selectivity yielded a single component (eigenvalue=3.39) that explained 85% of the variance. The four items all loaded on this component (loadings ranged from 0.44 to 0.53). Accordingly, we standardized each of the four indices and averaged them to form the final composite selectivity index for each student. The resulting scale had strong reliability [α=0.93].

Baseline Covariates

Pre-intervention Performance

As previously reported (*2*-*3*), there were three measures of pre-intervention performance for Study 2. These were a standardized measure of pre-intervention performance in 7th grade, GPA in core courses in 6th grade, and a standardized measure of state achievement test performance in 6th grade. Data for each prior performance indicator were obtained either from teachers' official gradebooks (in the case of 7th grade pre-intervention performance) or from official school records (all other indicators). As in Study 1, we simplified analyses by averaging all available indicators for each participant, after each had been standardized.

Again, as in Study 1, minority students had significantly lower baseline performance [M=-0.28, SD=0.83] than White students [M=+0.36, SD=0.74, $t(337)$=7.53, P<0.001]. For this reason, we mean-centered the composite on zero for each ethnic group, to preserve baseline differences in performance between the two focal ethnic groups. Representing prior performance with a single composite enabled us to use a parallel analytic model across the two studies and to retain all students with missing values on any of the pre-intervention performance indicators. In Study 2, this included one student who was missing 6th grade GPA and two students who had missing values for pre-intervention standardized test performance. Alternative models in which the baseline performance indicators were controlled separately and/or missing values were imputed using regression-estimated means (as done in *3*) did not change the key results, as reported in the Detailed Results section below (see SI Text p. 18).

Mediators

For Study 2, two candidate mediators of college enrollment were tested. These included (1) students' perceived sense of adequacy or belonging in the last two grades of middle school (i.e., 7th and 8th grade, after the commencement of the intervention), an indicator of their self-perceived "goodness of fit" in school; and (2) students' average GPA in the same last two grades of middle school. In addition to formal mediation, we tested associations between college outcomes and (3) whether or not students were placed in the remedial track in the last two grades of middle school (assignment to remediation was too rare to permit mediation tests). For each measure, affirmation had been shown to have a positive effect on the minority students in this study, as documented in previous reports (*3*, *6*). Each of these three measures is described below.

Academic Belonging in Middle School

As described in the 2009 report (*3*), a survey was used to measure academic "adequacy" or belonging. A scale that measured academic belonging was one of several scales included in the survey. This scale gauged students' self-perceived "goodness of fit" in school and consisted of two subscales. One subscale included five items to assess *self-perceived social belonging in school* (e.g., "People in my school accept me"). The second subscale included four items to assess *self-perceived ability to succeed in school* (e.g., "I know what I need to do to succeed at [school name]"). In order to shorten the scale due to time constraints, students in Cohort 3 were not presented with one item that had been found in later analyses to have the lowest correlation with the overall scale mean [R=0.49, P<0.001]. Thus, this item was not available to be included in the calculation of the composite scale score for this cohort. Students responded to each item using a 6-point scale ranging from 1(*strongly disagree*) to 6(*strongly agree*). As reported in *6*, the reliability of the entire scale was found to be high, $\alpha \geq 0.76$ at each assessment.

In the first year of the study (7th grade), academic belonging was assessed twice, once at the beginning of the academic year in the fall semester before the commencement of the intervention (time 1), and once at the end of the academic year in the spring semester (time 2). In the second year (8th grade for most students), academic belonging was again assessed twice, once at the beginning of the fall semester (time 3) and once at the end in the spring semester (time 4). In the second year, one cohort was assessed only at the end (not at the beginning), due to logistical constraints. Thus, baseline belonging and two measurements of post-intervention belonging were assessed for all three cohorts. An additional third measurement of post-intervention belonging, in the fall of year 2, was assessed for two cohorts.

Of the 339 African American and White students in this study, all but 10 (97% of the sample) completed the survey at time 1 and time 4, and all but 13 (96% of the sample) completed it at time 2. For time 3, offered to only two of the three cohorts, 231 students (68% of the sample) completed the assessment; however, this represented 98% of students in these two cohorts. We calculated post-intervention belonging in two ways: our primary analysis (referenced in the main text) averaged any available post intervention time points for each participant, including time 3; as a check on robustness, we also averaged scores at time 2 and time 4 only, because, as noted, one cohort did not complete the survey at time 3. This led to a total of 337 students (99% of the sample) with post-intervention belonging scores and 328 students (97% of the sample) with both post intervention belonging scores and baseline belonging scores.

Grade Point Average (GPA) in Middle School

The second mediator tested was the primary outcome in the 2009 report (*3*), core-course GPA in middle school. It consisted of the average GPA students earned in core courses (Math, Social Studies, Science, and English) in each of the four semesters of the 7th and 8th grade. It was available for all participants in the sample and was obtained from official records.

Remedial Placement in Middle School

As described in the 2009 report (*3*), there were three types of remediation indices provided by the school over the life of the project: (1) placement in a special school or tutoring program to help students catch up academically; (2) classification of a child as exhibiting learning/emotional/behavioral difficulties interfering with school work, typically accompanied by enrollment in a special assistance program; and (3) retention (being "held back") in either 7th or 8th grade. These three indices were used to form a dichotomous variable. A score of 1 was assigned to students who were classified into any one of these three categories and a score of 0 was assigned to the remaining students. As noted above, we assessed the association of remedial placement with college outcomes, but its rarity in the sample prevented us from assessing its role in mediating the effect of the affirmation intervention on these outcomes.

Was Random Assignment Successful?

Yes. Table S4 shows the means for the key baseline covariates by condition, for the Study 2 sample as a whole. T-tests revealed no significant differences in ethnicity, gender, or baseline performance by condition [ts<1, Ps>0.35]. Nor were there significant condition differences on any baseline variable among the focal ethnic group, African Americans [ts<1.35, Ps>0.17]. When each baseline covariate (the pre-intervention performance composite and gender) was regressed separately on ethnicity, condition, and ethnicity × condition in the full sample, there were no significant effects of condition [zs/ts<1, Ps>0.30] and no significant ethnicity × condition interactions [zs/ts <1.05, Ps>0.29]. The main effect of ethnicity was nonsignificant for baseline performance [t=0.75, P=0.45] but was significant for gender [z=2.50, P=0.012], indicating African Americans in the sample were more likely than Whites to be female. However, this did not bias treatment effects as gender did not differ by condition, either overall (as noted above) or in the sample of African Americans (see below). Exploratory analyses that include a gender × condition interaction in addition to an ethnicity × condition interaction are discussed in the Detailed Results section (SI Text p. 24).

In the subsample of African Americans in particular, there were no differences in either baseline performance or gender as a function of condition [zs/ts <1.35, Ps>0.17]. Applying the same omnibus technique used in Study 1 yielded similar null results for condition [ts<1.05, Ps>0.30] and its interaction with ethnicity [ts<1.08, Ps>0.28] in the full sample, and for condition in the subsample of African Americans [ts<1.40, Ps>0.17].

Data Analyses

<u>Analytic Model</u>

The same multiple regression model used in Study 1 was used in Study 2. It tested both the ethnicity × condition interaction and the treatment effect for each of the two ethnicities, computed as a simple effect of condition in the full model. Again, the gender × condition and gender x ethnicity terms were excluded from the final model because they did not consistently explain variance; the one exception is noted in Detailed Results section below (see SI Text p. 24). Ordinal regression was used to analyze the trichotomous outcome related to college enrollment, logistic regression to analyze the dichotomous outcomes related to college enrollment, and linear regression to analyze the continuous measure of college selectivity.

<u>Analytic Approach</u>

The analytic approach was the same as in Study 1, except that the ethnic minority group was African Americans rather than Latino Americans.

Detailed Results Section

Additional Figures for Main Text Results

In the main text we provided only a subset of the figures depicting the key results. Figures S1-S5 provide graphical depictions of results not illustrated in the main text, along with raw means (left panel of each figure) for each outcome as a function of condition.

Academic Belonging and GPA in Middle School Partially Mediate the Positive Effect of Affirmation on African American Students' College Enrollment in Study 2

Given prior findings that affirmed African Americans in Study 2 expressed higher levels of academic belonging than non-affirmed African Americans (*6*), we hypothesized that academic belonging in middle school might explain some of the positive intervention effects on college enrollment. To test this hypothesis, we used a model in which students' post-intervention sense of belonging in middle school served as a mediator of the affirmation effect for African American students on the two dichotomous college outcomes. The belonging mediator was the average of all three post-intervention measurement occasions, from the spring of year 1 to the spring of year 2. We also controlled for students' pre-intervention level of belonging (fall of year 1), centered on the grand sample mean (there was no ethnicity gap in baseline belonging). Using post-intervention belonging at two of three measurement occasions (because, as noted, one cohort was not assessed at time 3) yielded similar results.

The model included the following terms: the main effects of ethnicity, condition, gender, and baseline performance; the two-way interaction between ethnicity and condition; and the main effects of baseline and post-intervention belonging, with the latter designated as the mediator. Ethnicity was set as 0 for African Americans, 1 for Whites, so that the full model would permit us to test mediation of the simple effect of condition among African Americans. Through this moderated mediation test, the goal was to determine whether post-intervention belonging explained why African American students in the affirmation condition had better college outcomes than those in the control condition, over and above baseline belonging and baseline academic performance.

The model that tested the mediating role of middle school GPA was identical except that the post-intervention belonging variable was replaced with middle school GPA, and the covariate corresponding to baseline belonging was removed. Analyses were conducted in STATA, using a routine suited for mediational testing with dichotomous outcomes. Tests of significance and confidence intervals are based on 5,000 bias-corrected bootstrap samples. All but 11 participants were included in the belonging mediation analyses. Two were excluded because they had no post-intervention belonging data, and nine additional subjects were excluded because they were missing baseline belonging data, resulting in a final sample of 328, as reported above. Various strategies for handling missing baseline belonging data, including substituting the grand sample mean or means obtained from multiple imputation, yielded similar results, for the larger sample of 337 who had post-intervention belonging data.

Figure S6 displays the path coefficients from an analysis of whether middle school belonging mediated the condition difference in college enrollment (2- or 4-year versus none) for African American, but not White students. We also provide the proportion of the total effect mediated (indirect effect divided by the total effect), though it can be only imprecisely estimated when the total effect is small. For African Americans, the effect of condition on college enrollment was partially mediated (22.7%) by belonging in middle school [B=0.057, z=2.08, *P*=0.038], 95% CI [0.0135, 0.1219]. By contrast, belonging did not mediate affirmation's effect on African Americans' enrollment in a 4-year college in particular, relative to a 2-year college or no college [*B*=0.006, *z*=0.27, *P*=0.785], 95% CI [-0.0356, 0.0553].

Figure S7 displays the path coefficients for an analogous mediation analysis using middle school GPA, rather than belonging, as the mediator (involving all 339 participants). Relative to academic belonging, middle school GPA was a weaker (14.1%) and non-significant mediator of the effect of affirmation on African Americans' enrollment in any college [*B*=0.042, *z*=1.72, *P*=0.086], 95% CI [0.0038, 0.0993]. However, it did partially mediate (32.6%) the positive effect of affirmation on whether African Americans specifically enrolled in a 4-year college, relative to a 2-year college or no college [*B*=0.069, *z*=2.82, *P*=0.005], 95% CI [0.0273, 0.1246].

Placement in Remedial Courses in Middle School was Associated with Lower Levels of College Enrollment in Study 2

As reported in the main text, assignment to the remedial track was associated with a lower probability of enrollment in a 2-year or 4-year college. Nine African Americans and three Whites in Study 2 were assigned to a remedial program in middle school (one additional student in remediation was from another ethnic group). As reported in the 2009 foundational study (*3*), affirmed African Americans were less likely to be placed in remediation than nonaffirmed African Americans. As reported in the main text, we tested whether deflection from this institutional track could account for some of the long-term effects of affirmation on college enrollment. Because only a small number of students were assigned to remediation (3.5% of the sample of African Americans and Whites), we conducted regression rather than formal mediation analysis (i.e., the rarity of remediation prevented model convergence in a full mediation model). We tested a simple multiple regression model that included main effects of ethnicity, gender, and assignment to remediation (0=no, 1=yes) but excluded the main effect of condition and ethnicity × condition interaction. The goal was simply to assess whether the main effect of remediation predicted college enrollment.

Analyses support the hypothesis that affirmation facilitated college entry in part by reducing the probability of assignment to the remedial track. Remedial placement was associated with a significantly lower probability of enrolling in any college [*B*=-2.31, *z*=-3.60, *P*<0.0001, *OR*=0.099] and a lower probability of enrolling in a 4-year college specifically [*B*=-2.79, *z*=-2.65, *P*=0.008, *OR*=0.061]. Focusing on the subsample of African Americans, who comprised 75% of those assigned to remediation, the effect of remediation on enrolling in any college remained significant even when baseline performance [*B*=-1.85, *z*=-2.05, *P*=0.041, *OR*=0.16] or, alternatively, middle school GPA [*B*=-1.99, *z*=-2.23, *P*=0.025, *OR*=0.14], was controlled. By contrast, the effect on 4-year college enrollment in the African American subsample did not remain significant when either of these two predictors was controlled [*z*s<0.80, *P*s>0.45].

Placement in Remedial Courses in Middle School was Associated with Lower Academic Belonging in Middle School in Study 2

Beyond its effects on college outcomes, remedial placement was also associated with lower belonging in middle school, consistent with the stigmatizing effect of some forms of remediation. We tested a regression model in which post-intervention middle school belonging was predicted by pre-intervention belonging (time 1) and main effects of ethnicity, gender, and remediation, for the full sample of students with non-missing values for both pre and post belonging variables (*N*=328). As reported in the main text, assignment to the remedial track was a significant negative predictor of belonging in middle school, defined as the average of belonging in the spring of year 1 and the fall and spring of year 2 (times 2, 3, 4) [B=-0.43, t(323)=-3.07, P=0.002, d=-0.78]. This effect held even when baseline performance [B=-0.41, t(322)=-2.79, P=0.006, d=-0.75], or middle school GPA [B=-0.34, t(322)=-2.38, P=0.018, d=-0.62] was also controlled. Virtually identical results were obtained when belonging included only timepoints when all cohorts were assessed (times 2 and 4), and for various treatments of missing values for either of the two ways of computing belonging. Though correlational, results suggest that placement in remediation was associated with decrements in academic belonging in the remainder of middle school.

Robustness of Basic Analytic Effects to Various Model Specifications

Were Treatment Effects Robust to Controlling for Main Effects of Cohort and Teacher Team?

Yes. In each study, students occupied different clusters, as reported in the foundational papers. In Study 1, the clusters were based on students' cohort, i.e., their school grade at the time of intervention: 6th, 7th, or 8th (*1*). In Study 2, the clusters were based on two classifications: their cohort and teacher team at the time of intervention (*2-3*). In the latter case, cohort reflected the specific school year that students took part in the study (each cohort received the intervention in one of three consecutive years). One of three teacher teams taught the students in each cohort (the same three teacher teams taught 7th grade at the school site during the years of intervention administration). While all of these clustering variables were included in analyses of the foundational studies (*1-3*), we excluded them in the primary analyses reported in the main text in order to create a simple, common model across outcomes and studies.

To ensure basic effects reported in the main text were unchanged when controls for these clusters were introduced, supplemental analyses were conducted that included main effects of cohort and, additionally in Study 2, main effects of teacher team. Two orthogonal contrast codes represented the three cohorts of students in each study (with one code assigning weights of +1,+1,-2 and the other assigning weights of -1,+1,0 to each of the three cohorts, respectively). In Study 2, two additional contrast codes were used to represent teacher team, using the same numerical values given above. The codes were defined in this way—orthogonal and balanced on zero—so that the meaning of the intercept would not change relative to the model reported in the main text and so that lower-order interactions would remain interpretable when testing whether these contrasts moderated condition effects.

Tables S5-S7 show the coefficients for models with and without the clustering variables for both studies. In Study 1, adding the block of cohort (i.e., grade) dummy variables improved model fit for course levels [ΔR^2=0.041, F(2,152)=6.84, P=0.0014] and for enrollment in the mainstream high school [χ^2 (1)=4.4705, P=0.0345], but not for enrollment in AVID or remedial clinics [χ^2s<0.55, Ps>0.45]. In Study 2, adding the block of teacher team dummies improved model fit for enrollment in a 4-year college [χ^2(2)=6.5068, P=0.0386] and for college selectivity [ΔR^2=0.028, F(2,191)=4.04, P=0.0191], but not for enrollment in some college versus none, [χ^2(2)=4.096, P=0.129]. Once teacher team was controlled, adding the block of cohort dummies did not significantly explain further variance [Ps>0.25], though there was a trend in better model fit for college selectivity [P=0.065].

Even when the block of cohort and teacher team variables explained a significant proportion of the variance, the critical ethnicity × condition interaction always remained intact [Ps<0.040], and in some cases became stronger.

Were Treatment Effects Moderated by Cohort or by Teacher Team?

No. Beyond simply controlling for the main effects of cohort, we checked for cohort × ethnicity × condition interactions, in a model including all main effects of each cohort contrast code and all relevant two-way and three-way interactions between the cohort contrast codes on the one hand and ethnicity and condition on the other. In Study 2, we additionally checked for teacher team × ethnicity × condition interactions, using an analogous regression model that included the two orthogonal contrast codes representing teacher team.

In Study 1, adding the block of two- and three-way interactions between the two cohort contrast codes, ethnicity, and condition did not improve model fit for any outcome [*P*s>0.10]. In Study 2, treatment effects were not moderated by either teacher team or cohort. Adding the block of two- and three-way interactions between either teacher team or cohort on the one hand, and ethnicity and condition on the other, to a model already containing the main effects, did not explain significant additional variance in any outcome [*P*s>0.22].

Were Treatment Effects Robust to Alternative Models of Representing Prior Performance?

Yes. We tested models with different methods of representing pre-intervention performance. As noted previously, there were two such indicators in Study 1: pre-intervention GPA and standardized test scores. Study 2 had these two indicators plus a third: pre-intervention performance in the class where the intervention was administered (*2*-*3*). We tested models controlling for the main effect of each performance indicator separately; controlling for different combinations of the performance indicators; and centering each performance indicator on the grand mean rather than on the mean for each ethnic group. In Study 2, we additionally tested different strategies for handling missing values for each indicator, either imputing them using regression-estimated means or using only participants with all available data. All of these alternative methods had negligible effects on the model and in no case altered the statistical significance of any key result.

To check for whether treatment effects were moderated by baseline performance, we added two- and three-way interactions between the pre-intervention performance composite, ethnicity, and condition to the basic model for the full sample. With one exception, adding this block of three interaction terms did not improve model fit [*P*s>0.25]. The exception was enrollment in the mainstream high school in Study 1 [χ^2 (3)=9.173, *P*=0.027]. In this case, the three-way interaction was just significant [*B*=-1.72, *z*=-1.96, *P*=0.050], suggesting that affirmation tended to have greater benefit for low-performing Latino Americans on this outcome. However, because this interaction emerged only for one of six outcomes across both studies, and moreover the interaction between prior performance and condition was not significant in the subsample of Latino Americans [*B*=-2.60, *z*=-1.72, *P*=0.086], it should be interpreted with caution. Interestingly, while our previous report had found affirmation to benefit low-performing African Americans' GPA more than that of high-performing African Americans (*3*), this moderation was not apparent for African Americans on more distal outcomes (e.g., college enrollment or selectivity). It is possible that high-performing African Americans benefited in other ways that simply did not manifest in improved GPA; for instance, they did show improved belonging as a result of the affirmation in their last year of middle school (*6*), and belonging mediated college enrollment, as discussed above (SI Text pp. 15-16).

Do Separate Subsample Analyses of Affirmation Effects in Each Ethnic Group Yield the Same Results as Simple Effects Tests of Affirmation?

Yes. In the main text, we present treatment effects derived from simple effects tests conducted in the full sample, with the model that accounted for the ethnicity × condition interaction. Using an alternative strategy of computing the effect of condition separately for the two ethnic groups yielded similar results. For the subsample of Latino Americans in Study 1, the effects of condition for all outcomes remained significant [*zs*/*ts*>2.40, *P*s<0.016]. Among Whites, there was still no effect of condition [*zs*/*ts*<1.55, *P*s>0.12]. Likewise, among the African American subsample in Study 2, the effects of condition for all outcomes remained significant [*zs/ts*>2, *Ps*<0.042]. Among Whites, there was still no effect of condition [*zs/ts*<1.10, *P*s>0.26].

Were Treatment Effects Robust to Alternative Methods of Computing Outcomes?

Yes. Two of the four outcomes in Study 1 required some judgment in defining their values. Below we report credible alternative methods of constructing each outcome, and show that these alternatives do not change the key effects.

Course Difficulty Scores

In the main text we reported the effect of affirmation on course difficulty scores in the school grade in which students were enrolled two years after the intervention, Year 3. In this year, as noted previously, course variability was relatively high for most students. To check the robustness of the effect, we also computed course difficulty as the annual average of course levels over the two years after the intervention (Years 2 and 3 averaged). This permitted more participants to be retained in the analysis (94% of the original sample), as it included those participants who had course levels data for Year 2 but not for Year 3. Constructing course difficulty in this way yielded virtually identical results. The ethnicity × condition interaction remained significant [B=0.19, t(168)=2.38, P=0.018]. For Latino Americans, there was a significant positive effect of condition [B=0.31, t(168)=2.48, P=0.014, d=0.42]. The effect of condition for Whites was still not significant [B=-0.08, t(168)=-0.76, P=0.446, d=-0.11].

Because we examined remedial clinics as a separate outcome, we also assessed whether the effect of affirmation on course difficulty scores in Year 3 remained intact when remedial clinics were excluded from the composite. We did so by testing the course levels variable at the high school level only: first, by excluding students whose course difficulty scores reflected participation in remedial clinics in Year 3 (the 6th grade cohort, then in 8th grade), and, second, by using only high school courses (the average of 9th and 10th grade course levels for the 8th grade cohort, and 9th grade course levels for the 7th grade cohort). In each case, the ethnicity × condition interaction remained significant [B=0.24, t(106)=2.75, P=0.007; B=0.25, t(108)=3.03, P=0.003, respectively]. The positive effect of affirmation was significant for Latino Americans [*t*s>2.15, *Ps*<0.034]; as in the main analysis, there was no effect for Whites [*t*s<1.75, *P*s>0.086]. Thus, the effect of affirmation condition on course difficulty scores was unaffected by the removal of remedial clinics as a component of the course levels variable.

Finally, we also tested for the effect of condition on course difficulty scores in Year 3 (the outcome reported in the main text) when additionally controlling for pre-intervention course difficulty scores. The latter represented the difficulty of students' courses at the beginning of the school year immediately before the grade in which students began participating in the intervention (Year 1 for the younger two cohorts and Year 0 for the oldest cohort). Doing so strengthened both the ethnicity × condition interaction [B=0.28, t(152)=3.37, P=0.001] and the simple effect of condition among Latino Americans [B=0.42, t(152)=3.30, P=0.001, d=0.56].

The simple effect of condition among Whites remained non-significant [*B*=-0.14, *t*(152)=-1.31, *P*=0.191, *d*=-0.19].

Enrollment in the Mainstream High School

Effects on this outcome were robust to excluding participants who were lost to attrition, i.e., those who left the school district. In the analyses reported above and in the main text, we made the assumption that participants who left the district were not in a mainstream high school in Year 3. We felt this to be a safe assumption because leaving the school district can represent dropping out of school and because these students' GPA in the previous two school years (Years 1 and 2) (*M*=2.48, *SD*=0.81) was similar to the GPA of students who enrolled in alternative high schools (*M*=2.35, *SD*=0.71). Moreover, both groups had lower GPAs than students who enrolled in the mainstream high school for the district (*M*=2.91, *SD*=0.84) [*t*s>2.00, *P*s<0.043].

Nevertheless, to assuage any potential concerns about this assumption, we repeated the analyses using a modified version of this outcome that omitted students who left the school district. The ethnicity × condition interaction remained significant [*B*=1.95, *z*=2.35, *P*=0.019], as did the simple effect of condition among Latino Americans [*B*=2.39, *z*=2.00, *P*=0.045, *OR*=10.89]. The simple effect of condition remained non-significant for Whites [*z*<1.40, *P*>0.18].

College Enrollment

In a supplemental analysis for Study 2, we assessed whether affirmation increased the probability of enrolling in a 4-year college versus a 2-year college, among the subset of college-goers (*N*=274). It did not. The ethnicity × condition interaction was not significant [*B*=0.27, *z*=0.84, *P*=0.398] and though the effect of affirmation among African Americans was in the expected direction, it too was non-significant [*B*=0.63, *z*=1.38, *P*=0.167, *OR*=1.88].

Did the Intervention Reduce Achievement Gaps Between Ethnic Groups?

It did. While these results were summarized in the main text, we report the detailed results here. In each case, the intervention reduced achievement gaps by improving the outcomes of minority students. Examining raw means, affirmation reduced the gap between Latino Americans and Whites in Study 1 for two of the four primary outcomes (these analyses were not possible for AVID enrollment, as this program enrolled almost exclusively Latino Americans), and reversed it for a third. The latter effect is consistent with previous research suggesting that once stereotype threat is allayed, minority students sometimes exceed Whites in their performance. For dichotomous outcomes, we report the gap between Whites and minorities as both an odds ratio (White students' odds of obtaining a given outcome divided by minority students' odds of obtaining the same outcome, obtained by defining ethnicity as 0 for minority students and 1 for White students, for a given condition defined as 0) and percentage point difference. For positively valenced outcomes (outcomes for which a higher level or probability is a positive outcome; all outcomes except enrollment in remedial clinics), an odds ratio greater than 1 and a positive percentage-point difference indicates an achievement gap favoring Whites. For negatively valenced outcomes (enrollment in remedial clinics), an odds ratio less than 1 and a negative percentage-point difference indicates an achievement gap favoring Whites. For continuous outcomes, we report the gap in standard deviation units.

In Study 1, the raw gap in course difficulty was 1.49 standard deviations in favor of Whites in the control condition but only 0.93 standard deviations in the affirmation condition, a

38% decrease. For enrollment in remedial clinics, the raw White to Latino American odds ratio was 0.052 in the control condition (-60.2 percentage points, with Latino Americans having higher levels of remedial enrollment) and 0.343 (-21.1 percentage points) in the affirmation condition. For enrollment in the mainstream high school, the raw odds ratio of 6.800 in the control condition (37.2 percentage points in favor of Whites) declined to 0.505 (-9.0 percentage points) in the affirmation condition, reversing in favor of Latino Americans.

Examining the raw means in Study 2, affirmation reversed the gap for two of the three primary outcomes and reduced it for the third. The control-condition White to African American odds ratio in some college enrollment versus none was 1.48 (6.4 percentage points in favor of Whites); in the affirmation condition, the odds ratio reversed in favor of African Americans to 0.48 (-10.8 percentage points). For enrollment in 4-year college versus 2-year or no college, the odds ratio again reversed in favor of African Americans, from 1.62 (11.8 percentage points in favor of Whites) in the control condition to 0.88 (-3.1 percentage points) in the affirmation condition. For college selectivity, the raw control gap of 0.64 standard deviations declined by 72% to 0.18 standard deviations in the affirmation condition.

For more precise estimates of reduction in the gap attributable to the affirmation, we calculated the covariate-adjusted difference between ethnic groups in the control condition relative to the affirmation condition, by comparing the main effect of ethnicity when the control condition was defined as 0 (and affirmation as +1) to the main effect of ethnicity when the affirmation condition was defined as 0 (and control as +1). The results of these analyses were consistent with those of the raw achievement gaps reported above. In Study 1, the adjusted achievement gap favoring White students in the control condition was highly significant for all outcomes [*z*s/*t*s>2.90, *P*s<0.004] but was reduced or reversed in the affirmation condition. For course difficulty, the gap favoring White students was 1.53 standard deviations in the control condition [t=9.33, P<0.001]. Though still significant [t=5.08, P<0.001], the gap favoring Whites was only 0.83 standard deviations in the affirmation condition, a 46% decrease. For remedial clinic enrollment, the gap favoring White students was an odds ratio of 0.037 (-64.0 percentage points, with Latino Americans having higher levels of remedial enrollment) in the control condition [z=-4.39, P<0.001]. The gap was reduced to an odds ratio of 0.309 (-22.1 percentage points) in the affirmation condition and became non-significant [z=-1.82, P=0.068]. For mainstream high school enrollment, the gap favoring White students was an odds ratio of 6.39 (36.4 percentage points) in the control condition [z=2.95, P=0.003], but reversed to an odds ratio of 0.47 (-9.8 percentage points) in favor of Latino American students (non-significantly) in the affirmation condition [z=-1, P=0.316].

In Study 2, affirmation eliminated or reversed the gap in covariate-adjusted college outcomes in favor of African Americans, as found with the raw means. In the control condition, the ethnicity gap favoring White students was significant for all outcomes [*z*s>2.00, *P*s<0.04], except for enrollment in any college [z=1.34, P=0.18]. In the latter case, the control gap was still an odds ratio of 1.75 (8.1 percentage points) in favor of White students. By contrast, in the affirmation condition, the gap for this outcome reversed to favor African Americans by 12.2 percentage points (*OR*=0.34), a significant difference [z=-2.27, P=0.023]. For 4-year college enrollment, the gap favoring White students was an odds ratio of 2.18 (18.9 percentage points) [z=2.06, P=0.039], but this gap was eliminated in the affirmation condition [z=-0.80, P=0.426]. In fact, the gap (non-significantly) reversed to favor African Americans by 7.2 percentage points (*OR*=0.73). Among attendees of 4-year colleges, the gap in college selectivity between African American and White students was 0.71 standard deviations in the control condition [t=4.06,

P<0.001], but only 0.046 standard deviations in the affirmation condition [t=0.27, P=0.788], a 94% reduction.

Did African American Students Randomly Assigned to Receive Higher Levels of Treatment Dosage Experience Additional Benefit on College Outcomes?

No. As documented in our earlier report (*3*), in Study 2, after the first year of the intervention in 7th grade, half of the students in the affirmation condition were randomly assigned to receive two to four additional classroom administrations of the intervention in the second year when they were in 8th grade. The specific number of doses varied by cohort due to pragmatic constraints (*3*): those in the low dosage condition received 3 to 5 doses in the first year only, depending on cohort; those in the high dosage condition received 7 to 8 doses over 2 years, depending on cohort.

Two orthogonal contrast codes were created to test the effect of the dosage manipulation. The first tested the effect of affirmation condition, assigning -2 to the control condition and +1 to each of the two affirmation dosage conditions (low and high). The second tested the effect of dosage specifically, assigning -1 to affirmed participants in the low treatment dosage condition, +1 to affirmed participants in the high treatment dosage condition, and 0 to the control condition. The main effects of each contrast code were added to the basic model in place of the main effect of condition. The single ethnicity × condition term in the basic model was replaced by two terms to represent the interaction of each of these contrast codes with ethnicity.

Accounting for dosage condition did not alter the basic affirmation effects. The ethnicity × affirmation interaction (i.e., the interaction between ethnicity and the first contrast code) remained significant in all analyses [zs/ts>2.05, Ps<0.039]. In addition, the simple effect of affirmation remained significant for African Americans [zs/ts>2.20, Ps<0.026] and non-significant for Whites [zs/ts<1.05, Ps>0.29] for all outcomes.

Consistent with our earlier report (*3*), higher dosage did not improve outcomes. If anything, for African Americans, low dosage seemed more effective for enrollment, specifically in 4-year college. On this outcome alone, there was a significant ethnicity × dosage interaction (an interaction between ethnicity and the second contrast code) [B=-0.42, z=-2.11, P=0.034, OR=0.65]. Among African American students, those in the high dosage condition showed a trend to enroll in 4-year colleges at a lower rate than those in the low dosage condition (56% vs. 79%) [B=-0.55, z=-1.83, P=0.068, OR=0.33]. Among Whites, there was no effect of dosage [B=0.30, z=1.15, P=0.251, OR=1.82]. The ethnicity × dosage interaction was not significant for any other outcome [zs/ts<1.50, Ps>0.13]. The present study indicates that additional affirmations given in the final year of middle school yielded no incremental benefit on long-term educational outcomes. Previous research suggests that affirmation timing rather than dosage may be the more critical variable, as early outcomes have disproportionate influence in a recursive process (*3, 6*). Nevertheless, it is possible that additional doses of affirmations timed to later threatening transitions, such as the transition to high school or college, might be beneficial.

Did the Tailoring of Treatment Exercises Play a Role in Treatment Effects?

In both studies, students in the affirmation condition received 1-2 tailored doses that singled out a value they had previously indicated to be personally important. It is possible that some of the positive effects of affirmation could be explained by the feelings of individuation that may have resulted from this procedural feature (e.g., students could have interpreted such individuation as a sign that their teacher cared about them). If this relatively small gesture turns out to contribute to affirmation effects, this would be an interesting finding unto itself. It would suggest that small gestures of recognition can facilitate affirmation effects.

Nevertheless, there is no theoretical or empirical reason to suspect that this tailoring played a major role in driving affirmation effects. First, students in both the control and treatment conditions generally expected teachers to read the exercises because teachers presented the activities to students as regular classroom exercises (per the scripts we provided to teachers).

Second, prior studies have provided evidence that the effectiveness of affirmation does not hinge on tailoring. In virtually all laboratory and field experiments, affirmations were tested without the personalized affirmation intervention (see *7*, for a review), including field experiments in school studies (e.g., *8*). As one example, in Study 2 of *1*, the affirmation intervention significantly increased Latino Americans' GPA even though no personally tailored affirmation was provided. Likewise, in the original study involving African Americans, as documented in *2*, significant treatment effects on GPA—a key driver of 4-year-college enrollment--emerged before individuated affirmation was administered later in the year. The GPA data in *2* pertained to the first academic terms (quarters 1 and 2), before the tailored affirmation intervention took place.

To clarify any role of tailoring in the current paper, we compared the effect of condition on minority students' average GPA in quarters 1 and 2 of year 1 (in the fall semester) with the effect of condition on their average GPA in quarters 3 and 4 of year 1 (in the spring semester). We found that the affirmation effect on minority students' GPA in both of the reported studies was similar in magnitude in the two terms before the personalized affirmation was offered [Study 1: B=0.17, $t(128)$=1.40, P=0.163, d=0.22; Study 2: B=0.25, $t(213)$=2.89, P=0.004, d=0.26] as it was in the two terms after [Study 1: B=0.16, $t(128)$=1.34, P=0.184, d=0.22; Study 2: B=0.21, $t(211)$=1.86, P=0.064, d=0.22]. This analysis leveraged the fact that the tailored affirmation(s) took place in the third term of the first year, for four of the six cohorts across Studies 1 and 2. (In Study 1, dose 4 for the 6th and 7th grade cohorts was tailored. In Study 2, doses 3 and 4 for cohort 1 and dose 4 for cohort 3 were tailored). On the whole, these analyses suggest that a key outcome of affirmation effects, GPA in middle school, a key driver of affirmation's effect on African American students' enrollment in 4-year college (see earlier discussion on p. 15-16 of SI Text), was unaffected by tailoring.

For two of the six cohorts, it was not possible to assess the effect of tailoring within cohort, because it was either offered too early in year 1 (Study 1's 8th grade cohort) or was not offered at all, due to practical constraints (Study 2's cohort 2). However, the fact that one of the Study 2 cohorts did not receive a personalized affirmation in year 1 offered the opportunity to conduct further analyses to understand the possible role of tailoring. We assessed whether membership in this cohort moderated the treatment effect for minority students, that is, whether this treatment effect was larger for those in Study 2 who did receive a tailored affirmation in year 1 (cohorts 1 and 3) versus those who did not (cohort 2). To test this, we added to the final model a dichotomous predictor that distinguished between membership in cohort 2 versus membership in one of the other two cohorts, along with two- and three-way interactions between this variable and ethnicity, gender, and condition, with ethnicity defined as 0 for American Americans and +1 for Whites. For both middle school GPA and belonging, the primary mediators of college outcomes (see p. 15-16 of SI Text), neither the cohort × condition nor cohort × condition × ethnicity interaction was significant [*ts*<1.20, *Ps*>0.23].

After year 1, half of all treated students in all three Study 2 cohorts were offered a tailored dose in year 2. Consequently, there were too few participants who did not receive tailoring in any cohort to enable comparison of college outcomes by whether participants received a tailored dose or not.

Although we cannot rule out that tailoring the affirmation exercises had some enhancing effect, the preponderance of theoretical and empirical evidence suggests that tailoring was not a necessary factor in the effects observed both here and in previous research.

Exploratory Analyses

<u>Was Affirmation Also Effective for Female Students?</u>

Given prior theory and research (see *9* for a review), we also explored whether the effects of affirmation extended to other identity-threatened groups for whom the transition to middle school may be particularly stressful. We focused on female students, given gender stereotypes in math and science and the challenges girls in particular face as they transition to middle school.

Although exploratory, affirmation appeared to benefit females on some outcomes. We tested the gender × condition interaction for both studies, by adding relevant terms to the basic model. As the main effect of gender (contrast coded with values of -1 for males and +1 for females) was already in the basic model, we added its 2-way and 3-way interactions with ethnicity and condition.

In Study 1, the gender × condition interaction was significant for course difficulty scores [*B*=0.21, *t*(151)=2.42, *P*=0.017]. This interaction reflected a significant simple effect of condition for women [*B*=0.27, *t*(151)=2.42, *P*=0.017, *d*=0.36]. In Study 2, although the gender × condition interaction on the key outcome of enrollment in any college was only a trend [*B*=0.60, *z*=1.85, *P*=0.064, *OR*=1.82], the positive affirmation effect for women was significant [*B*=1.11, *z*=2.30, *P*=0.022, *OR*=3.03]. Affirmation had no effect for men in either case [*zs/ts*<1.20, *P*s>0.24]. The gender × condition interaction was non-significant for all other outcomes across both studies [*zs/ts*<1.75, *P*s>0.083].

<u>Did English Language Proficiency Moderate the Effect of Affirmation in Study 1?</u>

Because many Latino students in Study 1 were, as noted previously, immigrants from families whose primary language was not English, it seemed important to assess whether the impact of the affirmation intervention extended to those with relatively less English proficiency. English proficiency might be important both proximally, for following the instructions in the affirmation exercise, and longitudinally, for doing well in advanced courses later in middle school and high school. We obtained a measure of English proficiency by constructing a composite based on the average of students' responses to five items related to language use at home and at school, administered in a survey before the intervention (e.g., "It is easy for me to read English"). Because response options varied across items, the scores for each item were standardized before being averaged. We then tested the resulting English proficiency composite (α=0.79) as a moderator of the effect of condition in the subsample of Latino Americans by adding its main effect and interaction with condition to the basic model used for analysis of this subgroup. Because this model included baseline performance as a covariate, this analysis tested whether English proficiency moderated affirmation effects above and beyond prior academic performance.

With one exception, English proficiency did not moderate the effect of affirmation for Latino American students [$zs<1.75$, $Ps>0.081$]. The exception was course difficulty, where there was a positive English proficiency × condition interaction [B=0.55, $t(61)=2.41$, P=0.019]. This interaction indicated that it was high English proficient (+0.75 SD above the mean) Latino Americans who benefited from affirmation [B=0.63, $t(61)=3.63$, P=0.001, d=0.84], not low English proficient (-0.75 SD below the mean) Latino Americans [B=0.12, $t(61)=0.72$, P=0.474, d=0.16]. While these analyses require further explication in future papers, they suggest English proficiency could be an important moderator of higher-level institutional outcomes for Latino American students.

Supporting Information References

1. Sherman DK, et al. (2013) Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *J Pers Soc Psychol* 104(4): 591-618.
2. Cohen GL, Garcia J, Apfel N, Master A (2006) Reducing the racial achievement gap: A social-psychological intervention. *Science* 313(5791): 1307-1310.
3. Cohen GL, Garcia J, Purdie-Vaughns V, Apfel N, Brzustoski P (2009) Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science* 324(5925): 400-403.
4. Steele CM (1988) The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology Vo1 21*, ed Berkowitz L (Academic Press, New York), pp 261-302.
5. Dynarski SM, Hemelt SW, Hyman JM (2015) The missing manual: Using National Student Clearinghouse data to track postsecondary outcomes. *Educ Eval & Policy Analysis* 37: 53-79.
6. Cook JE, Purdie-Vaughns V, Garcia J, Cohen GL (2012) Chronic threat and contingency belonging: Protective benefits of values affirmation on identity development. *J Pers Soc Psychol* 102(3): 479-496.
7. Cohen GL, Sherman DK (2014) The psychology of change: Self-affirmation and social psychological intervention. *Annu Rev Psychol* 65(1): 333-371.
8. Bowen NK, Wegmann KM, Webber KC (2013) Enhancing a brief writing intervention to combat stereotype threat among middle-school students. *J of Educ Psychol* 105(2): 427-435.
9. Steele CM, Spencer SJ, Aronson J (2002) Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology Vol 34*, ed Zanna MP (Academic Press, San Diego), pp 379-440.

Table S1.

Study 1. Outcome availability across the three years of data collection, by cohort.

Grade Level in Year 1	6th Grade		7th Grade		8th Grade	
	Count	Percentage Retained	Count	Percentage Retained	Count	Percentage Retained
Total Randomly Assigned in Year 1	55	100%	79	100%	51	100%
Total with Course Levels Data in Year 3	48	87%	66	84%	46	90%
Total with Remedial Clinics Data in Year 2 or 3	48	87%	72	91%	--	--
Total with AVID Data in Year 2 or 3*	25	86%	28	88%	--	--
Total with High School Enrollment Data in Year 3	--	--	79	100%	51	100%

Study 1. The table presents the number of students in each cohort for whom data along each outcome are available, both as a count and as a percentage of students in the study. Participants were the same as the final sample in Study 1 of *1*, with the inclusion of one additional student who had been an outlier on the previously assessed GPA outcome but not on the ones featured here (see p. 1 of the SI Text). We define each cohort by their grade level in Year 1 of the study. Row 1 indicates the number of students in each cohort who were randomly assigned to condition (control or treatment) in Year 1. A dash indicates that the outcome was unavailable for anyone in the cohort. For all outcomes, attrition occurred within cohort only when students left the school district. Three participants did not have course levels data for Year 2 but did for Year 3, so they were not considered lost to attrition. Enrollment in AVID and in remedial clinics in middle school was assessed in 8th grade. This corresponded to Year 2 for the 7th grade cohort and Year 3 for the 6th grade cohort; data were unavailable for students in the 8th grade cohort on this outcome as they had left middle school. Enrollment in the district's mainstream high school was assessed in Year 3. Data for the 6th grade cohort were unavailable on this outcome because they were still in middle school. For the 7th and 8th grade cohorts, those who left the district and thereby failed to enroll in the district's mainstream high school were included; thus all students in these two cohorts were analyzed for this outcome.
*AVID was available only for Latino Americans and thus the percentage of the sample retained for that measure is based on the total number of Latino Americans in the sample within each cohort.

Table S2.

Study 1. Grade levels, dosage and attrition within cohort across the three years of data collection.

	Cohort Grade Level			Doses Offered			Number of Students Retained				
	6th Grade	7th Grade	8th Grade	6th Grade	7th Grade	8th Grade	6th Grade	7th Grade	8th Grade	Total	Attrition Rate
Year 1	6th	7th	8th	4	4	2	55	79	51	185	
Year 2	7th	8th	9th	0	0	0	52	74	48	174	
							(3)	(5)	(3)	(11)	-5.9%
Year 3	8th	9th	10th	0	0	0	48	66	46	160	
							(4)	(8)	(2)	(14)	-8.0%
							Number of Students Lost to Attrition				
Average							(3.5)	(6.5)	(2.5)	(12.5)	
Cumulative							(7)	(13)	(5)	(25)	

Study 1. The table presents the grade levels of each cohort during each year of the experiment (left panel), the number of intervention doses administered to each cohort of students during Year 1 (middle panel), and the number of White and Latino American students originally randomly assigned (*N*=185) who were retained each year of the study (right panel), as a function of the grade when they began the intervention (in Year 1). We refer to each cohort by their grade level in Year 1, the primary intervention year. In the right panel, parentheses directly below number retained indicate the number of students lost to attrition within each cohort in a given year. The 8th grade cohort also received 3 doses in Year 0, the pilot year immediately prior to the primary intervention year. The bottom panel provides the average and cumulative loss of participants within cohort across years. The number retained in Year 3 indicates the total number of students within each cohort for which we had course levels data (*N*=160). Three participants did not have course levels data for Year 2 but did for Year 3, so they were not considered lost to attrition.

Table S3.

Study 1. Baseline covariates and number of control or treatment doses by condition.

	Control		Treatment	
	Mean	SE	Mean	SE
Baseline Covariates				
Proportion Latino American	0.43		0.44	
Proportion female	0.55		0.57	
Pre-intervention academic performance	-0.01	0.06	0.01	0.07
Dosage				
Doses completed in Year 1 and Year 2	4.04	0.08	4.14	0.06
Proportion completed of all doses offered	0.94	0.01	0.97	0.01
Completed all doses offered	0.82		0.89	
Completed a tailored dose	0.93		0.96	
N	92		93	

Study 1. The table reports the mean and standard errors of the baseline covariates, as well as the number of control or treatment doses completed, separately for the control (*N*=92) and treatment (*N*=93) groups. Pre-intervention academic performance reflects a composite of two indicators and is mean-centered within ethnicity. For ethnicity and gender, the means represent the percentage of Latino American and female students in each condition, respectively. The last two rows in the Dosage panel indicate the percentage of students in each condition who completed the specified doses.

Table S4.

Study 2. Baseline covariates and number of control or treatment doses by condition.

	Control		Treatment	
	Mean	SE	Mean	SE
Baseline Covariates				
Proportion African American	0.47		0.47	
Proportion female	0.53		0.51	
Pre-intervention academic performance	0.04	0.06	-0.04	0.06
Dosage				
Doses completed in Year 1 and Year 2	6.98	0.08	7.02	0.06
Proportion completed of all doses offered	0.95	0.01	0.96	0.01
Completed all doses offered	0.80		0.75	
Completed a tailored dose	0.62		0.64	
N	172		167	

Study 2. The table reports the mean and standard errors of the baseline covariates, as well as the number of control or treatment doses completed, separately for the control (*N*=172) and treatment (*N*=167) groups. Pre-intervention academic performance reflects a composite of three indicators and is mean-centered within ethnicity. For ethnicity and gender, the mean represents the percentage of African American and female students in each condition, respectively. The last two rows in the Dosage panel indicate the percentage of students in each condition who completed the specified doses.

Table S5.

Study 1. Comparison of regression models with and without controls for student grade level, i.e., cohort.

	Outcomes							
	Course Levels		AVID		Remedial Clinics		Mainstream High School	
Variable	Basic Model	Controlling for Cohort	Basic Model	Controlling for Cohort	Basic Model	Controlling for Cohort	Basic Model	Controlling for Cohort
Baseline Performance	0.409***	0.468***	-0.969	-0.941	-0.924**	-0.952**	0.464	0.465
	(0.067)	(0.067)	(0.671)	(0.678)	(0.368)	(0.369)	(0.374)	(0.394)
Ethnicity	-0.574***	-0.608***			1.647***	1.670***	-0.927**	-0.983**
	(0.062)	(0.060)			(0.376)	(0.375)	(0.314)	(0.326)
Gender	0.023	0.013	-0.056	-0.059	0.096	0.113	0.103	0.065
	(0.044)	(0.043)	(0.386)	(0.388)	(0.245)	(0.247)	(0.243)	(0.253)
Condition	0.099	0.121	2.206**	2.160**	-0.433	-0.482	0.780	0.855
	(0.086)	(0.083)	(0.85)	(0.855)	(0.471)	(0.478)	(0.485)	(0.496)
EthnicityxCondition	0.264**	0.291***			-1.061*	-1.085*	1.305**	1.384**
	(0.087)	(0.084)			(0.475)	(0.478)	(0.484)	(0.495)
Grade dummy 1		0.071*		0.138		0.170		0.527*
		(0.031)		(0.36)		(0.24)		(0.262)
Grade dummy 2		-0.160**						
		(0.051)						
Constant (Control Mean)	-0.160**	-0.166**	-2.467***	-2.450***	-0.641	-0.641	0.929**	1.078***
	(0.062)	(0.060)	(0.755)	(0.754)	(0.362)	(0.361)	(0.312)	(0.332)
Sample Size	160	160	53	53	120	120	130	130
Adjusted R^2	0.488	0.524						
Deviance			49.910	49.762	116.402	115.889	121.771	117.301

In Study 1, simple linear regression was used for the continuous course difficulty outcome, logistic regression for the three dichotomous outcomes. Baseline performance is mean-centered within ethnicity. Condition is dummy coded (0=control, 1=affirmation). All other predictors, including those corresponding to ethnicity, gender, and cohort, are mean-centered on 0. The three levels of cohort are represented by two orthogonal contrast codes (+1,+1,-2 and -1,+1,0). Where two levels of cohort were included in analyses, only a single contrast code was used (+1,-1). Unstandardized regression coefficients are reported. Values in parentheses correspond to the standard error for the coefficient directly above. *$P \leq .05$; ** $P \leq .01$; *** $P \leq .001$.

Table S6.

Study 2. Comparison of regression models with and without controls for teacher team.

	Outcomes							
	Ordinal College Enrollment		Enrolled in 2- or 4-Year College		Enrolled in 4-Year College		College Selectivity	
Variable	Basic Model	Controlling for Teacher Team	Basic Model	Controlling for Teacher Team	Basic Model	Controlling for Teacher Team	Basic Model	Controlling for Teacher Team
Baseline Performance	1.530***	1.596***	1.196***	1.232***	1.882***	1.937***	0.837***	0.826***
	(0.168)	(0.173)	(0.195)	(0.200)	(0.221)	(0.226)	(0.099)	(0.098)
Ethnicity	-0.368*	-0.371*	-0.28	-0.289	-0.39*	-0.389*	-0.342***	-0.342***
	(0.166)	(0.167)	(0.208)	(0.210)	(0.189)	(0.191)	(0.084)	(0.083)
Gender	0.006	-0.005	0.076	0.059	-0.025	-0.048	-0.132*	-0.129*
	(0.121)	(0.123)	(0.155)	(0.156)	(0.138)	(0.140)	(0.061)	(0.060)
Condition	0.329	0.316	0.375	0.348	0.307	0.294	0.157	0.151
	(0.238)	(0.240)	(0.313)	(0.314)	(0.27)	(0.272)	(0.117)	(0.115)
EthnicityxCondition	0.604**	0.591**	0.816**	0.821**	0.545*	0.565*	0.320**	0.334**
	(0.239)	(0.241)	(0.315)	(0.317)	(0.27)	(0.273)	(0.117)	(0.116)
Teacher team dummy 1		0.179*		0.096		0.213*		0.057
		(0.085)		(0.109)		(0.098)		(0.043)
Teacher team dummy 2		-0.311*		-0.352		-0.241		-0.167**
		(0.148)		(0.190)		(0.167)		(0.067)
Constant (Control Mean)			1.542***	1.592***	0.256	0.275	-0.207**	-0.217**
			(0.214)	(0.220)	(0.186)	(0.189)	(0.084)	(0.082)
Sample Size	339	339	339	339	339	339	199	199
Adjusted R^2							0.291	0.313
Deviance	540.947	532.525	282.541	278.445	340.346	333.840		

In Study 2, students had one of three teacher teams. Ordinal regression was used for the college enrollment index, linear regression for the continuous college selectivity variable, logistic regression for the two dichotomous outcomes. Baseline performance is mean-centered on 0 within ethnicity. Condition is dummy coded (0=control, 1=affirmation). All other predictors, including those corresponding to ethnicity, gender, and teacher team, are mean-centered on 0. The three levels of teacher team are represented by two orthogonal contrast codes (+1,+1,-2 and -1,+1,0). Unstandardized regression coefficients are reported. Values in parentheses correspond to the standard error for the coefficient directly above. *$P \leq .05$; ** $P \leq .01$; *** $P \leq .001$.

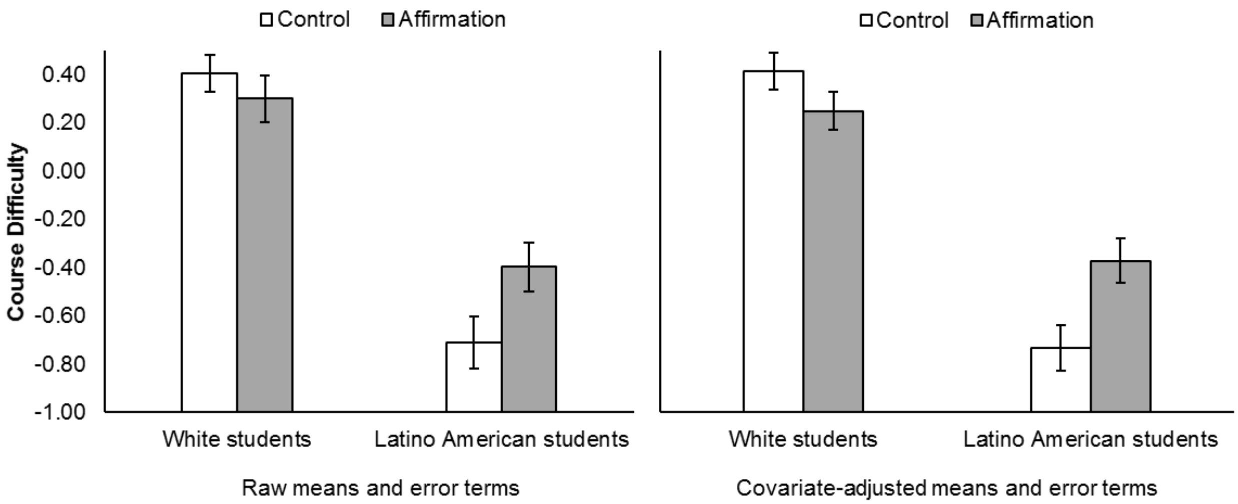
Table S7.

Study 2. Comparison of regression models with and without controls for teacher team (TT) and cohort.

	Outcomes							
	Ordinal College Enrollment		Enrolled in 2- or 4-Year College		Enrolled in 4-Year College		College Selectivity	
Variable	Basic Model	Controlling for TT and Cohort	Basic Model	Controlling for TT and Cohort	Basic Model	Controlling for TT and Cohort	Basic Model	Controlling for TT and Cohort
Baseline Performance	1.530***	1.596***	1.196***	1.228***	1.882***	1.963***	0.837***	0.830***
	(0.168)	(0.174)	(0.195)	(0.200)	(0.221)	(0.230)	(0.099)	(0.098)
Ethnicity	-0.368*	-0.373*	-0.280	-0.279	-0.390*	-0.401*	-0.342***	-0.339***
	(0.166)	(0.168)	(0.208)	(0.211)	(0.189)	(0.193)	(0.084)	(0.083)
Gender	0.006	-0.009	0.076	0.048	-0.025	-0.051	-0.132*	-0.122*
	(0.121)	(0.123)	(0.155)	(0.157)	(0.138)	(0.140)	(0.061)	(0.060)
Condition	0.329	0.301	0.375	0.337	0.307	0.271	0.157	0.158
	(0.238)	(0.240)	(0.313)	(0.315)	(0.270)	(0.274)	(0.117)	(0.115)
EthnicityxCondition	0.604**	0.583*	0.816**	0.813**	0.545*	0.563*	0.320**	0.325**
	(0.239)	(0.241)	(0.315)	(0.317)	(0.270)	(0.274)	(0.117)	(0.115)
Teacher team dummy 1		0.181*		0.103		0.211*		0.047
		(0.086)		(0.109)		(0.098)		(0.043)
Teacher team dummy 2		-0.327*		-0.347		-0.283		-0.157*
		(0.150)		(0.193)		(0.171)		(0.067)
Cohort dummy 1		0.081		0.011		0.145		-0.034
		(0.085)		(0.110)		(0.098)		(0.041)
Cohort dummy 2		-0.140		-0.170		-0.131		0.154*
		(0.150)		(0.190)		(0.168)		(0.069)
Constant (Control Mean)			1.542***	1.611***	0.256	0.297	-0.207**	-0.220**
			(0.214)	(0.222)	(0.186)	(0.191)	(0.084)	(0.082)
Sample Size	339	339	339	339	339	339	199	199
Adjusted R^2							0.291	0.326
Deviance	540.947	530.807	282.541	277.639	340.346	331.135		

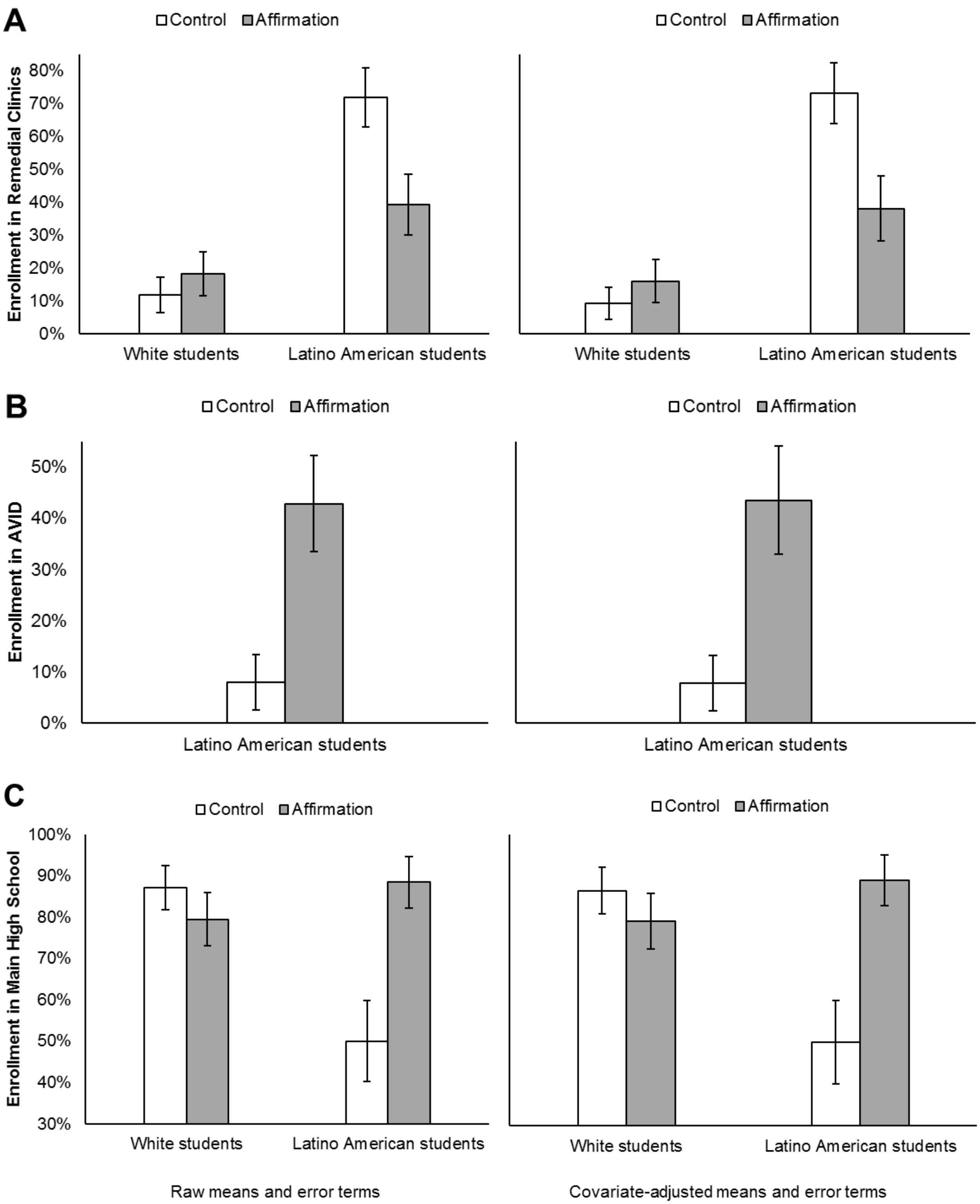
In Study 2, students participated in the intervention during the same grade (7th) but enrolled in different chronological years. Ordinal regression was used for the college enrollment index, linear regression was used for the continuous college selectivity variable, logistic regression for the two dichotomous outcomes. Baseline performance is mean-centered on 0 within ethnicity. Condition is dummy coded (0=control, 1=affirmation). All other predictors, including those corresponding to ethnicity, gender, teacher team, and cohort, are mean-centered on 0. The three levels of teacher team and cohort are each represented by two orthogonal contrast codes (+1,+1,-2 and -1,+1,0). Unstandardized regression coefficients are reported. Values in parentheses correspond to the standard error for the coefficient directly above. *$P \leq .05$; ** $P \leq .01$; *** $P \leq .001$.

Fig. S1



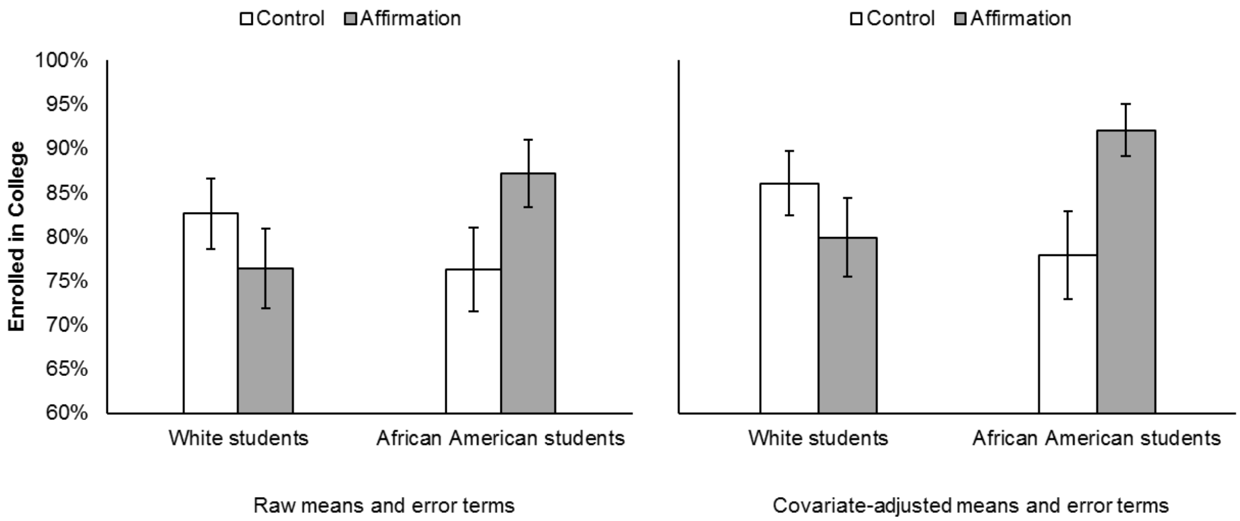
Course difficulty for all three cohorts in Study 1, two years after the intervention, as a function of ethnicity and condition, controlling for pre-intervention academic performance and gender. Error bars represent +/- 1 SE. The y-axis represents raw (**left**) or covariate-adjusted (**right**) mean levels of course difficulty, standardized within grade level. Adjusted means were obtained from linear regression. The right panel is the same as Fig. 2 in the manuscript. *N*=160.

Fig. S2


A
Control
Affirmation
Enrollment in Remedial Clinics
80%
70%
60%
50%
40%
30%
20%
10%
0%
White students
Latino American students
B
Enrollment in AVID
50%
40%
30%
20%
10%
0%
Latino American students
C
Enrollment in Main High School
100%
90%
80%
70%
60%
50%
40%
30%
White students
Latino American students
Raw means and error terms
Covariate-adjusted means and error terms

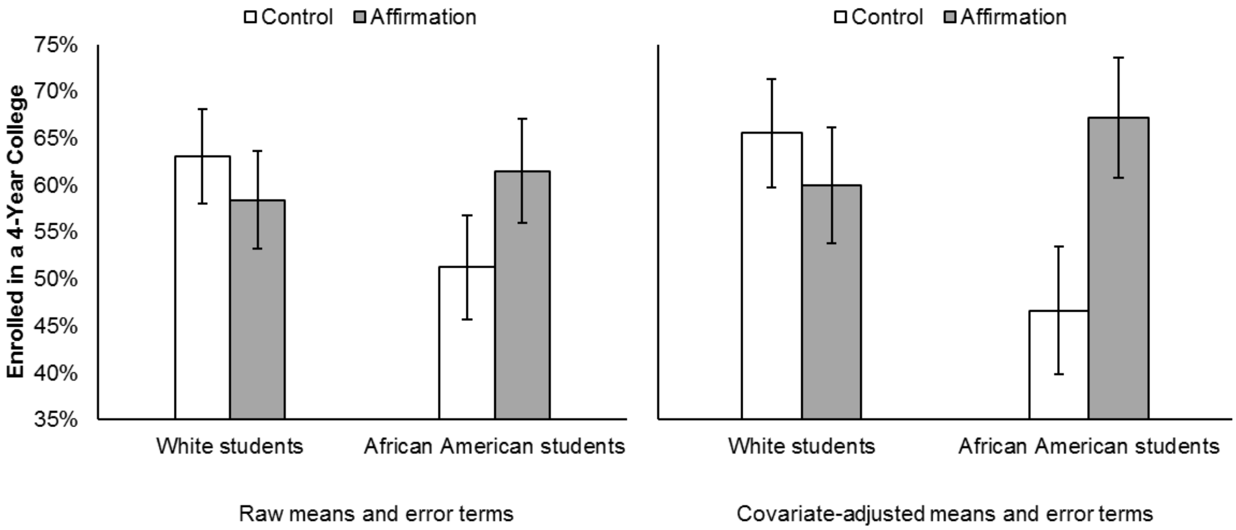
Outcomes corresponding to the remedial and college readiness tracks for students in Study 1, as a function of ethnicity and condition, controlling for pre-intervention academic performance and gender. Error bars represent +/- 1 SE. The y-axis in each panel represents the raw percentage of students (**left**) or adjusted probability (**right**) that a student obtained the given outcome: **(A)** enrolled in a remedial clinic during 8th grade, N=120; **(B)** enrolled in the college preparatory program AVID during 8th grade, N=53; and **(C)** enrolled in the mainstream high school for the district two years after the intervention, N=130. Outcomes **(A)** and **(B)** were only available for the 6th and 7th grade cohorts; outcome **(C)** was only available for the 7th and 8th grade cohorts (see SI Text and Table S1). Adjusted means were obtained from logistic regression.

Fig. S3



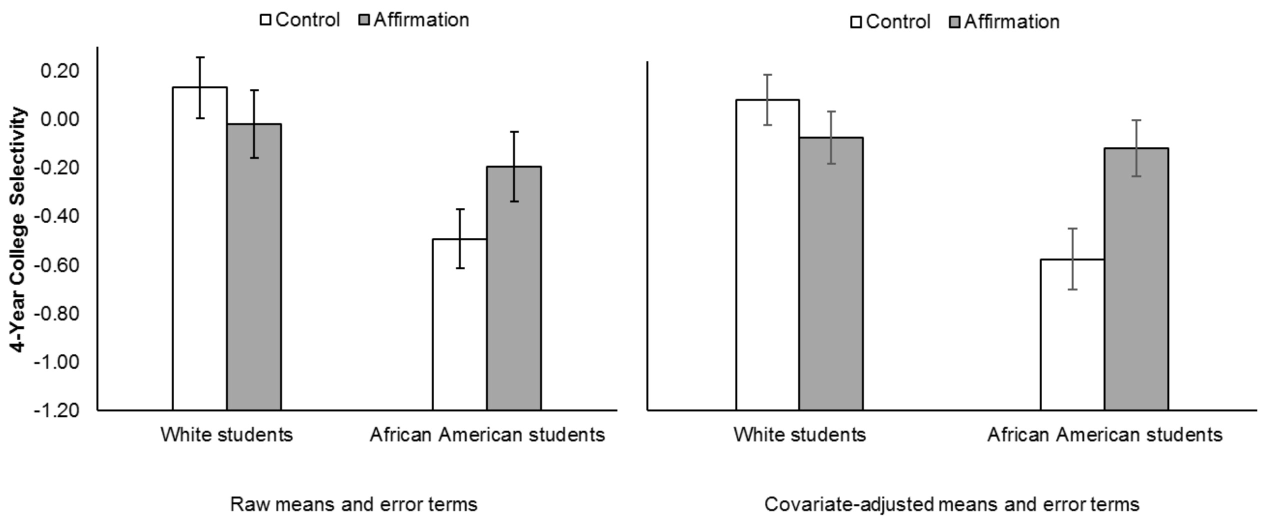
The figure depicts whether students in Study 2 ever enrolled in a 2- or 4-year college, relative to no college, as a function of ethnicity and condition, controlling for pre-intervention academic performance and gender. Error bars represent +/- 1 SE. The y-axis represents the raw percentage of students **(left)** or adjusted probability **(right)** of ever enrolling in college. Adjusted means were obtained using logistic regression. The right panel is the same as Fig. 3A in the manuscript. *N*=339.

Fig. S4



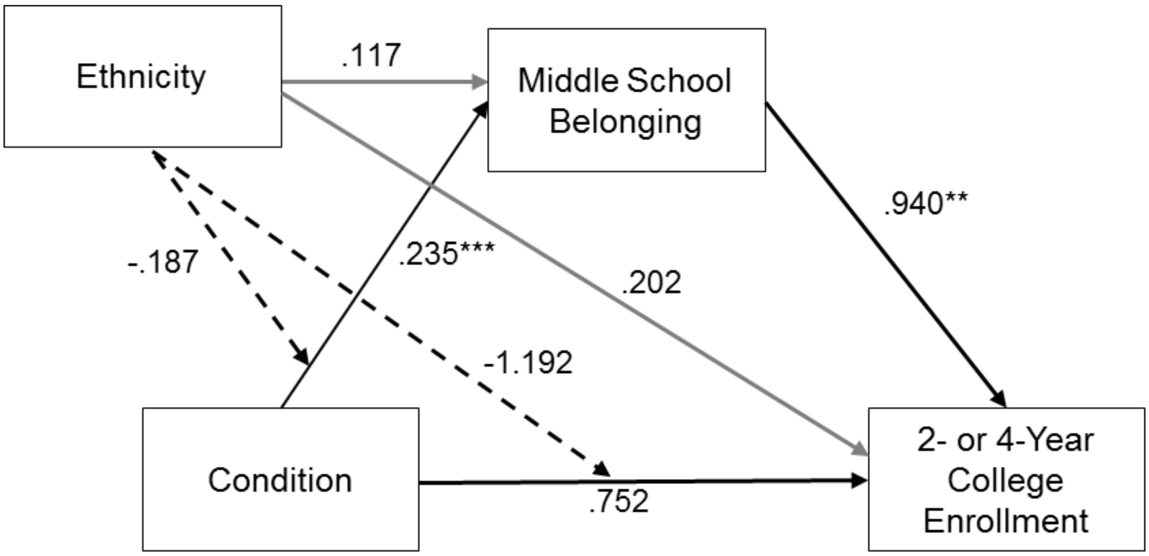
The figure depicts whether students in Study 2 ever enrolled in a 4-year college, relative to a 2-year college or not enrolling in college at all, as a function of ethnicity and condition, controlling for pre-intervention academic performance and gender. Error bars represent +/- 1 SE. The y-axis represents the raw percentage of students **(left)** or adjusted probability **(right)** of ever enrolling in a 4-year college. Adjusted means were obtained using logistic regression. *N*=339.

Fig. S5



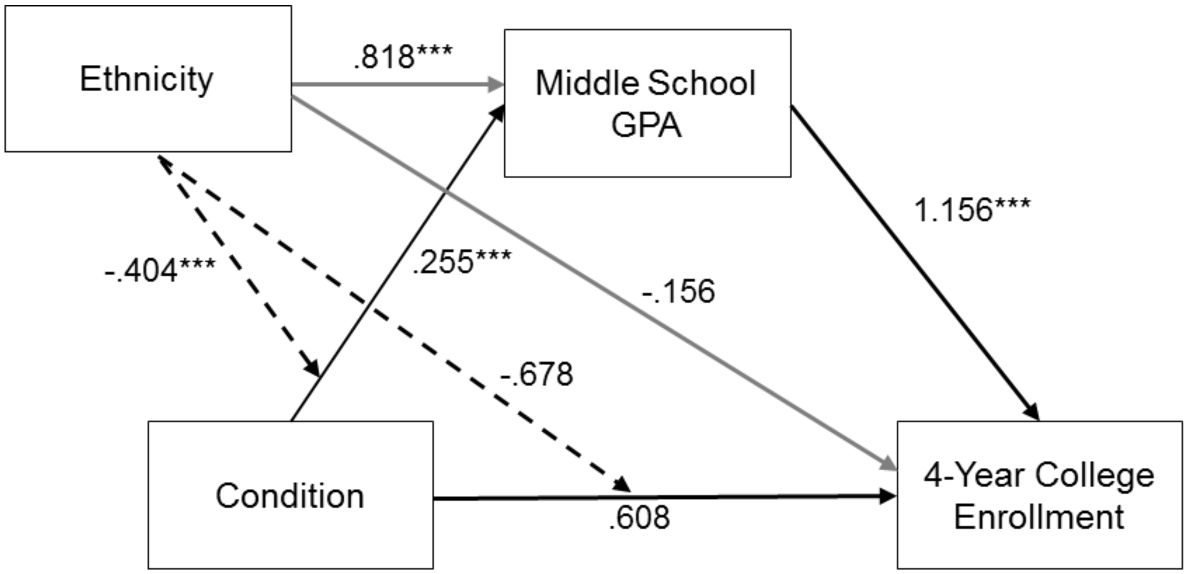
Selectivity of college choices for students in Study 2 who ever enrolled in a 4-year college, as a function of ethnicity and condition, controlling for pre-intervention academic performance and gender. Error bars represent +/- 1 SE. The y-axis represents raw (**left**) or covariate-adjusted (**right**) mean scores on a composite selectivity index, derived from publicly-available indicators. Adjusted means were obtained from linear regression. The right panel is the same as Fig. 3B of the manuscript. *N*=199.

Fig. S6



Study 2. Moderated (ethnicity) mediation (middle school belonging) of condition on 2- or 4-year college enrollment (versus no enrollment). For ethnicity, 0=African American, +1=White. Middle school belonging represents self-reported sense of belonging after the commencement of the intervention, the average belonging score in each semester from the spring of year 1 to the spring of year 2. For condition, 0=control, +1= affirmation. Dashed lines indicate moderator relations. Model controls for gender (+1=female, -1=male), baseline performance (mean-centered on zero within ethnicity), and baseline belonging (fall of year 1). Coefficients represent unstandardized estimates derived from maximum likelihood estimation. The direct effect of condition controls for the mediator and all other variables. Only study participants with non-missing values for baseline and post-intervention middle school belonging were included (N = 328). *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$.

Fig. S7



Study 2. Moderated (ethnicity) mediation (middle school GPA) of condition on 4-year college enrollment (versus enrollment in 2-year college or no college enrollment). For ethnicity, 0=African American, +1=White. Middle school GPA is the average GPA students earned over eight quarters after the commencement of the intervention, through years 1 and 2 of the study (7th and 8th grade for most students). For condition, 0=control, +1=affirmation. Dashed lines represent moderator relations. Model controls for gender (+1=female, -1=male) and baseline performance (mean-centered on zero within ethnicity). Coefficients represent unstandardized estimates derived from maximum likelihood estimation. The direct effect of condition controls for the mediator and all other variables (N = 339). *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$.