Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
2020, Vol. 31(9) 1059–1070

DOI: 10.1177/0956797620929984
www.psychologicalscience.org/PS
SAGE

Changing Social Contexts to Foster Equity in College Science Courses: An Ecological-Belonging Intervention

Kevin R. Binning[1,2], Nancy Kaufmann[3], Erica M. McGreevy[3], Omid Fotuhi[1], Susie Chen[2], Emily Marshman[4], Z. Yasemin Kalender[5], Lisa B. Limeri[6], Laura Betancur[2], and Chandralekha Singh[7]

[1]Learning Research and Development Center, University of Pittsburgh; [2]Department of Psychology, University of Pittsburgh; [3]Department of Biological Sciences, University of Pittsburgh; [4]Department of Physics, Community College of Allegheny County; [5]Department of Physics, Cornell University; [6]Department of Biochemistry and Molecular Biology, University of Georgia; and [7]Department of Physics and Astronomy, University of Pittsburgh

Abstract

In diverse classrooms, stereotypes are often "in the air," which can interfere with learning and performance among stigmatized students. Two studies designed to foster equity in college science classrooms (Ns = 1,215 and 607) tested an intervention to establish social norms that make stereotypes irrelevant in the classroom. At the beginning of the term, classrooms assigned to an ecological-belonging intervention engaged in discussion with peers around the message that social and academic adversity is normative and that students generally overcome such adversity. Compared with business-as-usual controls, intervention students had higher attendance, course grades, and 1-year college persistence. The intervention was especially impactful among historically underperforming students, as it improved course grades for ethnic minorities in introductory biology and for women in introductory physics. Regardless of demographics, attendance in the intervention classroom predicted higher cumulative grade point averages 2 to 4 years later. The results illustrate the viability of an ecological approach to fostering equity and unlocking student potential.

Keywords

diversity, intervention, education, stereotypes, open materials

Received 4/8/19; Revision accepted 4/17/20

In collaborative settings, demographic diversity can be a means to harness multiple perspectives to maximize human potential (e.g., Page, 2008). However, in practice this is often not the case. Several large meta-analyses and reviews of the literature indicate that demographic diversity does not generally improve performance in organizational and educational collaborations (e.g., Eagly, 2016; van Dijk, van Engen, & van Knippenberg, 2012; Williams & O'Reilly, 1998). In fact, when group-based stereotypes are "in the air" (Steele, 1997), diverse contexts are sometimes harmful to performance, particularly for people who are subject to negative stereotypes in those contexts. For example, women engineering students who were assigned to a majority-male work group participated less, were more anxious, and, if they held masculine stereotypes about engineering, had lower career aspirations following the interaction (Dasgupta, Scircle, & Hunsinger, 2015).

In the present research, we argue that problems with diversity are not simply psychological but also ecological. Stereotypes gain power from their intersubjectivity: People know about stereotypes, and they know that other people know about stereotypes (Steele, 1997).

Corresponding Author:
Kevin R. Binning, University of Pittsburgh, Learning Research and Development Center, 3939 O'Hara St., Pittsburgh, PA 15260
E-mail: kbinning@pitt.edu

This has consequences for all involved. For targets of negative stereotypes, the fear of being stereotyped by others taxes working memory and interferes with the learning of new material (Rydell, Shiffrin, Boucher, Van Loo, & Rydell, 2010; Taylor & Walton, 2011). For students who are not subject to negative academic stereotypes (e.g., White students, men in engineering), knowledge of stereotypes can create anxiety and discomfort when working with stigmatized others (Richeson & Shelton, 2003), which can harm their own cognitive performance (Blascovich, Mendes, Hunter, Lickel, & Kowai-Bell, 2001), increase discrimination (Goff, Steele, & Davies, 2008), and reduce their desire for future interactions (West, Koslov, Page-Gould, Major, & Mendes, 2017).

The present research tested an intervention designed to help realize the promise of diversity. Our goal was to change the social norms of contexts in which competence stereotypes are salient. When it is widely known or suspected that certain groups underperform, stereotypes help construct the meaning that students make when poor performance does occur ("maybe this unexpectedly poor grade is evidence that people like me/them do not belong here"; Walton & Wilson, 2018). By providing students an alternative narrative—one that holds adversity as both normative and surmountable—we aimed to render stereotypes unnecessary for understanding why underperformance occurs. We did so by taking methods from previous social-belonging interventions—interventions that have successfully reduced college attainment gaps (Walton & Cohen, 2007, 2011; Yeager et al., 2016)—and applying them at the ecological (classroom) level.

Whereas previous work has focused on changing the experience of individual students with individualized intervention activities (e.g., by delivering the intervention in lab settings or over the Internet; e.g., Dasgupta, 2011; Walton & Cohen, 2007, 2011; Yeager et al., 2019; Yeager et al., 2016), we targeted the intersubjective space that students and teachers share. We sought to collectively teach students a norm that adversity—both their own adversity and adversity among their peers—is not due to internal and fixed causes (e.g., "Some people just aren't smart enough") but rather to universal and temporary causes ("Everyone struggles sometimes, but you can improve by persisting"; Walton & Cohen, 2007). By establishing these beliefs as shared social norms, we sought to create scaffolding to support and reinforce the intervention message over the semester.

Two studies employed a brief, evidence-based intervention strategy. While incorporating elements of previous social-belonging interventions, such as reflective writing and exposure to testimonials from older students (e.g., Yeager et al., 2016), the intervention also engaged students in a classroom discussion with the peers they would work with over the ensuing semester. The discussion was designed to leverage maxims of communication (Grice, 1975) in an effort to establish the intervention message as common ground (Clark & Brennan, 1991). It was intended to use communications with peers as a form of social proof (Cialdini, 2001) of the intervention message that would help normalize adversity and pop a bubble of pluralistic ignorance (Miller & McFarland, 1987), wherein students see their own challenges and adversity as unique when in fact adversity itself is quite common (Walton & Cohen, 2007). As students then collaborated to solve difficult problems over the semester, we wanted them to adopt the intervention message and, to address stereotypes in the air, to know that their peers had done so, too.

Below, we report the results of interventions conducted in two college science gateway courses. We proceed in three steps. First, looking at studies that drew on 3 years of historical data for two science courses, we examine which groups tended to underperform in each course. Second, we examine whether an ecological-belonging intervention tailored for each course context attenuated underperformance and thereby fostered greater equity in academic outcomes. Finally, we report a series of exploratory analyses to understand how the intervention may have affected student outcomes. These analyses revolved around the possibility that establishing a classroom climate in which adversity is normal and likely to be overcome may (a) increase student engagement (e.g., higher attendance) and (b) allow students to take greater advantage of learning opportunities that are presented to them (e.g., collaborative learning).

Pre-Study: Understanding Course Contexts

We assumed that contextual norms shape the relevance and outcomes of stereotypes in the classroom. That is, stereotypes can be made relevant or irrelevant by the norms of the social context. Negative stereotypes against women, for example, appear to depress women's performance and participation in some learning contexts (e.g., physics and engineering) more than in others (e.g., biological sciences; Cheryan, Ziegler, Montoya, & Jiang, 2017; Leslie, Cimpian, Meyer, & Freeland, 2015). This suggests that the relevance and applicability of academic stereotypes vary across college classrooms.

Indeed, historical analyses of the university under study—a large, public research university in the mid-Atlantic region of the United States—found that

Table 1. Regression Results for Historical Analyses of Data From the 3 Most Recent Years Preceding the Current Study

Analysis and variable	Basic Physics for Engineers (*N* = 2,389)			Foundations of Biology (*N* = 4,393)		
	b	*SE*	*p*	*b*	*SE*	*p*
Analysis 1						
Semester Code 1	−0.09	0.04	.025	−0.32	0.04	.001
Semester Code 2	−0.09	0.04	.022	0.00	0.04	.931
SAT Verbal	−0.04	0.03	.136	0.08	0.02	< .001
SAT Math	0.44	0.03	< .001	0.14	0.02	< .001
High school GPA	0.61	0.04	< .001	0.23	0.04	< .001
Male	0.12	0.04	< .001	0.01	0.04	.821
White	0.13	0.05	.005	0.13	0.04	< .001
Analysis 2						
Semester Code 1	−0.09	0.04	.026	−0.32	0.04	< .001
Semester Code 2	−0.09	0.04	.027	0.00	0.04	.924
SAT Verbal	−0.04	0.03	.184	0.08	0.02	< .001
SAT Math	0.42	0.03	< .001	0.14	0.02	< .001
High school GPA	0.61	0.04	< .001	0.24	0.04	< .001
Male	0.13	0.04	< .001	0.01	0.04	.825
Asian	−0.02	0.05	.713	−0.13	0.04	.003
Black	−0.22	0.07	.002	−0.09	0.06	.131
Latinx	−0.17	0.10	.098	−0.04	0.09	.684

Note: In Analysis 1, ethnicity was used as a dichotomous variable (0 = non-White, 1 = White). In Analysis 2, dummy-coded estimates were used for each of the three main ethnic-minority groups, with White as the reference group. GPA = grade point average.

performance gaps varied across courses. Analyses examined course grades in Foundations of Biology and Basic Physics for Engineers—both of which were first-year-level gateway courses. After controlling for prior SAT/ACT scores and high school grade point average (GPA), analyses of the three most recent years of historical data (reported in Table 1) found that biology classrooms were associated with an emergent ethnicity grade gap ($b = 0.13$, $p < .001$, $d = 0.11$): White students earned higher course grades than non-White students after analyses controlled for high school preparation. However, there was no gender gap in biology ($b = 0.01$, $p = .821$, $d < 0.01$). By contrast, the same analysis in physics showed both gender and ethnicity gaps: Men and White students outperformed women and ethnic-minority students, *b*s = 0.12 and 0.13, *p*s < .006, *d*s = 0.12 and 0.13, for gender and ethnicity (White vs. non-White), respectively. Notably, in biology, students of Asian descent tended to underperform relative to White students ($b = -0.13$, $p = .003$, $d = 0.13$). However, this was not the case in physics, as there was no difference in average physics course grades between Asians and Whites ($b = -0.02$, $p = .713$, $d = -0.01$).

These analyses showed how patterns of underperformance varied across two science courses in ways that were consistent with research on inequity and stereotypes in the broader college culture in the United States (e.g., Cheryan et al., 2017). Thus, each context received a context-specific intervention. Whereas the intervention materials and analyses focused solely on ethnicity in biology (Study 1), they focused primarily on gender (and secondarily on ethnicity) in physics (Study 2).

Study 1

Method

Participants. Study 1 was facilitated by one instructor in the same course, Foundations of Biology, over four consecutive semesters with approximately 300 students each semester. The instructor taught the lecture and associated weekly discussion sections (also called *recitation sections*), in which the lecture course was divided into four sections of approximately 75 students each. Each semester, the research team randomly picked two sections to receive the intervention and two to receive business-as-usual control activities. We closed the study when, after four semesters, analyses found that the intervention was consistently effective and that a control condition was no longer ethically justifiable. Overall, 608 students from eight discussion sections were assigned to the *ecological-belonging* intervention (177 non-White, 431 White), and 607 students from eight contemporaneous discussion sections received *business-as-usual control* activities (171 non-White, 436 White). The full sample was 69% White, 21% Asian/Asian American, 7% Black/African American, and 4% Latinx American. The sample was 66% women.

Data sources and missing data. Data were obtained from instructor grade books that gave a comprehensive list of students enrolled at the beginning of the course, their attendance in discussion sections, their exam performance, and their course grade. Students completed an end-of-year survey in exchange for course credit. Finally, individual cases were anonymized and linked to the registrar's student data warehouse by a campus representative who was not associated with the study. This allowed us to obtain students' high school preparation records (high school GPA and standardized test scores) as well as their college enrollment records and cumulative GPA while safeguarding the confidentiality of students' records. The analyses used an intention-to-treat approach (Hollis & Campbell, 1999), meaning that if a student dropped the class either formally or informally after the intervention, any data that had been recorded up to that point were included in the analyses. All inferential tests were based on maximum likelihood estimation, which is a full-information approach that utilizes all available data.

Timing and random assignment. The intervention was delivered during the first week of classes for the semester. Checks on random assignment revealed no main or interactive effects of condition assignment on average levels of high school academic preparation across classrooms (i.e., standardized test scores and high school GPA; see the Supplemental Material available online).

Procedure.

Control condition. During the week of the intervention, students in the business-as-usual control condition participated in icebreaker activities to help them form social bonds with their peers. Evidence suggests that even superficial similarities with others can exert reliable effects on people's emotional and cognitive connection with those others (Brannon & Walton, 2013; Cwir, Carr, Walton, & Spencer, 2011). However, the business-as-usual activities lacked the intervention content that sought to foster an intersubjective understanding of the nature of belonging and adversity. Students formed groups of four on the basis of a superficial similarity (e.g., liking the same quote on the wall), created a biology-related team name together (e.g., "The Heterozy-goats"), drew a picture of their team's mascot, and presented their mascot to the class.

Intervention condition. The intervention condition differed from the business-as-usual control condition in that it was intended to change students' collective understanding of the nature of belonging, competence, and adversity in college. It did so through a carefully constructed series of activities involving a reflective writing exercise, exposure to student testimonials, and a semistructured group discussion. As in the control classrooms, students formed groups of four. The instructor then introduced and expressed the core intervention message that adversity early in college is quite common and can be overcome with persistence. Students then completed a reflective writing exercise on the challenges they had encountered so far in the transition to college and to consider how those challenges may improve over time (essays were anonymous to avoid making them evaluative; cf. Walton & Cohen, 2007). Next, students were exposed to three written stories attributed to older students. Adapted from prior research (Yeager et al., 2016), the stories were told in the first person and consistently involved a narrative in which students experienced initial adversity (e.g., getting a poor grade or having a hard time making friends), eventually began to turn things around (e.g., by finding a core group of friends to study with), and are now happy and successful. Because this study was focused on the ethnicity performance gap, the stories were attributed to students from different ethnic groups to help convey the universality of struggling across the spectrum of competence stereotypes. Thus, one quote was arbitrarily attributed to an African American student (i.e., "African American, [University name] Senior"), one to an Asian American student, and one to a White American student.

After reflecting on their own experiences and after learning about the normalcy of adversity from the stories, students engaged in a structured group discussion to help solidify the new norms in the same group setting in which they would work throughout the term. Because students knew their peers had received the instructions and testimonials together, we assumed they would proceed in their discussion with the intervention message being "given" in the background (Clark & Brennan, 1991) and that students would use this shared information in their conversations as a jumping-off point for sharing their own narratives (cf. Binning & Sherman, 2011; Schwarz, 2014). Groups were asked to discuss (a) why people often do not realize that so many students struggle with the transition to college and (b) how their lives might be different when they are juniors and seniors. Whereas the first prompt tacitly suggested that adversity is both common and often hidden beneath the surface, the second prompt suggested that students will progress through college and change along the way. By implying this knowledge in the premise of the discussion questions, we expected participants' answers to adopt the premises in order to cooperatively answer the question (Grice, 1975), and observations of the research team supported this assumption. After several minutes of discussion, the instructor reinforced the activity by asking volunteers to share with the full class what their group had discussed.

Group diversity. Across the four semesters, students formed 302 three- to five-person groups (149 intervention, 153 control; average group size = 4.03, *SD* = 0.52). To measure group diversity, we assigned a score reflecting the proportion of the group that was non-White (*M* = .28, *SD* = .23). Groups at −1 standard deviation in composition were effectively all White (28% of groups were homogenously White), whereas groups at +1 standard deviation in composition were effectively half minority (33% of groups had at least two minorities).

Survey measures. At the end of each semester, students completed a 10-min online survey in exchange for extra credit (*n* = 1,039). The analyses below focused on the measures that were included to assess students' perceptions of their collaborative work groups. See https://osf.io/cjqs6/ for the complete list of measures.

Perceived group competence. Three questions tapped perceptions of group competence (1 = *strongly disagree*, 5 = *strongly agree*; e.g., "My recitation team was effective at accomplishing its goals"; *M* = 4.23, *SD* = 0.70, α = .79).

Perceived group warmth. Perceptions of warmth from the group were assessed as the mean of four items (1 = *strongly disagree*, 5 = *strongly agree*; e.g., "Most members of my recitation team like me"; *M* = 4.20, *SD* = 0.65, α = .85).

Results

Analyses with Optimal Design 3.0 (Spybrook et al., 2011) revealed that the obtained design, modeled with two levels (groups of four students nested within 302 student work groups), yielded .80 power to detect effects (e.g., control vs. belonging) as small as *d*s of 0.17. To further increase the sensitivity of the analyses, we included controls for semester at Level 2 and controls for gender, ethnicity, and high school preparation (SAT math, SAT verbal, and high school GPA) at Level 1. All analyses reported below used these same control variables.

Intervention outcomes.

Course grades. Following the focus on differences between White and non-White students (including Asians and Asian Americans), we first examined the effect of the intervention on course grades as a function of ethnicity. A two-level regression analysis (students at Level 1 nested within their four-person work group at Level 2) revealed a main effect of the intervention, indicating that students in the ecological-belonging condition had higher overall course grades than students in the control condition (*b* = 0.15, *p* = .021, *d* = 0.12). Consistent with expectations, this effect was further moderated by participant ethnicity (i.e., an Ethnicity × Condition interaction; *b* = −0.40, *p* = .005, *r* = .08). As depicted in Figure 1, the intervention effect (control vs. belonging) was stronger among non-White (*b* = 0.44, *p* = .001, *d* = 0.36) than among White students (*b* = 0.08, *p* = .326, *d* = 0.07). Further simple-slopes analyses also revealed that in the control condition, there was a performance gap favoring White students (*b* = 0.34, *p* = .001, *d* = 0.28), but this gap was absent in the intervention condition (*b* = −0.05, *p* = .623, *d* = 0.04).

Analyses of the intervention effect as a function of gender revealed that, as in historical analyses, there was no gender gap in the control condition (*b* = −0.08, *p* = .601, *d* = 0.04). The intervention effect was similarly weak for both genders (Gender × Condition interaction, *b* = 0.08, *p* = .539, *r* = .02). Consequently, the intervention in biology addressed underperformance where it existed, among ethnic-minority students, but it had no effect as a function of gender, for which historical analyses showed no performance gap. We further parsed course grade into its two main components, exam performance and attendance, to see where the effects were strongest.

Exam performance. The primary component of student grades, average exam performance (85%–90% of course grade), showed no main effect of condition (*b* = 0.79, *p* = .276, *d* = 0.06), but once again there was an Ethnicity × Condition interaction (*b* = −4.19, *p* = .007, *r* = −.08). Simple-slopes analyses revealed that the intervention effect (control vs. belonging) was stronger among non-White students (*b* = 3.47, *p* = .014, *d* = 0.26) than White students (*b* = −0.71, *p* = .772, *d* = −0.05). Further simple-slopes analyses also revealed that in the control condition, there was a performance gap favoring White students (*b* = 3.81, *p* = .001, *d* = 0.29), and this gap was absent in the intervention condition (*b* = −0.37, *p* = .775, *d* = −0.03).

Attendance. A minor component of course grade (10%–15%) was attendance in the discussion section where the intervention took place. This was a counted variable with

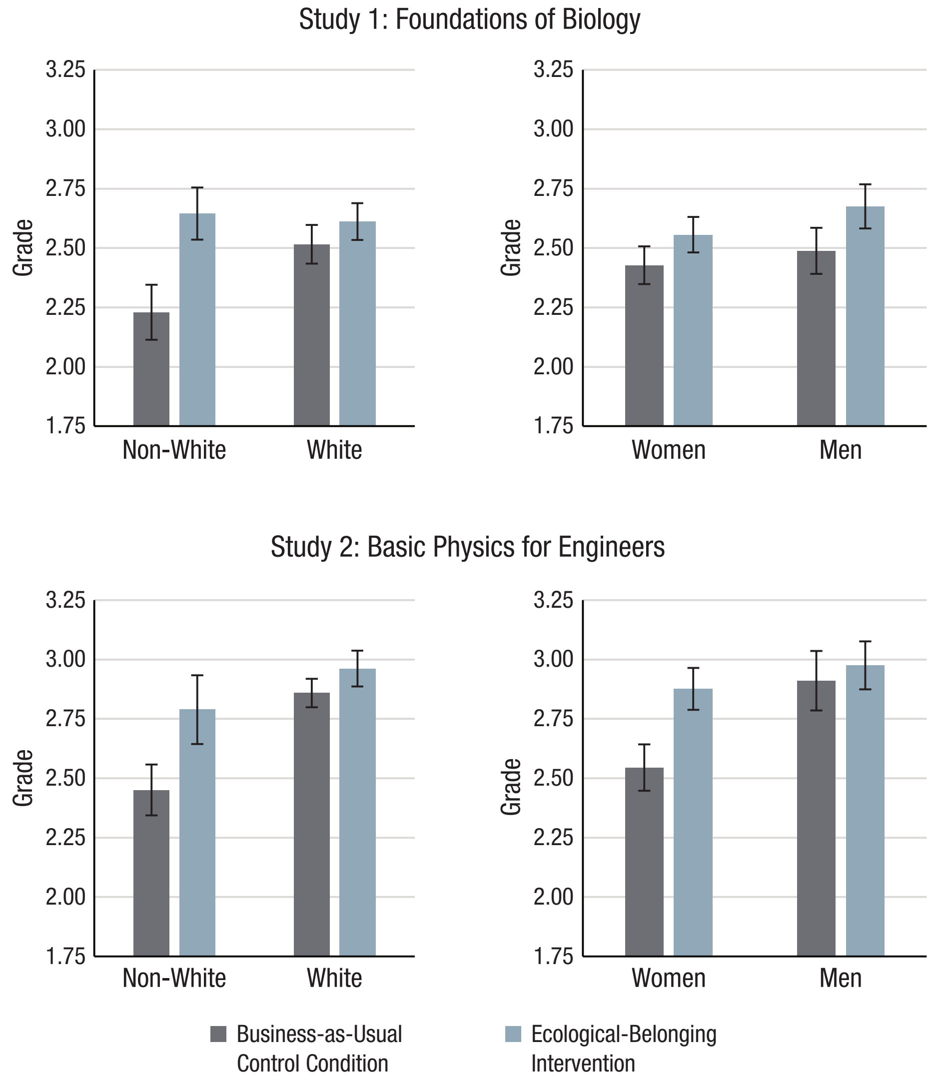

Fig. 1. Results from Study 1 and Study 2: average course grade on a 4-point scale (0.00 = F, 4.00 = A or A+) among students assigned to a business-as-usual control condition and an ecological-belonging intervention. Results are shown separately for ethnicity (left column) and gender (right column). Error bars reflect ±1 *SE*.

12 to 13 possible days, depending on the semester. We used the same predictors and approach as above, but we analyzed attendance with a two-level Poisson regression model using a log-link function with the variable-exposure assumption. The analysis revealed that students in the ecological-belonging condition had slightly higher average attendance than students in the control condition (b = 0.02, p = .078, d = 0.12). There was no Ethnicity × Condition interaction (b = −0.03, p = .267, r = −.03).

One-year persistence. To assess whether students persisted in college over the year following the intervention, we used the same predictors with a two-level binomial (Bernoulli) model (0 = *not persisting*, 1 = *persisting*). This revealed a main effect of condition on 1-year persistence (b = 0.70, p < .001, h = .22), indicating that students in the ecological-belonging condition were more likely to complete their courses (i.e., with a GPA greater than 0.00) in the two semesters following the intervention. There was no Ethnicity × Condition interaction (b = −0.15, p = .716, r = −.01), indicating that the effect held regardless of participants' ethnic group.

Cumulative GPA. From the registrar, we obtained students' most recent cumulative GPA, which ranged from 4 years (eight semesters) after the intervention for the first cohort to 2.5 years (five semesters) after the intervention for the most recent cohort. However, results showed no main effect (b = 0.07, p = .240, r = .07) and no Ethnicity × Condition interaction (b = −0.07, p = .303, r = −.03).

Exploring how the intervention affected learning. The weekly discussion sections where the intervention took place were rich with learning opportunities, some of which we were able to measure and explore. Learning opportunities included physically attending the discussion sections where the intervention took place (greater attendance is associated with improved academic outcomes; e.g., Credé, Roch & Kieszczynka, 2010), the presence of diversity in collaborative work groups (e.g., Page, 2008), and effective collaborations with peers (e.g., Kyndt et al., 2013). To examine whether intervention students were able to take better advantage of these learning opportunities, we explored whether each feature was more predictive of students' outcomes in the intervention condition compared with the control condition.

Attendance × Condition. We first explored whether attendance in the discussion sections was differentially predictive of long-term outcomes (retention and cumulative GPA). Analyses found that discussion-section attendance was more predictive of students' outcomes in the ecological-belonging condition than in the control condition and that these effects were not moderated by participant ethnicity. First, attendance predicted higher 1-year retention in the control condition ($b = 0.45$, $p = .023$, odds ratio, or $OR = 1.36$), but it was a slightly stronger predictor in the ecological-belonging condition ($b = 0.71$, $p < .001$, $OR = 1.51$; Attendance × Condition: $b = 0.26$, $p = .100$, $OR = 1.26$). Analyses revealed a similar pattern on cumulative GPA (Attendance × Condition: $b = 0.07$, $p = .012$, $r = .07$). This indicated that although attendance predicted cumulative GPA in the control condition ($b = 0.16$, $p < .001$, $r = .24$), it was a stronger predictor in the ecological-belonging condition ($b = 0.23$, $p < .001$, $r = .27$). Moreover, when we controlled for exam performance, the interaction effect remained ($b = 0.06$, $p = .025$, $r = .08$; simple slopes: *b*s = 0.01 and 0.06, *p*s = .603 and .004 for control and belonging, respectively).

Group Diversity × Condition. Despite evidence that group diversity does not generally improve performance (e.g., Eagly, 2016), it has the potential to be beneficial (e.g., Page, 2008). Exploratory analyses on course grades showed that in the intervention condition, the grades of ethnic-minority students benefited from greater group diversity. There was a three-way interaction (Group Diversity × Condition × Participant Ethnicity: $b = -1.70$, $p = .021$). For non-White students, there was no effect of diversity on grades in the control condition ($b = -0.14$, $p = .735$, $r = -.02$). But in the ecological-belonging condition, greater diversity was associated with higher grades ($b = 1.00$, $p = .038$, $r = .12$). There was no effect of diversity on grades among White participants (*b*s = 0.19 and −0.38, *p*s = .501 and .155, *r*s = .04 and −.08, for control and intervention, respectively). There were also no group diversity effects (main or interactive) on 1-year retention or cumulative GPA.

Analyses of the two survey measures revealed a Group Diversity × Condition interaction on perceptions of group competence ($b = 0.50$, $p = .026$, $r = .12$). A breakdown of this interaction showed that in the control condition, the more diverse a four-person group was, the less competent its members perceived it to be ($b = -0.29$, $p = .073$, $r = .10$). But in the ecological-belonging condition, this association was eliminated ($b = 0.18$, $p = .225$, $r = .07$). Neither factor interacted with participant ethnicity, nor was there a three-way interaction (Group Diversity × Condition × Participant Ethnicity: $b = -0.68$, $p = .157$, $r = -.04$). A parallel analysis on perceived warmth found no interaction effect (Condition × Diversity: $b = 0.26$, $p = .201$, $r = .07$). Thus, the diversity benefit to students in the belonging condition held on group competence (but not group warmth), and it held regardless of students' own ethnic group.

Group Effectiveness × Condition. Finally, a variety of evidence indicates that students benefit from working in effective collaborations (Kyndt et al., 2013). To see whether this was more likely to be the case in the belonging condition, we simply used students' survey measure of group competence as a proxy of group effectiveness. There was a Group Competence × Condition interaction on final course grades ($b = 0.21$, $p = .019$, $r = .08$). This showed that in the control condition, perceived competence was not associated with class grades ($b = -0.05$, $p = .394$). But in the intervention condition, higher competence was associated with higher grades ($b = 0.16$, $p = .010$). This effect still held ($b = 0.20$, $p = .024$, $r = .07$), even after we controlled for the Ethnicity × Condition interaction on course grades, and it did not interact with group diversity (Condition × Group Diversity × Competence: $b = -0.16$, $p = .458$).

Similarly, analyses found a Group Competence × Condition interaction on cumulative GPA ($b = 0.07$, $p = .012$, $r = .07$). As seen with course grades, group competence had no effect on long-term GPA in the control condition ($b = -0.03$, $p = .317$, $r = -.03$), but it predicted higher cumulative GPA among students who received the ecological-belonging intervention ($b = 0.07$, $p = .022$, $r = .07$). Together, these exploratory analyses suggest that the intervention improved students' ability to benefit from learning opportunities in their discussion section.

Study 2

Historical analyses of the last 3 years of student data revealed that, unlike in Foundations of Biology, there was a clear gender gap in Basic Physics for Engineers, with men consistently outperforming women. Also unlike in biology, in which women comprised a majority of students (66%), women were a minority in physics (32%), which is consistent with broader societal trends in engineering-focused courses (Cheryan et al., 2017). Thus, we sought to customize the intervention to this context. We conducted a focus group with women graduate students in physics to generate content for the student stories presented in the second phase of the intervention. See the Supplemental Material for more details and sample scripts of the physics intervention.

Study 2 was conducted during a single semester, after the conclusion of Study 1. Discussion sections were led by three different instructors (two Asian males and one White male), each of whom had two discussion sections that received the intervention and at least three other discussion sections that used the business-as-usual approach. Rather than the course instructor leading the intervention, as in biology, White women graduate students in physics visited the intervention sections. They delivered the intervention during the second week of classes to sections associated with each instructor (with their remaining sections serving as control conditions). The physics classrooms featured a smaller classroom environment (20–30 students vs. 70–80 students in biology). Unlike in biology, students' work groups in physics were not formalized, and so we did not obtain data on the groups' demographic composition.

Method

Participants and design. The physics sample was 32% women and 82% White, 12% Asian/Asian American, 4% Black/African American, and 4% Latinx American. The intervention was delivered to 169 students across six discussion sections (47 women, 122 men), with 438 students from 14 contemporaneous discussion sections serving as controls (149 women, 288 men, 1 unknown). As in Study 1, data were obtained from instructor grade books for course outcomes and from the registrar for 1-year persistence.

Analytic approach. Information on whom students worked with during discussion sections was not collected by instructors. Thus, we opted not to conduct multilevel modeling because of the small number of sections (six) that received the treatment. Because each instructor's lecture course had two treatment and at least three control discussion classrooms, we simply controlled for mean differences across students' instructors using dummy codes. Analyses used a nonparametric bootstrapping regression procedure that was robust to violations of normality assumptions.

Results

Course grades. As in Study 1, analyses revealed a main effect of the intervention, indicating that students in the ecological-belonging condition had higher overall course grades than students in the control condition (b = 0.16, p = .023, d = 0.11). Consistent with expectations, this effect varied across genders (Gender × Condition interaction: b = −0.30, p = .050, r = −.08). As depicted in Figure 1, the intervention effect (control vs. belonging) was stronger among women (b = 0.37, p = .004, d = 0.43) than men (b = 0.07, p = .326, d = 0.09). Further simple-slopes analyses also revealed that in the control condition, there was a performance gap favoring men (b = 0.37, p < .001, d = 0.46), which was reduced in the intervention condition (b = 0.07, p = .400, d = 0.08).

Analyses of ethnicity effects revealed that the intervention effect did not statistically differ between White students (b = 0.11, p = .100, d = 0.13) and non-White students (b = 0.31, p = .173, d = 0.36; Ethnicity × Condition interaction: b = −0.23, p = .191, r = .05). However, the pattern of results was similar to that seen in Study 1 (Fig. 1).

Exam performance. Analyses on students' average exam performance revealed that the main effect of condition (b = 0.14, p = .142, d = 0.12) was stronger among women (b = 0.43, p = .017, d = 0.35) than men (b = 0.03, p = .823, d = 0.02; Gender × Condition interaction: b = −0.41, p = .060, r = −.08). Further simple-slopes analyses also revealed that in the control condition, there was a performance gap favoring men (b = 0.62, p < .001, d = 0.50), which was reduced in the intervention condition (b = 0.21, p = .261, d = 0.17).

Attendance. Analyses of students' average attendance found a main effect of the intervention: Students in the ecological-belonging condition had higher average attendance than students in the control condition (b = 0.65, p = .004, d = 0.22). There was no Gender × Condition interaction (b = −0.71, p = .148, r = −.04).

One-year persistence. Results were consistent with Study 1 in that 1-year persistence was higher in the ecological-belonging condition than in the control condition (b = 0.84, p = .074, h = .16). There was no Gender × Condition interaction (b = −0.41, p = .736, r = −.01).

Attendance × Condition. As above, we examined the effect of attendance on students' long-term outcomes. First, attendance predicted 1-year retention in the control condition (b = 0.58, p < .001, OR = 1.79) and in the belonging condition (b = 1.28, p = .037, OR = 3.50). Despite being in the hypothesized direction, these slopes did not differ statistically (Attendance × Condition: b = 0.67, p = .265, OR = 1.96). However, replicating Study 1, analyses of cumulative GPA revealed a Condition × Attendance interaction (b = 0.37, p < .001, r = .15). This indicated that attendance was a stronger predictor of cumulative GPA in the ecological-belonging condition (b = 0.49, p < .001, r = .20) than in the control condition (b = 0.12, p < .001, r = .20). Moreover, when we controlled for exam performance in the model, the interaction effect was unchanged (b = 0.37, p < .001, r = .16), as attendance predicted GPA in the belonging condition (b = 0.38, p < .001, r = .16) but not in the control condition (b = −0.01, p = .781, r = .01).

General Discussion

An ecological-belonging intervention fostered equity in two college science gateway courses by lifting the course performance of previously underperforming students. Specifically, Study 1 found that the intervention lifted the performance of ethnic minorities in first-year biology, whereas Study 2 found it lifted the performance of women in first-year physics. These groups had been found to underperform in recent years compared with majority students, but the intervention statistically eliminated performance gaps. To help illustrate how these effects unfolded, we present evidence that the intervention had both psychological and ecological impacts. Psychologically, the intervention fostered engagement, as evidenced by greater attendance in the discussion section where the intervention occurred and greater 1-year college persistence. These findings replicate those of studies showing that social-belonging interventions can enhance college engagement, such as the number of office hours students attend and how often they participate in class (Walton & Cohen, 2007; Yeager et al., 2016). Notably, in the present research, the effects on engagement did not differ across demographic groups, meaning that the intervention increased engagement even among students who had not previously been underperforming.

By *ecological effects*, we mean that the intervention actually modified or interacted with elements in students' learning environments. Specifically, we found several pieces of evidence suggesting that the intervention allowed students to take better advantage of the learning opportunities they encountered over time. Our reasoning built on research showing that threatening social contexts can interfere with effective learning (Rydell & Boucher, 2017; Rydell et al., 2010) but that social-psychological intervention can lift this interference (Taylor & Walton, 2011). Evidence came from three sets of analyses.

First, Studies 1 and 2 found similar Attendance × Condition interactions on students' long-term GPA. Here, attendance in the discussion section was the learning opportunity (Credé et al., 2010), and attendance in the intervention condition was especially predictive of students' long-term college GPA. Notably, this finding held independently of students' exam performance in the course. This suggests that the intervention did not just impact what students learned in their course (as assessed by exam performance), but by changing the classroom ecology, it also may have changed how students learned, and this change may have endured long after the course ended.

Second, the intervention appeared to activate benefits of diversity in learning contexts (Page, 2008), as only in intervention classrooms did the diversity of students' four-person work groups predict higher perceived group competence and, among ethnic minorities, higher course grades. Third, being part of an effective work group can itself be thought of as a learning opportunity (Kyndt et al., 2013). And indeed, only in the intervention classrooms did group competence predict higher course grades and, more distally, students' cumulative college GPA.

Together, these findings suggest that the intervention increased engagement and that students' experiences in these more engaged settings predicted their academic outcomes. We argue that the ecological-belonging intervention may have changed the classroom context in ways that promoted student learning.

Changing the context of learning

Stereotype threat and belonging uncertainty can undermine students' ability to take advantage of learning opportunities. In one study, when students were reminded of negative stereotypes about their group during a memorization task, they had poorer memory for the learned items on a subsequent low-threat recall task (Taylor & Walton, 2011). In the present work, historical analyses were consistent with the possibility that stereotypes in the air may have hindered the performance of negatively stereotyped students. These students underperformed, even after analyses controlled for their high school preparation (e.g., high school GPA), suggesting that the gaps emerged during the course itself. We reasoned that if these students were underperforming because of stereotypes in the air, then addressing those stereotypes could help students realize their potential (Walton & Spencer, 2009).

The current work adds to research showing how social-psychological interventions may mitigate the impact of threat on learning and performance (for reviews, see Cohen & Sherman, 2014; Harackiewicz & Priniski, 2018; Walton & Yeager, 2020). We showed for the first time how this may be accomplished, not just by targeting students' subjective experience but also by targeting the social ecology. Research has highlighted the importance of social-ecological factors for shaping how people experience diverse contexts (Greenaway & Turetsky, 2020). By teaching students norms that encouraged them to see their own and their peers' adversity as normative and temporary, the ecological-belonging intervention appeared to enhance student engagement and create a context to support and reinforce the intervention message over time. The long-term intervention benefits on GPA—benefits that were not moderated by ethnicity or gender—suggest that students actually took the social ecology with them out of the context where the intervention took place. Recent evidence suggests that the benefits of social-belonging interventions may even persist after students leave college (Brady, Cohen, Jarvis, & Walton, 2020).

Future directions and limitations

The mechanisms by which more generalized benefits of interventions materialized (e.g., on long-term GPA) are not yet clear, but the ecological-level benefits of interventions have been documented in prior research (Powers et al., 2016). This prior work found that when classrooms had a high number of African Americans who had benefited from a self-affirmation intervention the previous year, White students got higher grades, too. This suggests that benefits from interventions may be contagious. Moreover, benefits of interventions may persist over time when they are reinforced by others in the social context (Cohen & Sherman, 2014; Yeager et al., 2019). Understanding how intervention effects catalyze, generalize, and persist are pressing topics for additional research.

Echoing Yeager and Walton (2011), we wish to emphasize that the ecological-belonging intervention cannot magically improve equity. Rather, the results prompt the need for research to understand whether and how classroom ecologies can influence the intervention's effectiveness. For example, we stressed above that understanding the different groups that underperform in different classrooms is a prerequisite for designing and evaluating effective social-psychological interventions that address underperformance (see Binning & Browman, 2020). One limitation here is that across both of our studies the intervention facilitators were White women, as facilitators from other genders or ethnicities may change students' experiences with the intervention (Dennehy & Dasgupta, 2017; Marx & Goff, 2005). The numerical and structural distribution of minorities in the setting may also be crucial moderators (Binning & Unzueta, 2013; Walton, Logel, Peach, Spencer, & Zanna, 2015).

Finally, we note the limited scope of the present studies. We did not have sufficient statistical power to address intersectional identities, nor did we focus on less visible identities, such as those related to sexual orientation and social class, that can bear on students' sense of belonging in classroom and college settings. By better understanding how the salience of various identities in classrooms can impact performance, we may be able to engineer diverse contexts to unlock students' latent potential.

Transparency

Action Editor: Leaf Van Boven
Editor: D. Stephen Lindsay
Author Contributions
K. R. Binning, N. Kaufmann, E. M. McGreevy, and O. Fotuhi developed the study concept. All authors contributed to the study design. E. M. McGreevy, E. Marshman, and Z. Y. Kalender executed the interventions. K. R. Binning, S. Chen, and L. Betancur analyzed the data. K. R. Binning drafted the manuscript. All authors provided critical revisions. All authors approved the final version of the manuscript for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This research was supported by the National Science Foundation Division of Undergraduate Education (Grant No. 1524575).
Open Practices
Data for these studies are available on request to the corresponding author. Materials have been made publicly available via OSF and can be accessed at https://osf.io/cjqs6/. The design and analysis plans for the studies were not preregistered. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620929984. This article has received the badge for Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

Kevin R. Binning https://orcid.org/0000-0002-5396-4183
Laura Betancur https://orcid.org/0000-0002-0412-2529

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620929984

References

Binning, K. R., & Browman, A. S. (2020). Theoretical, ethical, and policy implications for conducting social-psychological interventions to close educational achievement gaps. *Social Issues and Policy Review*, *14*, 182–216.

Binning, K. R., & Sherman, D. K. (2011). Categorization and communication in the face of prejudice: When describing perceptions changes what is perceived. *Journal of Personality and Social Psychology*, *101*, 321–336.

Binning, K. R., & Unzueta, M. M. (2013). Perceiving ethnic diversity on campus: Group differences in attention to hierarchical representation. *Social Psychological and Personality Science*, *4*, 500–507.

Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social Psychology*, *80*, 253–267.

Brady, S. T., Cohen, G. L., Jarvis, S. N., & Walton, G. M. (2020). A brief social-belonging intervention in college improves adult outcomes for Black Americans. *Science Advances*, *6*(18), Article eaay3689. doi:10.1126/sciadv.aay3689

Brannon, T. N., & Walton, G. M. (2013). Enacting cultural interests: How intergroup contact reduces prejudice by sparking interest in an out-group's culture. *Psychological Science*, *24*, 1947–1957.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, *143*, 1–35.

Cialdini, R. B. (2001). Harnessing the science of persuasion. *Harvard Business Review*, *79*, 72–81.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, *65*, 333–371.

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*, 272–295.

Cwir, D., Carr, P. B., Walton, G. M., & Spencer, S. J. (2011). Your heart makes my heart move: Cues of social connectedness cause shared emotions and physiological states among strangers. *Journal of Experimental Social Psychology*, *47*, 661–664.

Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry*, *22*, 231–246.

Dasgupta, N., Scircle, M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences, USA*, *112*, 4988–4993.

Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences, USA*, *114*, 5964–5969.

Eagly, A. H. (2016). When passionate advocates meet research on diversity, does the honest broker stand a chance? *Journal of Social Issues*, *72*, 199–222.

Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: Stereotype threat and distance in interracial contexts. *Journal of Personality and Social Psychology*, *94*, 91–107.

Greenaway, V. P., & Turetsky, K. M. (2020). Socioecological diversity and inclusion: A framework for advancing diversity science. *Current Opinion in Psychology*, *32*, 171–176.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics. Vol. 3: Speech acts* (pp. 41–58). New York, NY: Free Press.

Harackiewicz, J. M., & Priniski, S. J. (2018). Improving student outcomes in higher education: The science of targeted intervention. *Annual Review of Psychology*, *69*, 409–435.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*, *319*, 670–674.

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning: Do recent studies falsify or verify earlier findings? *Educational Research Review*, *10*, 133–149.

Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*, 262–265.

Marx, D. M., & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology*, *44*, 645–657.

Miller, D. T., & McFarland, C. (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and Social Psychology*, *53*, 298–305.

Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Powers, J. T., Cook, J. E., Purdie-Vaughns, V., Garcia, J., Apfel, N., & Cohen, G. L. (2016). Changing environments by changing individuals: The emergent effects of psychological intervention. *Psychological Science*, *27*, 150–160.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*, 287–290.

Rydell, R. J., & Boucher, K. L. (2017). Stereotype threat and learning. In J. M. Olson (Ed.), *Advances in experimental social psychology* (Vol. 56, pp. 81–129). San Diego, CA: Academic Press.

Rydell, R. J., Shiffrin, R. M., Boucher, K. L., Van Loo, K., & Rydell, M. T. (2010). Stereotype threat prevents perceptual learning. *Proceedings of the National Academy of Sciences, USA*, *107*, 14042–14047.

Schwarz, N. (2014). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. New York, NY: Psychology Press.

Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., & Raudenbush, S. (2011). Optimal design plus empirical evidence: Documentation for the “Optimal Design” software. Retrieved from http://hlmsoft.net/od/od-manual-20111016-v300.pdf

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.

Taylor, V. J., & Walton, G. M. (2011). Stereotype threat undermines academic learning. *Personality and Social Psychology Bulletin*, *37*, 1055–1067.

van Dijk, H., van Engen, M. L., & van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, *119*, 38–53.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, *92*, 82–96.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, *331*, 1447–1451.

Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a “chilly climate” transform women’s experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, *107*, 468–485.

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, *20*, 1132–1139.

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, *125*, 617–655.

Walton, G. M., & Yeager, D. S. (2020). Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science*, *29*, 219–226.

West, T. V., Koslov, K., Page-Gould, E., Major, B., & Mendes, W. B. (2017). Contagious anxiety: Anxious European Americans can transmit their physiological reactivity to African Americans. *Psychological Science*, *28*, 1796–1806.

Williams, K., & O’Reilly, C. (1998). The complexity of diversity: A review of forty years of research. *Research in Organizational Behavior*, *21*, 77–140.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., . . . Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*, 364–369.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They’re not magic. *Review of Educational Research*, *81*, 267–301.

Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., . . . Dweck, C. S. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences, USA*, *113*, E3341–E3348.