Changing Social Norms to Foster the Benefits of Collaboration in Diverse Workgroups

Kevin R. Binning, a,b,1 Nancy Kaufmann, c Erica McGreevy, c Omid Fotuhi, b Susie Chen, a Emily Marshman, d Z. Yasemin Kalender, d Lisa Limeri, e Laura Betancur, a Chandralekha Singh d

a University of Pittsburgh, Department of Psychology, Pittsburgh, PA 15260
b University of Pittsburgh, Learning Research and Development Center, Pittsburgh, PA 15260
c University of Pittsburgh, Department of Biological Sciences, Pittsburgh, PA 15260
d University of Pittsburgh, Department of Physics and Astronomy, Pittsburgh, PA 15260
e University of Georgia, Department of Biochemistry and Molecular Biology, Athens, GA 30602

1 Corresponding author: Kevin R. Binning, kbinning@pitt.edu, Phone: 412-624-7473
Changing Social Norms to Foster the Benefits of Collaboration in Diverse Workgroups

Although collaboration is essential for advancing science and maximizing human performance, collaboration in demographically diverse groups has often proven ineffective and sometimes harmful for negatively stereotyped group members. Here we report the results of a two randomized field experiments that sought to change social norms in an effort to realize the benefits of demographic diversity in collaborative workgroups. Separate experiments were conducted in undergraduate Biology (N=1215) and Physics (N=607) courses that were already employing small-group collaboration (3-4 students) during weekly study sections. At the beginning of the semester, study sections were assigned to receive either the intervention or business-as-usual control activities. The 30-minute intervention used narrative writing, peer testimonials, and small group discussion to establish a local norm that social and academic struggles are normal, anxiety about belonging is common, and most students eventually overcome these challenges. At the end of the semester, students who worked in diverse groups reported more positive social experiences in the intervention compared to control condition. Behaviorally, average attendance was higher in study sections that received the intervention, as was persistence in college after one year. Finally, students in each context theorized to be high in belonging uncertainty showed performance benefits, as the intervention closed the ethnic group performance gap in Biology classrooms and the gender performance gap in Physics classrooms. The results illustrate how social experiences in collaborative groups can be engineered to help realize the benefits of diversity.

Collaboration in Demographically Diverse Workgroups

Humans have always walked a fine line between reaping the benefits of collaboration and suffering their consequences. ‘Two heads are better than one,’ but, ‘Too many cooks spoil the
broth’ (see Heywood, 1562/1867; Koriat, 2012). One way to realize benefits of collaboration is
to have non-redundancy among collaborators, such as non-redundancy of perspectives,
knowledge, or skills. If the members of a group can complement each together, like pieces of a
jigsaw puzzle, then each contributes to a larger whole and the benefits of collaboration can be
realized (Aronson, 2002). Following this reasoning, demographically diverse workgroups, by
bringing together diverse perspectives, have the potential to outperform more homogenous
workgroups. This possibility is often cited as one of the virtues of diversity (e.g., Page, 2008).
However, several large meta-analyses and reviews of the literature suggest demographic
diversity does not generally improve performance (e.g., Eagly, 2016; Williams & O’Reilly,
1998; Van Dijk, Van Engen, & Van Knippenberg, 2012). In fact, in some circumstances
collaboration can even be harmful, particularly when members of the group are subject to
negative stereotypes in that context (e.g., women in a majority male engineering group; Dasgupta
Scircle, & Hunsinger, 2015).

The present research proposes that social psychological processes are partly responsible
for the failure of demographic diversity to fulfill its potential. We argue this is because such
diversity introduces a layer of ambiguity into the social interactions that define the collaboration,
particularly when diversity occurs within categories, such as ethnicity and gender, where
stereotypes about competence vary across subgroups. For people who are subject to negative
competence stereotypes (e.g., ethnic minorities, women in engineering), any instance of
adversity or difficulty with performance-relevant behavior can be seen as a confirmation of the
stereotype. That is, in addition to all the typical reasons people struggle completing tasks (e.g.,
because the material or procedures are highly complex), the existence of competence stereotypes
provides a heuristic shortcut, an easy explanation for why stereotyped individuals struggle (Walton & Cohen, 2007).

Attributing adversity to a fixed, essential feature of a person, such as their ethnicity or gender, can have pernicious consequences. Such attributions imply that no matter what the person does or how hard they try, their essential character makes them unlikely to succeed in the long run (Yeager & Dweck, 2012). The fear of being targeted by negative stereotypes is known as stereotype threat (Steele, 1997). Importantly, people experience stereotype threat in performance settings, such as when taking academic tests (Steele & Aronson, 1995). But people also experience stereotype threat during collaborative interactions with outgroup members (Dasgupta et al., 2015). In fact, both minority and majority group members in intergroup interactions can experience stereotype threat during social interaction. Whites may experience the threat of appearing prejudiced (Goff, Steele, & Davies, 2009), while minorities experience the threat of appearing incompetent. This matters because regardless of its source, anxiety can be contagious (Hatfield, Cacioppo, Rapson, 1994). Anxiety during social interactions causes interaction partners to be more anxious (West, Koslov, Page-Gould, Major, & Mendes, 2017), which can increase avoidance and reduce the desire for future contact (West, Shelton, & Trail, 2009).

By understanding and managing the ambiguity created by demographic diversity it may be possible to improve the outcomes of collaboration in diverse groups. Our intervention approach borrows from research on the power of personal and social narratives to shape social norms and behavior (Walton & Cohen, 2007; Walton & Wilson, 2018; Wilson & Linville, 1985). Using a field experiment with random assignment, we sought to teach students in the intervention condition how to perceive adversity in a way that closes off the use of stereotypes to
explain it. That is, rather than attributing adversity to internal and fixed causes (e.g., ethnicity, gender), we sought to teach students to view adversity – both their own adversity and adversity among their peers – as universal and temporary (“Everyone struggles sometimes, things will improve if you keep at it”). We sought to instill these narratives not just at the individual level, as in previous research (e.g., Walton & Cohen, 2007), but at an ecological (small group and classroom) level as well. We did so by delivering the intervention in a collaborative group setting, in the early stages of the group’s formation. In this way, we sought to change the local norms for collaboration and thereby improve a variety of personal, group, and classroom outcomes over time.

Across two studies we examined the effect of the intervention on three main types of outcomes among students enrolled in Introductory Biology (Experiment 1) and Basic Physics for Engineering (Experiment 2). That is, we examine their self-reports of their social experiences (e.g., the efficacy of their group and the quality of their social interactions on campus), behavioral indicators of engagement (i.e., study section attendance and one-year college persistence), and their performance (i.e., course grades and one-year college GPA).

**Control Condition**

Both of the courses under study were employing collaborative learning activities regularly during required weekly study sections (i.e., classroom settings associated with a larger lecture). The use of collaborative and cooperative learning in the classroom is a well-established pedagogical tool, and it is widely used in classrooms around the world (Johnson, Johnson, Roseth, & Shin, 2014; Kyndt et al., 2013). We hypothesize that the intervention operates by changing collective norms and narratives about the nature of adversity, which in turn changes the nature of social interactions over the course of the semester. To make this inference, we wanted
to be certain that it was not simply a sense of group affiliation or students’ sense of “mere belonging” that drove any effects. Notably, evidence suggests that incidental features of a context that enhance belonging (e.g., answering the same way on a survey) exert reliable effects on people’s emotional and cognitive connection with others (Brannon & Walton, 2013; Cwir, Carr, Walton, & Spencer, 2011). Conveniently, the activities students were already doing at the time of the intervention were structured to allow this type of inference. That is, during the week of the intervention the business-as-usual condition encouraged students to form social bonds with their peers. Critically, however, the activities lacked the intervention content theorized to change the way students think about belonging and adversity.

The intervention was randomly assigned to eight Biology classrooms and six Physics classrooms (Ns=608 and 169 for Biology and Physics, respectively). Meanwhile, there were eight contemporaneous business-as-usual control classrooms in Biology and 14 in Physics (Ns=607 and 438 for Biology and Physics, respectively). It was delivered on either the first (in Biology) or second (in Physics) week of classes for the semester (see SOM for sample intervention script and materials). The control activities during the week of the intervention included “ice-breaker” activities designed to help students get to know one another and feel comfortable in their group. For example, in Biology classrooms, the business-as-usual activities involved forming groups of four based on a superficial similarity (e.g., liking the same quote on the wall or living in the same dorm), creating a Biology-related team name together (e.g., “The Heterozy-goats”), drawing a picture of their team’s mascot, and presenting their mascot to the class. Physics classrooms also featured classroom introductions and an initial small group problem solving set.

**Intervention Condition**
The facilitator introduced the intervention as a break from regular classroom activities and asked students to engage in a brief reflective writing task in which they considered their own transition to college and the stressors and challenges they had overcome. This was meant to help students organize their own narratives about their transition in preparation for the upcoming discussion. The facilitator then presented testimonials of upperclassmen that conveyed the intervention message, namely, that adversity during the transition to college is both normal and temporary (Walton & Cohen, 2007). We tailored the quotes to the respective social contexts under study and designed them to provide students a model of the types of narratives we wanted them to adopt. In Biology, the quotes did not draw explicit attention to ethnicity, but each quote was attributed to members of different ethnic groups (one African American, one Asian American, and one White American) to help illustrate the universality of struggling during the transition to college. The quotes used in Biology were largely adapted from prior research (Yeager et al., 2016; Walton & Cohen, 2007). Physics testimonials were created for that context following a focus group with two female Physics graduates. Physics quotes were attributed to two men and two women, and one woman mentioned her initial discomfort with being a woman in the majority-male classroom (see supplementary materials for the actual quotes used).

The next phase involved group discussion. Our goal with the group discussion to provide social proof of the intervention message from students’ peers (Cialdini, 2001). Social psychological interventions in education have been theorized to be especially powerful and enduring when they influence students’ interaction with others (Goyer et al., 2017). Students were encouraged to both say and hear that struggling in college is common but surmountable, not from an instructor or experimenter but from actual peers who were in the same situation. To accomplish this, we designed the discussion questions to have intervention messages pre-
supposed, so that students would be encouraged to adopt the intervention messages as “given” in their discussions (Schwarz, 2014). That is, groups were asked to discuss a) why people often do not realize that so many students struggle with the transition college and b) how their lives may be different when they are juniors and seniors. Whereas the first topic assumes that adversity is both common and often hidden beneath the surface, the second topic assumes that students will progress through college and change along the way. After several minutes of discussion, the facilitator reinforced the activity by asking volunteers to share with the full class what their group had discussed. Just as in the control condition, after the intervention, students continued to see and interact with their group members throughout the remainder of the semester—a fact that we expected would help reactivate over time any norms established by the intervention (Aarts & Dijksterhuis, 2003).

**Self-Reported Social Experiences**

We hypothesized that demographically diverse groups should be more strongly affected by the intervention than homogenously White groups. If there is no diversity in a group because all its members are White, for example, then ethnicity is a constant and loses much of its social explanatory power in the group setting (Nisbett, 2009). It is when diversity is present that the ambiguity targeted by the intervention are most likely to be in play. In Biology, students were arranged into 302 (149 treatment; 153 control) four-person groups and interacted with their groups over the term. Biology students’ small-group memberships were recorded by the course instructor, which allowed for analyses of group composition effects. Unfortunately, records of small groups were not obtained from Physics classrooms, and thus this one set of analyses was limited to the Biology study. Although the vast majority (218 groups; 72%) of the small groups had at least one ethnic minority, a number of groups were homogenously White (84 groups;
To measure group diversity, students were assigned a score reflecting the proportion of their workgroup who were non-White (M=.28, SD=.23). Although this approach to defining diversity does not differentiate among ethnic minority groups, it follows from historical data and the present results showing underperformance among all ethnic minority groups, including students of Asian descent, in the targeted Biology course (see supplementary materials for additional details).

Using multi-level modeling to account for the nested nature of the data, we then tested whether workgroup diversity moderated the effect of the intervention on students’ social experiences. Using registrar data, this and all below analyses controlled for individual participants’ SAT/ACT math, SAT/ACT verbal, and their high school GPA. Although all inferential tests were based on the full sample, the simple effects of diversity were also examined separately for White and non-White students. While the effects were generally stronger for minority students, White students trended in the same direction and there were no ethnicity differences in the strength of Condition X Composition interactions (i.e., there were no three-way interactions). Thus we simply controlled for participants’ ethnicity using dummy codes. As plots of the results in Figure 1 indicate, analyses revealed interactive patterns on students’ quality of campus interactions (B=.64, two-tailed p=.030), perceived efficacy of their small workgroup (B=.42, p=.025), and liking of their course instructor who had lead the intervention (B=.85, p=.005).

Simple slopes analyses revealed that in the control condition, increasing diversity was associated with lower quality of interactions, (B = -.44, p=.050), lower perceived group efficacy (B= -.26, p=.070), and lower instructor liking (B= -.52, p=.019). By contrast, among students who received the social belonging intervention, these relationships were opposite in sign and
were not significant ($B$s=.20, .16, .32, $ps>.170$, for quality, efficacy, and liking, respectively).

These results suggest the intervention fostered more positive social experiences for students who worked together in diverse workgroups but not for students who worked in homogenously White groups.

*Figure 1. Aggregated survey responses of students’ perceptions and evaluations of the quality of interactions with others students and faculty on campus (I), the perceived effectiveness of their small work group (II), and their level of liking for the course instructor (III), scaled from 0 (low) to 1 (high). Error bars reflect +/-1 standard errors.*

**Attendance**

In both courses we found evidence of a main effect of the intervention on discussion attendance, although the effect was stronger in Physics than in Biology (Figure 2). In Biology, students who received the social belonging intervention attended discussion ($M=.94$) marginally more than students who received business-as-usual control activities ($M=.92$; $B=.02, p=.077$). In Physics, students who received the social belonging intervention attended discussion ($M=.96$) significantly more than students who received business-as-usual control activities ($M=.89$; $B=.07, p<.001$). Although the attendance effects appeared to be larger for the stereotyped groups in each context (ethnic minorities and women for Biology and Physics, respectively), effects on attendance were general main effects that did not differ significantly between stereotyped and
non-stereotyped groups. In Biology, the intervention effect on attendance was significant among minority students ($B=.04$, $p=.043$) but not significant among White students ($B=.01$, $p=.354$). In Physics, the effect on attendance was significant among both women ($B=.10$, $p=.009$) and men ($B=.06$, $p=.020$). A breakdown of attendance scores for each course is displayed in Figure 2, revealing a pattern consistent with the notion that the intervention boosted behavioral engagement in the same classroom where the intervention was administered.

**Persistence**

Our analyses further revealed a main effect of the intervention on college persistence after one year. Using students’ academic transcripts, we created a dummy variable to capture persistence for the year of the intervention (the intervention semester and one subsequent semester). Students were coded as 1 (Persisting) if they were listed as having received a GPA score greater than 0.00 in each of the two terms. They received a 0 (Not-persisting) if they did not have a GPA reported in the student database after the intervention term (a sign of either dropout or transfer) or received a GPA of 0.00 in either of the two terms (indicative of either
formal or informal withdrawal from all courses). The overall rates of persisting were higher among Physics students \((M=94.5\%)\), who were primarily students in the Engineering College, than among Biology students \((M=88.6\%)\), who were primarily in the College of Arts and Sciences. However, aside from this main effect, multilevel binomial regression analyses on the combined sample revealed persistence was significantly moderated by the intervention. The effect approached significance in Physics \((B=.84, p=.074)\) and significant in Biology \((B=.70, p<.001)\), although these slopes did not differ \((B=.14, p=.770)\), and the combined slope collapsing across both courses was significant \((B=.73, p<.001)\). Students in the intervention condition had significantly higher college persistence \((M=.953)\) than students in the control condition \((M=.909)\). As depicted in Figure 2, like attendance, the pattern held regardless of whether students were members of historically underperforming groups in their respective contexts (interaction \(ps>.260)\).

**Course Performance**

Multilevel linear regression analyses also revealed a main effect of the intervention on performance, whereby students in the intervention condition had significantly higher course performance compared to students in the control condition. While controlling for participants’ ethnicity, gender, high school GPA, and SAT/ACT math and verbal scores, adding the condition variable \((0=Control, 1=Social Belonging)\) to the model predicting course grades (measured 0-4.0 GPA scale) yielded small and similarly-sized effects in Physics, \((B=.14, p=.033)\) and Biology \((B=.15, p=.021)\). That is, with course performance, being in the intervention condition was associated with .14-.15-point boost in students’ course grade.

As hypothesized, the intervention also closed achievement gaps. As depicted in Figure 1, in control Physics classrooms, there was gender gap favoring men \((B=.37, p<.001)\), which was
Figure 3. Average course grade on a 4 point scale (0.00=F; 4.00=A or A+) among students assigned to business-as-usual control conditions or intervention conditions within Basic Physics for Engineers as a function of both gender* (I) and ethnicity (II), and within and Introductory Biology as a function of both gender (III) and ethnicity* (IV). Error bars reflect +/-1 standard errors (*=interaction p<.05).

reduced to non-significance in the classrooms that received the intervention condition \((B=.06, p=.628)\). Likewise, in the control Biology classrooms, there was an achievement gap favoring White students \((B=.34, p=.001)\), which was reduced to non-significance in the intervention condition \((B=-.05, p=.623)\). Further analyses indicated the achievement gaps were closed not by lowering the grades of higher performing groups, but by improving the grades of the underperforming groups. In Physics, women experienced a particularly strong performance boost
in the intervention condition ($B=.37, p=.004$), compared to men ($B=.06, p=.484$). Likewise, in Biology, the intervention effect was especially strong among minority (non-White) students as a whole ($B=.44, p=.001$) compared to White students ($B=.08, p=.326$). The intervention effect among minority students was consistent across all minority groups, including Asian students ($B=.48, p=.001$).

Also of note, unlike in Physics, in Biology there was no gender gap in the control condition ($B=-.05, p=.601$) and the intervention effect was similar for both genders (Gender x Condition interaction, $B=-.06, p=.652$). As such, the intervention in Biology addressed underperformance where it existed, among ethnic minority students, but it had no effect where no performance gap was present. In Physics, the intervention effect was not statistically different between White ($B=.11, p=.100$), and non-White students ($B=.31, p=.173$).

**College GPA**

Next we tested whether the intervention effects in each course had a general effect on students’ overall term GPA. Results differed between Biology and Physics. Among Biology students, there was a significant ethnicity gap in the intervention-term GPA of students in the control condition ($B=.40, p<.001$), with White students having a higher GPA ($M=3.07$) than minority students ($M=2.68$). However, this achievement gap was significantly smaller among students in the intervention condition ($B=.13, p=.083$). As was true for course and exam performance, the gap was closed via improved grades among minorities in the intervention condition. That is, minority students in intervention condition had a significantly higher GPA ($M=2.99$) compared to minority students in the control condition ($M=2.68; B =.31, p=.005$), whereas White students performed similarly in the intervention condition ($M=3.12$) and control condition ($M=3.07$), $B=.05, p=.373$. 
Among Physics students, by contrast, there was no gender gap in overall GPA in either the control condition ($B = -.10, p = .126$) or the intervention condition ($B = -.03, p = .829$). This indicates that at the GPA level, women were not collectively underperforming in the control condition, meaning their underperformance in Physics appears to have not been characteristic of women’s broader academic underperformance. Perhaps, then, the intervention had no effect on term GPA for women because there was no broader underperformance to address at this level.

**Discussion**

Research has documented a number of benefits of working in collaborative learning groups, including that it increases critical thinking (Gohkale, 1995), reinforces cognitive connections with course material (Kim & Sax, 2011; Walker, 2003; Cabrera et al., 2002; Cross, 1999), increases academic achievement (Kyndt et al., 2013; Stump et al., 2011), and increases retention into the second year of college (DeAngelo, 2014; Tinto, 1999; Chickering & Gamson, 1987). However, large-scale analyses of the literature suggest that while collaboration is generally better than working alone, the benefits are not enhanced by demographic diversity (Eagly, 2016). The present research argues that demographic diversity does have the potential, not only to foster equity and fairness, but also to enhance the effectiveness of collaboration while boosting behavioral engagement and performance.

To realize the promise of diversity, however, we argue that it is necessary to understand the social psychological processes that pose problems for diverse social interactions. Here we reported the results of a brief intervention, tested and replicated in two college STEM courses, which sought to foster effective collaboration in demographically diverse workgroups. The intervention targeted the social norms and narratives that students socially construct to evaluate themselves and their collaborative partners, not just at the individual level but at group and
classroom level, too. In particular, we targeted students’ norms and narratives about adversity, stereotypes, and belonging. Using a combination of instruction, reflective writing, student testimonials, and semi-structured small-group discussion, we trained students to collectively view adversity as both normal and temporary with sustained effort.

The present research imported into the classroom several lessons from prior work that was first honed in lab settings (Walton & Cohen, 2007) and then delivered at scale over the internet (Yeager et al., 2016). It did so following the observation that social interactions in diverse classroom settings can be a source of threat for both majority and minority groups alike (Dasgupta et al., 2015). We used a semi-structured discussion with classroom peers designed to elicit and reinforce the intervention message. In this way, the present intervention sought to shape local classroom norms around the meaning of adversity, not just during the intervention but over the ensuing semester. We sought to leverage the processes inside the classroom to establish the intervention message as conversational “common ground” (Binning & Sherman, 2011; Stalnaker, 2002).

Consistent with our aims, students’ survey responses showed that the social benefits of the intervention were concentrated in more diverse work groups. Students across the intervention classrooms tended to come to their study sessions more frequently, and they were more likely to have achieved full-time enrollment after one-year. Finally, the intervention closed the ethnicity performance gap in Biology and the gender performance gap in Physics. Both historical analyses (reported in SOM) and the results from the business-as-usual control conditions indicated that women enrolled in Physics and ethnic minorities enrolled in Biology tended to underperform their peers in the class, even after controlling for proxies for each student’s high school preparation - that is, their standardized test scores and high school GPA. It was hypothesized that
the emergence of performance gaps was driven by processes related to students’ social interactions in diverse classroom contexts, and thus an intervention tailored to secure belonging in this context should shrink or close those gaps. Indeed, the brief, scalable intervention effectively eliminated the emergent group-based achievement gaps in performance, not by bringing down higher-performing groups but by lifting up the underperforming groups.

Additional research is needed to understand the mechanism by which the intervention impacted diverse collaborations. For example, one line of research has found that although not everyone may benefit from a particular social psychological intervention, those who do benefit can make others around them better off too (Powers et al., 2016). Thus, if women or ethnic minorities in the classroom benefited from the intervention by rendering stereotype threat and belonging uncertainty less impactful, other students may have benefited from their benefit, for example, by reciprocating their heightened engagement or openness caused by higher belonging security. We argue that while the intervention had a particularly favorable impact on negatively stereotyped groups, we also note the evidence of a general benefit of the intervention, as evidenced by the main effects of the intervention on attendance, performance, and one-year persistence. Additional research is need to understand how these general benefits come about.

More research is also needed to understand the impact of who delivers the classroom intervention exercises. For logistical (not theoretical) reasons, all the facilitators of the present intervention studies were early-career White females. In Biology, the facilitator was the course instructor. In Physics, by contrast, the facilitators were visiting the courses as representatives of the Physics Department for just the day of the intervention. The fact that similar results were found in both courses suggests the intervention may be robust to different facilitator characteristics, although there are reasons to expect facilitator characteristics to matter. For
example, facilitators from other genders or ethnicities may change students’ experiences with the intervention (Egalite, Kisida, & Winters, 2015). Women engineering students who worked with female peer mentors fared better than those who worked with male peer mentor (Dennehy & Dasgupta, 2017). Other research has shown that student-teacher relationships that occur across ethnic group divides can be a source of distrust and insecurity (Cohen, Steele, & Ross, 1999). The survey results above showed that students’ ratings of the instructor in diverse groups were significantly moderated by the intervention. Varying the characteristics of who delivers the intervention, and more generally seeking to understand the role of facilitators and instructors in the intervention effects, will help illuminate factors that may catalyze and inhibit the intervention effects.

With a recent review of the literature urging more research on when the benefits of diversity emerge (Eagly, 2016), the results suggest the insights from the present intervention can be leveraged to enhance the benefits of diversity. Fostering a climate that mitigates context-specific concerns about stereotypes and belonging in the classroom may yield long-term benefits for the student, university, and society.

**METHODS**

*Missing Data.* Estimates were based on maximum likelihood estimation, which is a full information approach that utilizes all available data. The analyses used an intention-to-treat approach (Hollis & Campbell, 1999). This approach is common in clinical trials to account for possible differential attrition rates between conditions, as it includes for analysis any participant who was assigned to a treatment or control classroom at the time of the intervention (‘once randomized, always analyzed’). We opted for this approach given evidence that social psychological interventions can impact dropout and retention (Goyer et al., 2017). Thus if a
student dropped the class either formally or informally at a later point, whatever data they had
provided were included in the analyses.

Timing and Delivery. Random assignment to conditions occurred not at the lecture level
but at the level of the smaller satellite discussion sections associated with each lecture course.
That is, associated with each lecture were required discussion sections, with approximately 75
students in each Biology section and 30 students in each Physics section, that met once a week
for 50 minutes. The Biology intervention was conducted by the same female instructor over four
consecutive semesters with approximately 300 students each semester. The Physics intervention
was conducted over a single semester with three instructors who each had approximately 200
students in their lecture. All three instructors were males, but the intervention was always
delivered by one of two female Physics graduates (one graduate student and one post-doc in the
department) who visited the discussion section on the day of the intervention. In both conditions
and in both courses, students engaged in cooperative learning activities each week after the
intervention (12-13 meetings total), except for holidays or instructor absences, as they worked
together in groups to solve problem-sets related to the current class material (statistical models,
while always modeling students nested within workgroups in Biology, also accounted for
semester-to-semester variation in grades with course dummy codes at Level 2.)

As a quality improvement project approved by the institutional review board, all students
who were enrolled in class during the week of the intervention participated in the study. In
Biology there was an even split, as 608 students from eight discussion sections were assigned to
the social belonging intervention (177 non-White; 431 White) and 607 students from eight
contemporaneous discussion sections received the control activities (171 non-White; 436 White).
Overall the sample was 69% White, 21% Asian/Asian American, 7% Black/African American,
and 4% Latino. For each semester in Biology, two sections were randomly selected to receive the
treatment and two received the business-as-usual activities. In Physics, the intervention was
delivered to 169 students across six discussion classrooms in Physics (47 women; 122 men),
with 438 students from 14 contemporaneous discussion sections serving as controls (149 women;
288 men; 1 unknown). The Physics sample was 82% White, 12% Asian/Asian American, 4%
Black/African American and 4% Latino/a American. Each participating Physics instructor had
two sections that received the intervention and 3-4 control sections. Please see the supplementary
materials for the intervention script and a detailed discussion about the development, design and
implementation of the intervention.
References


Supplementary Online Materials:

Changing Social Norms to Foster the Benefits of Collaboration in Diverse Workgroups

Understanding the Classroom Contexts.

The present approach assumes that the sources of students’ anxiety are context and domain specific. Although we suggest the action of stereotyping tends to occur in classroom settings, we used two methods to understand the nature of the context where these interactions took place: 1) the existence of broad cultural stereotypes and 2) the actual group performance trends in the classes under study using historical student records.

First, evidence indicates that different class settings are likely to cause different students to question their belonging. For example, one study found that the more people associated a particular discipline with an intellectual “brilliance” stereotype (i.e., that brilliance is required for success and belonging), the greater the gender imbalance against women in that academic discipline (36). Physics and engineering were at the high end of the brilliance spectrum and had a high gender-imbalance, while biological sciences were on the lower end and were more gender-balanced. These imbalances correspond to differences in numerical representations at the undergraduate level, as women tend to be underrepresented in physics and engineering majors but overrepresented in biological sciences (16, 18). Thus, women may experience more belonging uncertainty when they are a negatively stereotyped minority in engineering courses than when they are a non-stereotyped majority in biology courses. Context also matters for performance of racial minorities, although evidence suggests that difficult classroom content in general tends to place minorities under stereotype threat (37). We argue in the present research that well-timed social belonging interventions can counteract the negative outcomes in such settings and close achievement gaps (Cohen & Sherman, 2014).
Historical Student Trends. Given these differences in stereotypes and demographic representation across disciplines, our next step was to examine if they corresponded to differences in the historical performance of students who took Physics and Biology courses at the university under study—a large, public research university in the mid-Atlantic region of the United States. The historical analyses examined patterns in performance gaps in two courses, Physics 1 for Engineers and Introductory Biology—both of which were freshman-level courses (see SOM for more details). After controlling for prior SAT scores and high school GPA, analyses of the most recent three years of historical data found that the Physics classrooms were the site of emergent gender and race gaps, in which men and White students consistently outperformed women and minority students ($B_s=.12$ and $.13; \ p_s < .006$, for gender and race [White vs. non-White], respectively), controlling for prior performance. The Biology classrooms were associated with an emergent race gap ($B=.13, \ p < .001$), as White students outperformed non-White students controlling for prior performance, but there was no gender gap ($B=.01, \ p=.821$). Notably, in Biology, students of Asian descent tended to underperform relative to White students ($B=-.13, \ p=.003$). This effect replicated a similar analysis among Asian students’ psychology course grades at the same university (38, 39). However, this was not the case in Physics, as there was no difference in average Physics course grades between Asians and Whites ($B=-.02, \ p =.713$).
Table S1. Regression model output for historical analyses of three previous years of data. Panels I and II depict the estimates using race as a dichotomous variable (0=Non-White; 1=White).

Panels III and IV depict the dummy-coded estimates for each of the three main racial minority groups with Whites as the reference group. The pattern of results indicates that gender gaps emerged in Physics but not in Biology. Conversely, a race gap between Asian and White students emerged in Biology but not Physics.

<table>
<thead>
<tr>
<th>Panel</th>
<th>Course</th>
<th>B</th>
<th>SE</th>
<th>p</th>
<th>B</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Basic Physics for Eng. (N=2389)</td>
<td>-.09</td>
<td>.04</td>
<td>.025</td>
<td>-.32</td>
<td>.04</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Semester Code 1</td>
<td>-.09</td>
<td>.04</td>
<td>.022</td>
<td>-.09</td>
<td>.04</td>
<td>.026</td>
</tr>
<tr>
<td></td>
<td>SAT Verbal</td>
<td>-.04</td>
<td>.03</td>
<td>.136</td>
<td>.04</td>
<td>.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>SAT Math</td>
<td>.44</td>
<td>.03</td>
<td>&lt;.001</td>
<td>.14</td>
<td>.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>HS GPA</td>
<td>.61</td>
<td>.04</td>
<td>&lt;.001</td>
<td>.23</td>
<td>.04</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>.12</td>
<td>.04</td>
<td>&lt;.001</td>
<td>.01</td>
<td>.04</td>
<td>.821</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>.13</td>
<td>.05</td>
<td>.005</td>
<td>.13</td>
<td>.04</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>II</td>
<td>Introductory Biology (N=4393)</td>
<td>-.32</td>
<td>.04</td>
<td>.001</td>
<td>.92</td>
<td>.02</td>
<td>.924</td>
</tr>
<tr>
<td></td>
<td>Semester Code 1</td>
<td>-.09</td>
<td>.04</td>
<td>.027</td>
<td>.00</td>
<td>.04</td>
<td>.924</td>
</tr>
<tr>
<td></td>
<td>SAT Verbal</td>
<td>-.04</td>
<td>.03</td>
<td>.184</td>
<td>.08</td>
<td>.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>SAT Math</td>
<td>.42</td>
<td>.03</td>
<td>&lt;.001</td>
<td>.14</td>
<td>.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>HS GPA</td>
<td>.61</td>
<td>.04</td>
<td>&lt;.001</td>
<td>.24</td>
<td>.04</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>.13</td>
<td>.04</td>
<td>&lt;.001</td>
<td>.01</td>
<td>.04</td>
<td>.825</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>-.02</td>
<td>.05</td>
<td>.713</td>
<td>-.13</td>
<td>.04</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>-.22</td>
<td>.07</td>
<td>.002</td>
<td>-.09</td>
<td>.06</td>
<td>.131</td>
</tr>
<tr>
<td></td>
<td>Latino/a</td>
<td>-.17</td>
<td>.10</td>
<td>.098</td>
<td>-.04</td>
<td>.09</td>
<td>.684</td>
</tr>
</tbody>
</table>
Appendix S1. Additional Background on the Development of the Classroom-Based Intervention

The present research agenda arose when colleagues in the Department of Biology (Kaufmann & McGreevy) noticed that their small-group, cooperative learning activities were not garnering full participation among group members. Often times, it seemed, students from minority backgrounds were participating the least. They approached their colleagues (Binning & Chen in Psychology; Fotuhi in LRDC) to work together to devise an intervention that merged insights from prior social belonging interventions (Walton & Cohen, 2007; Yeager et al., 2016), and imported them into the existing group activity structure to help capitalize on social processes in an important academic social setting. The Physics intervention was then developed in close collaboration with colleagues in Physics. This work followed the observation that women tended to underperform in the calculus-based physics course targeted by the intervention, but not in a similar, algebra-based physics course in which women were generally a majority of students. Each of these interventions borrowed from the general social-belonging intervention approach developed by Greg Walton and colleagues (Walton & Brady, 2019; Walton, Murphy, Logel, Yeager & The College Transition Collaborative, 2017).

Additional Reading


Appendix S2. Additional Details on Logistics of Classroom-Based Intervention.

The intervention was introduced by the facilitator as a break from typical classroom activities and consisted of a series of individual and group exercises. The procedure began with the facilitator explicitly verbalizing the intervention message—that the transition to college can be stressful, that doubts about belonging are normal, and that if students are experiencing such doubts, these doubts will likely improve. After this introduction, students completed an independent reflective writing task, in which they described a recent challenge related to college and how they can (or did) overcome it. Students’ essays commonly described feelings of being homesick, balancing course work, and having trouble with roommates. Physics students also frequently referred to doubts about their math ability. To make the essays non-evaluative, students were asked to not write their names or any identifying information on them.

Tailoring of testimonials. The facilitator then presented a series of written testimonials (via either a handout in Physics or overhead projection in Biology) attributed to upperclassman from diverse backgrounds. With the historical analyses documenting the emergence of specific achievement gaps in each course, we sought to tailor the intervention materials to each context. This was done by changing the content of the testimonials students read before the group discussion. In both courses the general message of all the testimonials was the same, but in Physics, two of the four student testimonials that students received were attributed to female students (“Allison” and “Aniyah”), while two were attributed to male students (“Nathan” and “Anil”). The inclusion of two male and two female quotes, in which both genders reported on their doubts and struggles that they overcame, was meant to convey to students that adversity is universal and not unique to specific groups. Following a focus group conducted with female physics graduates to generate authentic content for the testimonials, one of the female quotes
referred to her experience specifically as a woman in the majority-male class (e.g., “I was one of
just a handful of women in the course, and sometimes I felt a little embarrassed to ask
questions”). However, the testimonial further described how the student eventually got past her
anxiety and had a successful term (“I quickly learned that other students usually had the same
question I did, and we all benefitted from working with each other and learning from each
other”). Given the prevalence of stereotypes about the need for innate brilliance to succeed and
belong in Physics (36), the Physics quotes also incorporated the importance of an intelligence
growth mindset (55). For example, one of the female quotes stated, “I remember there wasn’t
always an ‘a-ha!’ moment where everything clicked. It was usually much more gradual, with
some concepts only becoming clear after lots of practice and discussion with my study group."

In Biology, students received three testimonials adapted from prior research (35, 31),
which were originally designed to address racial differences in belonging uncertainty in college.
Rather than attributing the quotes to different genders, at the bottom of each quote the student
was identified in terms of their race (e.g., “Asian American, [University name] Senior”). Again,
to help convey the universality of struggling, one quote was arbitrarily attributed to an African
American student, one to an Asian American student, and one to a White American student (e.g.,
“African American Senior at [University]”). As in Physics, all testimonials shared the same
message subtext: the student described social anxieties and academic fears that, with time and
effort, were eventually overcome and replaced with feelings of connection, integration, or
success. One example quote explained, “I didn’t go to a very good high school and worried that
my high school courses had not prepared me well for college.” The student found professors
intimidating and “worried a lot about how they and other students would evaluate me.” But
around the sophomore year, “I started to feel more comfortable in my classes and began to ask
my friends to edit my papers for me.” Further, the quote described, “Even when professors are
critical, or their grading harsh, it didn’t mean they looked down on me or that I didn’t belong.”

As described above, students then proceeded to engage in small group discussion that was
designed to build on the common ground established by the facilitator’s opening remarks, the
reflective writing task, and the testimonials. After completing the small group discussion, the
facilitator opened up the discussion to the full class to share their experiences.
Appendix S3. Intervention script for Introductory Physics classrooms.

**I. Introduction  3 minutes**

“Hi, my name is ______. I am a graduate student (post-doc) working in the Physics Department. Today we are going to take a break from the regular classroom activities to talk about going to college at ______. We are always trying to make things better for our students, and so today the Physics Department is asking for your help to tell us a little about your experiences since you have come to ______. For many of you, this is the first time you are leaving home, you are meeting a lot of new people, taking on a lot of new courses, and trying to find your place here at ______. It can be easy to feel overwhelmed and to ask yourself, “Do I really belong here?” and “Am I smart enough to make it?”

These kinds of experiences are normal in the transition to college. Everyone goes through them, and they get better with time as you adjust to college.”

**II. Writing  7-10 minutes**

“Today, we’d like each of you to reflect on some of the experiences you have had so far in coming to college. So, please take about ten minutes to write on the sheet of paper I passed out to you about the experiences and challenges you have had since coming to ______. What has been difficult or challenging for you? You can think about your experiences meeting other students, making friends, taking classes, adjusting to dorm life: look, coming to college is a big transition. Please write about some of the difficulties you have experienced in the transition to college and how some of these difficulties and challenges have begun to improve with time as you have spent more time in college.

Please don’t include your name or other identifying information in what you write.

When the ten minutes are up, we will collect what you’ve written. Please write on your own, individually. We’ll talk about this later.”

*At the end of 7-10 minutes, students writing samples are collected.*

**III. Reading  5-7 minutes**

After skimming several student examples of writing, the facilitator stands at the podium in front of the class:

“As I’m looking few some examples of what you all have written, I see a lot of very common concerns that you have. I am also not surprised that I had some of the same concerns when I took freshmen physics. In preparation for today’s exercise, a team interviewed a number of upperclassmen who described their transition to college and some of their experiences in the Physics 174 course. I’d like to take a few minutes to provide some examples that these students have provided.”

Facilitator then hands out the quotes. Facilitator reads each quote in full (see Appendix S6).
IV. Small group discussion 7-10 Minutes

“For the next ten minutes, I ask that you please discuss with the students sitting near you (3-4 students in each group) what you wrote about and the quotes you have just read. Please answer the following questions as a group: (Facilitator writes these questions on the board, may need to help students get into small groups).

• What are some common themes across several of the quotes we read?
• Why do you think that sometimes students don’t realize that other people are also struggling with the course?
• How do you think your lives will be different when you are a Junior or Senior?

V. Closing 3-5 minutes

“All right, let’s come back together. I’ve been overhearing some great discussions and I’d love to hear your thoughts. I think your feedback is going to be very helpful for the Physics Department. Would anyone like to share what your group discussed with the class?”

Facilitator calls on groups if no one raises their hand. Include references to own experience as appropriate, emphasizing that struggles during the transition are normal and temporary.
Appendix S4. Below are the quotes that were provided to students as a handout as part of the intervention protocol.

Quotes from former Physics 174 students

“I remember taking my first Physics class as a freshman. Before coming to college, I didn’t worry much about grades, so I felt unprepared for the increased workload and differences in grading. I remember being surprised after getting burned grade-wise several times, and feeling stressed as a result. But then I got some help from the instructor and the TA, found a study group, and was able to turn things around. Looking back now, I think my struggles were pretty normal. Even though people don’t like to admit it, basically everyone has trouble with certain concepts. Although it was a somewhat rocky start, it felt good to learn from my mistakes, and I am proud of the success I have had.”

-Nathan, Pitt Bioengineering Senior

I was one of just a handful of women in one of my intro physics study groups, and sometimes I felt a little embarrassed to ask questions. However, I quickly learned that other students usually had the same question I did, and we all benefitted from working with each other and learning from each other. Sometimes I had difficulty with an idea that my classmates understood. Other times, they struggled with concepts that I understood. I remember there wasn’t always an “aha!” moment, where everything clicked. It was usually much more gradual, with some concepts only becoming clear after lots of practice and discussion with my study group. I realized that everyone struggles some times, and the important thing is to not give up and help each other out.”

-Allison, Pitt Electrical Engineering Senior

“When I first got here, I was worried because I was really struggling with some of the physics concepts. It felt like everyone else was doing just fine, but I just wasn’t sure if I was cut out for a physics course. At some point during the first semester, I came to realize that, actually, a lot of other students were struggling, too. And I started to look at struggling as a positive thing. After I struggled with a hard problem and then I talked to other classmates and my TA about the solution—I realized that all that effort was worth it because it helped me learn and remember much more.”

-Aniyah, Pitt Chemical Engineering Graduate

“I didn’t go to a very good high school, and was I worried that my high school courses had not prepared me well for college. Honestly, when I got here, I thought professors were scary. I thought they were critical and hard in their grading, and sometimes it felt like they put things on the quizzes or exams that we hadn’t discussed in class. But then I realized that the professor wanted me to be able to apply the physics concepts in many different situations. So I started to study in a way that would help me do that, and I did my best to learn from my mistakes on quizzes and exams. And I saw that even when the professors’ grading seemed tough, it didn’t mean they looked down on me or that I didn’t belong. It was just their way of motivating high achieving students.”

- Anil, Pitt Civil Engineering Senior
Appendix S5. Checks on random assignment.

Within each course there were from four to six discussion sections. Random assignment always took place at the discussion section level, and each participating instructor always had at least two discussion sections in the treatment condition and at least two sections in the control condition during the semester in which the intervention was delivered. There were seven different participating courses (4 Biology and 3 Physics). As such, to check the effectiveness of random assignment across sections, we subjected students SAT/ACT math score, SAT/ACT verbal score, and their high school GPA to a 7 (Course) x 2 (Condition) factorial MANOVA. The results revealed that while there were mean differences between courses, there was never a main effect of condition ($F$s=0.67, 0.01, and 0.34; $p$s=.414, .962, and .551, for SAT/ACT math, SAT/ACT verbal, and HS GPA, respectively). Moreover, there were no significant two-way interactions (Course*Condition), ($F$s=1.40, 1.34, and 1.11; $p$s=.212, .237, and .352, for SAT/ACT math, SAT/ACT verbal, and HS GPA, respectively). As such, random assignment appeared to be effective in equating students on their high school preparation, as within each course students in sections who received the treatment did not differ from students who received the control activities across three measures of academic performance. We nevertheless controlled for individual students’ high school preparation below.
Appendix S6. Additional details on Biology survey.

Survey. During the second-to-last week of the semester, all students in Biology were offered the opportunity to complete an online survey in exchange for course credit. Averaging across semesters, approximately 82% of students who were enrolled at the beginning of the semester completed the survey, which asked students a variety of questions about their social and academic experiences in the classroom. The measure of perceived group efficacy was asked each time the survey was given (N=1029), while the measure of quality of students’ interactions was asked just three out of four surveys (N=767), and the measure of instructor liking was only administered in the final survey (N=247).

Quality of social interactions on campus. Three items were taken from the National Survey of Student Engagement (Kuh, 2001), “Indicate the quality of your interactions with” … “Students,” “Faculty,” from 1 (Very poor) to 7 (Excellent). Alpha=.69.

Team efficacy. Composed of 3 items: My recitation team was effective at accomplishing its goals. My recitation team was effective at solving problems. [REV] My recitation team was NOT effective/useful for me. Alpha=.79.

Instructor liking. Composed of 3 items. Overall, I was satisfied with my course instructor. I generally liked the instructor for the course. I feel warmly toward the course instructor. Alpha=.92.
Appendix S7. Additional Details on Attendance Analyses.

Students’ discussion section attendance grade was typically a small portion of the course grade (5-10%, depending on the course) with the bulk of students’ grades coming from their exam scores. Nonetheless, this measure was important for assessing the effect of the intervention, as it reflected students’ behavioral engagement in the same classroom setting in which the intervention was delivered. The measure of discussion section attendance was assessed as the proportion of attended sections in which the instructor collected attendance data. Depending on the course and instructor, there were between eight and 13 sections in which attendance was collected. The attendance scores were highly skewed (with scores piled up at the high end of the distribution); however, the analytic approach for each course was robust to violations of normality because we employed non-parametric bootstrapping to estimate the effects.

We sought to compare the relative influence of the course attendance variable and the exam performance variable as predictors of persistence. When entered into a regression model simultaneously (including all previously-stated controls), the attendance variable was a powerful predictor of persistence \( (B=.49, \ p < .001) \), whereas exam performance was a marginally-significant predictor \( (B=.18, \ p=.065) \). As such, the analyses suggest that while attendance only constituted a small portion of students’ actual course grades, course attendance was a strong predictor of college persistence and a better predictor than course exam performance.
Table S2. Pearson correlations among main outcomes of interest.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Class grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Avg. exam scores</td>
<td>.881**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Attendance score</td>
<td>.594**</td>
<td>.493**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>One-Year Persistence</td>
<td>.247**</td>
<td>.171**</td>
<td>.287**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Quality of interactions</td>
<td>.196**</td>
<td>.195**</td>
<td>.129**</td>
<td>.042</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Group efficacy</td>
<td>.044</td>
<td>.014</td>
<td>.144**</td>
<td>-.036</td>
<td>.115**</td>
</tr>
<tr>
<td>7</td>
<td>Instructor liking</td>
<td>.314**</td>
<td>.306**</td>
<td>.205**</td>
<td>.085</td>
<td>.388**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).