

Intervening to Support Belonging: A Randomized Controlled Trial of Strength-Promotion
and Threat-Reduction Approaches

Christopher Leland Williams
Saint Augustine, Florida

B.S.Ed., University of Virginia, 2018

A Dissertation presented to the Graduate Faculty
Of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Psychology

University of Virginia
May, 2023

Abstract

The higher education landscape in the United States is characterized by ongoing disparities in both academic (e.g., achievement) and well-being (e.g., psychosocial health) outcomes between students with underrepresented identities—including racially-ethnically minoritized (URM; Black, Hispanic, Native American, and/or Pacific Islander) students and first-generation students—and those with overrepresented identities—racial-ethnic (White and/or Asian) students and continuing-generation students. There are well-evidenced approaches to reducing these gaps with various social psychological belonging interventions. However, these interventions overwhelmingly focus on reducing threats, which do not work in all contexts and may produce unintended negative consequences. A promising alternative approach is to focus on promoting strengths, especially those that have developed as a function of students' underrepresented social identities and associated life experiences (e.g., resilience). In this study, I developed a novel strength-promoting intervention and tested it alongside a more traditional threat-reducing intervention in a randomized controlled trial ($N = 1,291$; 56 class sections). I found that that both interventions promoted mentorship (Cohen's $d = 0.54$) and reduced stress ($d = -0.29$) among URM students. I also found that the interventions exhibited heterogeneous effects across class sections, and some of this variance was explained by students' perceptions of their instructor's growth mindset. Exploratory analyses revealed additional positive effects of both interventions on the grades of URM participants in particular contexts, as well as positive effects of the strength-promoting intervention on the grades of first-generation students ($d = 0.49$). I discuss potential reasons for null effects on some primary hypotheses and how this study contributes to the wider literature on social psychological interventions.

Keywords: belonging, social psychological interventions, strength-based approaches

Intervening to Support Belonging: A Randomized Controlled Trial of Strength-Promotion and Threat-Reduction Approaches

In the United States, there are persistent and well-documented disparities in outcomes between students with social identities and backgrounds that have been underrepresented in the higher education system—including Black, Hispanic, Native American, and/or Pacific Islander and first-generation—relative to those with social identities and backgrounds that have been overrepresented in the higher education system—including White, Asian, and continuing-generation. These outcomes include educational achievement and attainment (Engle & Tinto, 2008; Museus & Quaye, 2009; National Center for Education Statistics, 2021, 2022a; Reardon, 2011, 2013) and physical, mental, and psychosocial well-being (Destin, 2019b; Destin & Debrosse, 2017; Gopalan & Brady, 2020; Lipson et al., 2016; Ma et al., 2016; McGee et al., 2019). These trends are especially pronounced in science, technology, engineering, and mathematics (STEM) fields (Barber et al., 2020; McGee, 2020), which support not only the life success of individuals themselves (Fayer et al., 2017; Langdon et al., 2011; National Center for Education Statistics, 2022bs) but also the economic well-being of the US more broadly (Giffi et al., 2018; President’s Council of Advisors on Science and Technology, 2012; Xue & Larson, 2015).

Although these gaps are complex and multiply determined, the role of student psychology is non-trivial (Cohen & Sherman, 2014; Garcia & Cohen, 2013). In fact, an increasingly robust body of evidence suggests that interventions that target the way in which students interpret their academic and social realities can help address long-standing disparities in education (Harackiewicz & Priniski, 2018; Yeager & Walton, 2011). Many of these interventions are effective because they target a basic human need: the need to belong (Walton &

Wilson, 2018). However, as the field of social psychological belonging interventions matures, there are at least two domains in which the research is underdeveloped. First, most of these interventions target the reduction of psychological threat. By psychological threat, I am referring to particular thoughts and beliefs—oftentimes instilled by an unjust society—that restrain individuals from being able to fully avail themselves of their learning environment. For example, the social-belonging intervention (Walton & Brady, 2020) reduces the threat of inferring that social and academic challenges experienced in college are indicative of one’s inability to belong in college. To do so, it communicates to students that such challenges are common and temporary. Another example of a threat targeted by these interventions is being aware of negative stereotypes about the intelligence of one’s social identity group. Recent theorizing suggests there may be negative and as-yet unexplored consequences of focusing entirely on threat-reduction (Covarrubias & Laiduc, 2022; see also Brannon & Lin, 2021). For example, it is possible that the social-belonging intervention’s message (that most challenges in college are common and temporary) supports students’ motivation and achievement, but also leads them to believe that structurally-determined inequities are instead individually determined. Concurrently, a nascent body of empirical research offers a promising alternative approach: the promotion of strength. To what extent should belonging interventions in education reduce threats to versus promote strengths of students with underrepresented social identities? When and where may one approach be more effective than the other?

Second, many existing belonging interventions exhibit heterogeneous treatment effects. That is, similar (or identical) interventions have impacts of varying sizes on similar (or identical) outcome variables (e.g., Borman et al., 2018; Walton et al., 2023). However, the contextual factors that explain this heterogeneity are not well-understood. What are these contextual

factors? Do they vary for a threat-reduction approach to intervention as opposed to a strength-promotion approach to intervention? What can these factors teach us about the structural changes that are necessary for long-term solutions to the existing inequities in higher education?

The present study provides insight into both of these domains. First, it develops and directly tests a novel strength-promoting belonging intervention against a traditional threat-reducing belonging intervention in a randomized controlled trial. Second, it directly tests and compares the contextual moderators of each intervention.

Social Psychological Belonging Interventions

One of the key contributions of the field of social psychology is its ability to be applied to improve lives (e.g., in educational contexts) through theoretically grounded and rigorously tested interventions (Wilson, 2006, 2011). Some of the most potent interventions operate by addressing individuals' persistent questions about their present and future levels of belonging in a given context (Walton & Brady, 2017). Importantly, the individuals most likely to face uncertainty about their belongingness are those whose identities are marginalized within those contexts, e.g., underrepresented and racially-ethnically minoritized (URM; Black, Hispanic, Native American, and/or Pacific Islander) individuals. This dynamic uniquely positions social psychological belonging interventions as tools for promoting equity and reducing disparities in educational outcomes between marginalized and privileged individuals. In the domain of postsecondary education, there are three examples of social psychological belonging interventions (for more, see Walton & Wilson, 2018, Table S3) that provide an illustration of the unresolved issues in the literature. These include interventions targeting social-belonging, self-affirmation, and difference-education.

The social-belonging intervention helps students reconstrue belonging-relevant experiences of adversity by communicating that many of the challenges faced in college are shared by most others and tend to improve with time (Walton & Brady, 2020; Walton & Cohen, 2007, 2011). The self-affirmation intervention offers students the chance to affirm their value in domains outside of education ((Layous et al., 2017; Sherman et al., 2021; Shnabel et al., 2013). The difference-education intervention provides students with a framework for understanding the different experiences they encounter in college relative to their peers as being a function of their different backgrounds (Stephens et al., 2014, 2015).

From the Reduction of Threat to the Promotion of Strength

Reduction of Threat

The social-belonging, self-affirmation, and difference-education interventions are all focused on reducing psychological threat. This broad approach has been shown to be very effective in promoting academic and well-being outcomes for many populations of students across a number of diverse contexts (Harackiewicz & Priniski, 2018). However, it may not capture the full potential of social psychological belonging interventions. Further, interventions borne of this approach may produce unintended negative consequences, as recently outlined by Covarrubias and Laiduc (2022). For example, they suggest that the social-belonging intervention may promote a “color-evasive” ideology due to its emphasis on the commonality of adversity experienced in college. To the extent that such a message de-emphasizes the role of legitimate and structural identity-based discrimination (such as via structural racism), it may lead students to see existing identity-based disparities through an individualistic lens. If all students experience the same levels of adversity, but there are still group-specific differences in achievement, it is not

unreasonable (the logic goes) to conclude that such differences are due to individual shortcomings.

With regard to the difference-education intervention, Covarrubias and Laiduc (2022) suggest that it promotes a multicultural ideology—by highlighting group-specific differences in experiences—that could result in its own negative consequences. For example, research has found multicultural messages to increase endorsement of race essentialism, which is the idea that the differences between individuals of different races are inherent to them as human beings (Wilton et al., 2019). Race essentialism is associated with racial stereotyping, justification of existing (and inequitable) social hierarchies, and other negative intergroup attitudes and behaviors (for a review, see Roth et al., 2023).

Despite these possibilities, to my knowledge the potentially negative consequences of threat-reducing belonging interventions have yet to be assessed. As such, it is possible that the benefits of these interventions accrue alongside as-yet unmeasured and un-tested detriments. This type of phenomenon is not unprecedented; research on resilience in individuals with marginalized racial-ethnic identities provides an example of how seemingly positive adjustment can come with a cost. In one study, low-income African American youth who demonstrated high academic and social competence were found to have high levels of allostatic load, an indicator of chronic physiological stress and dysregulation that is associated with longer-term negative health outcomes (Brody et al., 2013; Evans & Kim, 2012). Applying this to interventions, it is possible that the demonstrated positive effects of threat-reducing interventions on outcomes such as persistence (e.g., Yeager et al., 2016) and achievement (e.g., Walton & Cohen, 2011) occur alongside higher levels of mental and physiological stress, which may yield more distal negative effects on health that are not captured by the majority of intervention studies (Destin, 2019a).

Promotion of Strength

In contrast to the threat-reducing approach to addressing inequity, recent theorizing in social psychology has focused on the importance of recognizing, validating, and promoting the inherent value of groups that have faced marginalization and underrepresentation in mainstream societal institutions (Destin, 2019a; Destin et al., 2022). This perspective draws on scholarly traditions outside of social psychology, including models of community cultural wealth (Yosso, 2016), evolutionary-developmental psychology (Ellis et al., 2017), humanization (del Carmen Salazar, 2013), and validation theory (Rendón, 1994; Rendón Linares & Muñoz, 2011). This theorizing suggests that one approach to promoting marginalized individuals' inherent value is by identifying and lifting up the unique strengths that they develop as a function of their marginalized backgrounds. For example, individuals with marginalized backgrounds may develop high levels of resilience and creativity by regularly having to navigate societal institutions that were not developed with them in mind (e.g., historically and predominately White universities). Researchers are now beginning to test the causal effects of promoting individuals' "background-specific" strengths (BSS) on academic and well-being outcomes (Hernandez et al., 2021). These strengths are unique assets developed by individuals as a direct function of their marginalized backgrounds and the associated lived experiences.

For example, one set of studies experimentally tested the effects of a brief reflection activity in which students from low socioeconomic status (SES) backgrounds considered how those backgrounds may have allowed them to develop unique strengths (Hernandez et al., 2021). Relative to a control condition, students in the strengths condition reported more positive feelings about their social class background, greater inclinations to persist in the face of academic challenges, and higher levels of self-esteem. Another study conducted with refugees

found that a similar strength-based manipulation increased refugees' expectations of academic success, challenge-seeking, and engagement with online college courses (Bauer et al., 2021). In yet another study, low-SES college students briefly exposed to a professor who endorsed background-specific strengths were more likely than low-SES students in a control condition to see their backgrounds as a source of strength, interpret academic difficulty in a productive manner, and persist on a challenging academic task (Silverman et al., 2022, Study 3). Together, these studies provide proof-of-concept evidence for the causal relationship between reflecting on background-specific strengths and improved academic and well-being outcomes. However, there are important differences between the manipulations tested in these prior studies and the intervention that I designed and tested in the present study. For a summary of these differences, see Table 1.

As seen in Table 1, the existing literature on background-specific strengths has largely focused on students from low socioeconomic (Hernandez et al., 2021; Silverman et al., 2022) or refugee (Bauer et al., 2021) backgrounds. In the present study, I aimed to widen this literature by specifically focusing on students with underrepresented and racially-ethnically minoritized (URM; Black, Hispanic, Native American, and/or Pacific Islander) identities. There is correlational evidence that experiences with background-specific strength promotion among individuals with such identities, perhaps counterintuitively, may promote belonging in spaces in which belonging traditionally has been withheld from them, such as mainstream institutions of higher education (Brannon & Lin, 2021). However, as far as I know, there have not been any experimental tests of background-specific strength promotion specifically targeting URM students in higher education.

Table 1

Key Differences Between Prior Strength-Promoting Studies and the Present Study

	Sample(s)	Context	Experimental Manipulation	Primary DVs	Heterogeneity
Present study	Undergraduate students	In person, introductory STEM courses	30-45 minutes; 5 student anecdotes; 2 writing prompts; race-specific	Achievement and well-being	Heterogeneity analyses based on hypothesized contextual moderators
Bauer et al. (2021)	Refugees	Study 1: Online, imagined university Study 2: Online university	10 minutes; 3 student anecdotes; 2 writing prompts; refugee-specific	Online course engagement	No heterogeneity analyses
Hernandez et al. (2021)	Study 1: Lower-SES undergraduate students Study 2: Lower-SES Black and Latinx middle school students	Study 1: Online (via Prolific) Study 2: Online	10 minutes; no anecdotes; 2 writing prompts; SES-specific	Inclinations to persist	No heterogeneity analyses
Silverman et al. (2022)	Study 1: Educators Study 2: High school and undergraduate students Study 3: Undergraduate students	Study 1: Online, live professional development session Study 2: Online Study 3: Online, introductory psychology class	Excerpt from ostensible professor’s lecture; 1 writing prompt; SES-specific	Task persistence	No heterogeneity analyses

The Role of Context in Social Psychological Belonging Interventions

A key issue to understand with regard to social psychological belonging interventions is treatment effect heterogeneity, that is, the variance in the impact of a particular intervention on a particular outcome across different contexts (Weiss et al., 2014; Weiss et al., 2017). In what contexts do belonging interventions work? Some progress has been made in answering this question. As would be expected, traditional threat-reducing belonging interventions have been found to be more effective in contexts with higher levels of psychological threat (Borman et al., 2018; Hanselman et al., 2014). Further investigation has suggested in particular that threat-reducing interventions produce the largest positive effects when the contexts in which they are implemented have preexisting psychological threat (e.g., larger achievement gaps between URM and non-URM students) coupled with psychological opportunities (e.g., a supportive social environment) for a sense of belonging to be developed (Hammarlund et al., 2022; Walton & Yeager, 2020).

In my own research, I have identified contextual heterogeneity effects in line with this prior literature. For example, I found that an online, group discussion-based threat-reduction intervention improved the academic achievement of students in lower-level calculus classes but not higher-level classes (Williams et al., 2023). The former classes were a setting that may present higher levels of psychological threat due to the students' relative lack of familiarity and comfort with the quantitative material. This effect was stronger in classes where students perceived their peers to be especially warm and competent. These perceptions of the social context may represent psychological opportunities for the development of a sense of belonging (Walton & Yeager, 2020). In support of this interpretation, a measure of students' sense of classroom belonging mediated the intervention's positive effects on students' end-of-semester

course grades, such that students in the lower-level courses who were assigned to the intervention condition reported higher levels of classroom belonging than those in the control condition, and this higher belonging was then associated with improved grades (Williams et al., 2023).

Despite these advances, more work is needed to understand contextual heterogeneity, which is important not only because it offers insight into where student-level interventions will be most effective, but also because it can reveal structural markers of (dis)advantage. Because social psychological interventions target students' subjective construal of the objective world, features of the objective environment that enable or disable the interventions carry information about how that environment is subjectively experienced by students. In other words, objective indicators from the educational context influence students to the extent that they are perceived by those students, and changing those indicators could also change student perceptions. Consider a student who perceives that their instructor (in this case, an objective feature of the environment) does not believe that all students can do well in school. One way to change the student's perception is by directly intervening with the student, such as with a social psychological intervention. Another way to change the student's perception is to change the instructor's communication with the student. These two approaches—individual-level psychological change and context-level structural change (see Walton & Wilson, 2018, Figure 1)—are complementary, and both have their place in educational reform efforts to promote equity (Garcia & Cohen, 2013; Cohen & Sherman, 2014). In these ways, student-level social psychological belonging interventions can both mitigate some of the consequences of preexisting inequities while also providing a roadmap for addressing the system-level features that allow such inequities to persist.

However, because models of background-specific strength promotion have not yet been fully leveraged as social psychological belonging interventions, the contextual moderators of their (potential) treatment effects are still unknown. Do they differ from those of threat-reducing interventions? If so, how and why? The present study specified a priori hypotheses about the contextual moderators of the intervention effects of both a (novel) strength-promoting and (traditional) threat-reducing intervention, allowing for a direct test of heterogeneity and its determinants as well as a comparison of those determinants across the two psychological approaches.

To summarize, the goals of the present study were to (a) directly compare a threat-reducing and strength-promoting intervention in college STEM classes and (b) assess the moderating roles of student-level (i.e., URM status) and context-level (i.e., perceptions of instructors and peers) factors for both interventions. I conducted a randomized controlled trial (RCT) with undergraduates in engineering and calculus classes at a selective, predominately White public university. There were three conditions: a novel strength-promoting belonging intervention, a traditional threat-reducing belonging intervention, and a control condition. To produce the materials for the intervention and control conditions, I conducted a pilot study that followed a process informed by principles of design thinking (Razzouk & Shute, 2012; Yeager, Romero, et al., 2016; Rosenzweig et al., 2019a) and improvement science (Bryk et al., 2015; Rosenzweig et al., 2019b). The process centered undergraduate students' perspectives and engaged in rapid prototyping, iteration, and small-scale testing. In the following sections, I provide an overview of the pilot study, articulate how it informed the proposed study, describe each condition of the RCT, present the study's hypotheses, describe the study's method and

results, and conclude with a discussion of the study's contributions, limitations, and insights for further research.

Pilot Study

In the pilot study, my goal was to develop the materials for two social psychological belonging interventions tailored to the educational context in which I would ultimately test the interventions. The first intervention took the strength-promoting perspective of validation (Rendón, 1994; Rendón Linares & Muñoz, 2011). I consider it a strength-promoting intervention because the messages of acknowledging students' identity-relevant challenges were coupled with messages that affirm students' academic strengths. In particular, I aimed to create an intervention with the following core components:

1. Validation of challenging identity-based experiences allowed by structural inequality
2. Validation and affirmation of academic ability
3. Provision of strategies to deal with challenges

The second intervention took the threat-reducing perspective of the traditional social-belonging intervention (Walton & Cohen, 2007, 2011; Walton et al., 2017; Walton & Brady, 2020). The core components of this intervention are:

1. Communication that challenges are shared by others and temporary in nature
2. Communication that challenges do not mean one does not belong
3. Provision of strategies to deal with challenges

Both interventions were structured in accordance with the design principles of social psychological interventions in education (Yeager & Walton, 2011) and followed a format that is typical of belonging interventions (Walton & Brady, 2017). That is, they included three distinct

elements: narrative text that introduces the intervention's messages and frames the intervention activity as an effort on the part of the research team to better understand students' experiences; a series of anecdotes from actual upper-year undergraduate students that contain the intervention's core components; and a writing prompt that asks students to reflect on and more deeply process the messages communicated by the text and anecdotes. Each intervention required between 30 and 45 minutes to complete.

Intervention Prototyping and Development

To develop the validation (strength-promoting) and social-belonging (threat-reducing) interventions, my research team and I first conducted a series of focus groups with undergraduate students in the university's school of engineering. These focus groups were recorded and transcribed, and then two members of the research team reviewed the transcripts for student experiences that could be used as anecdotes in the intervention materials. To find anecdotes for the validation intervention, my team and I searched for student reports of social identity-relevant experiences in the school of engineering. At this stage in the study, I was interested in a broad array of social identities that tend to be underrepresented in STEM contexts, including race, social class, and gender. An example quote from this first round of review is "I think I've become a lot more feminist being in the engineering school because I notice that I'm alone often in my internships. I've been the only girl and you kind of have to figure out how to elbow your way out. Be loud and be in people's faces."

To find anecdotes for the social-belonging intervention, another researcher and I searched for student reports of difficulties they faced in the school of engineering and how those difficulties were shared by others and temporary in nature. An example quote is "... don't think you're alone in any aspect. Like you think you're going to fail this test, there're probably other

people who feel the same way... [it can] be a deterrent if you kind of feel like you're the only one."

After culling the raw qualitative data for relevant experiences, my team and I curated a series of anecdotes that stayed true to students' experiences while communicating the key messages of each intervention. We also produced the narrative text for each intervention that would serve as the framing. With my research team, I then followed an iterative design process in which I solicited multiple rounds of feedback on the framing and anecdotes from an interdisciplinary group of scholars and a diverse group of undergraduate students. After each round of feedback, we updated and refined the text and anecdotes until the research team and I felt confident that the materials of each intervention effectively communicated their intended core components.

Intervention Testing

Once the materials for each intervention were finalized, I ran a relatively small-scale field experiment to evaluate the interventions. At this stage, my team and I were unable to recruit a large enough sample to have adequate power to test both the validation and social-belonging interventions against a control condition. Given the novelty of the validation intervention and the large body of evidence supporting the social-belonging intervention (Walton & Brady, 2020), we prioritized the validation intervention in this first test. Another consequence of the relatively small sample size is that I was unable to conduct properly-powered heterogeneity analyses.

For the control condition, I developed a set of materials similar to those used in RCTs of the social-belonging intervention in the past (e.g., Yeager, Walton, et al., 2016). This condition mirrored the format of the validation and social-belonging interventions in its materials (narrative text, anecdotes, and writing prompt) and duration. It differed, however, in its content.

In particular, the control materials discussed the challenges of adjusting to the physical landscape of the university, rather than the challenges of adjusting to more belonging-relevant situations, such as making friends and performing well on exams.

Participants

The total sample size was 524 first-year (or transfer) undergraduate engineering students across 20 individual sections of an introduction to engineering class (11% URM [Black, Hispanic, Native American, and/or Pacific Islander]; 35% female; 10% first-generation). Of these students, 468 engaged with either the intervention or control materials, and 363 responded at all three time points (baseline survey, intervention/control activity, follow-up survey).

Procedure

All students in an introductory engineering class ($N = 739$ across 20 sections) were recruited to participate in the study in the fall semester of the academic year. At Time 1 (within the first two weeks of the semester), the students were given consent forms in class. If they consented to participate, they were given a link to a brief (ten minute) online baseline survey. At Time 2 (near the middle of the semester, in October), consented students were given a link in class to engage with either the validation intervention or control materials (about 45 minutes). They were randomized to condition ahead of time, and the randomization was blocked on race, gender, and class. If students had not yet made a consent decision (e.g., if they were absent at Time 1), they were given the opportunity to consent at Time 2 and still participate in the intervention or control activity. At Time 3 (within the last two weeks of the semester, at the end of November), the consented students were again given a link in class to a brief (ten minute) online follow-up survey. After the conclusion of the semester, administrative data on the consented students' GPA and demographic information were merged onto their survey

responses, and the data set was anonymized before being shared with the research team for analysis.

Results

We found no main intervention effect of the validation intervention on students' Time 2 belonging or their end-of-semester GPA. Similarly, we did not find any moderation of the intervention effect by students' identification with a group that is underrepresented in engineering (Black, Hispanic, Native American, and/or Pacific Islander; female; first-generation).

Follow-Up Investigation

In further investigating the results of this preliminary test of the validation intervention, I identified four factors that may explain the null effects and inform the next iteration of the intervention: sample size, intervention processes, control condition, and contextual heterogeneity.

Sample Size. The study's power to detect intervention effects was undermined by the relatively small sample size, which was further reduced by an unexpectedly low consent rate. The total number of students who were recruited was 739, yet only 524 of them consented and provided data for at least one timepoint (71%). As noted above, only 468 (63%) were present during the administration of the intervention or control materials (at Time 2). This starkly-reduced sample made my ability to test moderation by social identity (a key hypothesis) very underpowered. For example, only 48 students who participated in the study at Time 2 (when the intervention and control activities were administered) identified as Black, Hispanic, Native American, and/or Pacific Islander. After the block randomization, there were only 25 of these students in the intervention condition relative to 23 in the control condition.

Intervention Processes. To investigate the role of the theorized intervention processes in the null effects of the intervention, I conducted a qualitative analysis of the essays that students were asked to write as part of the intervention or control condition using established intervention fidelity methods (Brisson et al., 2020; Hulleman & Cordray, 2009; Murrah et al., 2017). The research team and I developed a set of codes designed to capture the key intervention processes, which map onto the core components described previously: Acknowledge Inequality, Ability Validation, and Strategy. In addition to the validation intervention processes, I included two codes to assess the presence of messages that come from the social-belonging intervention—Common and Temporary—because the control condition I used in the study asked participants to consider the challenges of adjusting to the physical landscape of the university, which is very similar to the social-belonging intervention’s prompt for students to consider the challenges of adjusting to the social and academic realities of the university. To evaluate the presence of the various codes in the participants’ writing, two research assistants (RAs) independently coded the essays. Of the 524 total participants, only 453 were present at Time 2 and had non-missing essays. To establish reliability, the RAs first coded 100 essays each. Their agreement ranged from 60% (on the Strategy code) to 92% (on the Ability Validation code). After meeting to discuss and resolve the discrepancies, the two RAs then coded another 100 participants. After this round of coding, their agreement ranged from 96% (on the Strategy code) to 100% (on all of the other codes). They then coded the remaining 256 participants with agreement ranging from 90% (Strategy) to 100% (Ability Validation). Discrepancies were resolved through discussion.

Based on the final set of codes, I found evidence that only two (out of three) of the core components of the validation intervention were communicated to students: Acknowledge Inequality and Ability Validation. Moreover, these components were not associated with

belonging and academic achievement in the hypothesized ways. The intervention had a positive effect on the students' acknowledgement of inequality in their essays, especially for students with marginalized racial-ethnic identities, indicating that students (especially Black, Hispanic, Native American, and/or Pacific Islander students) in the intervention condition were more likely to acknowledge the existence of structural inequality at the university in their written reflections, relative to control participants. However, contrary to my hypotheses, this acknowledgement was negatively associated with GPA and belonging. There is no evidence that the intervention lowered GPA, but the patterns associated with the coding suggested that a key component of this strength-promoting intervention—validation of negative experiences allowed by structural inequity—may not be a beneficial message for students to receive as they enter a new educational context.

The intervention had a positive effect on students' writing about ability validation, indicating that students in the intervention condition were more likely to endorse the existence of their own academic abilities. However, ability validation did not exhibit a significant relationship with grades or belonging. This is the most direct form of strength-promotion in the intervention, suggesting that an alternative or more varied set of strengths ought to be evoked among participants.

Also contrary to my hypotheses, the intervention did not have an effect on students' writing about strategies, indicating that there was no distinguishable difference between the writing of intervention versus control participants on strategies they could use to deal with challenges that they experience in college. However, among intervention participants, this code was positively associated with both grades and belonging. Although correlational, this suggests

that better communicating a more effective set of strategies may be an important component of a revised version of the intervention.

Control Condition. I found some evidence that the control condition I used was not fully inert. That is, the control condition in this study may have boosted the outcomes of students who were randomly assigned to it, undermining our ability to accurately detect the true effect of the validation intervention. Although this possibility cannot be directly tested, converging evidence suggests it may be true. For example, control condition students in this study reported higher levels of psychological well-being at Time 3 relative to two cohorts who completed surveys at a similar time point one and three years prior. Additionally, the control participants' mean level of belonging increased from the beginning to the end of the semester by 0.39 *SD*, whereas the intervention participants' mean level of belonging only increased by 0.28 *SD*.

Control participants were also more likely than intervention participants to write about the temporary nature of difficulties. Of the control participants, 68% included a description of difficulties as temporary in their essay, whereas only 53% of intervention participants did so. This is important because communicating the temporary nature of difficulties is a core component of the social-belonging intervention. As such, the control condition may have been operating to some extent as a more traditional threat-reducing social-belonging intervention, rather than a passive activity that is unrelated to students' sense of belonging.

Moderation by Student Subgroup. I did not find evidence for moderation of the validation intervention's (null) effects by any underrepresented student subgroup. One potential reason for this is that in attempting to represent as many underrepresented subgroups as possible, no one group in particular was impacted strongly enough to experience meaningful intervention effects.

Contextual Heterogeneity. Finally, as noted earlier, the sample size (of both students and class sections) was too small to conduct appropriate heterogeneity analyses. Hox and colleagues (2018) suggest that fewer than 20 level 2 units (here, the class sections) will produce potentially inaccurate fixed effect estimates and standard errors (p. 216), an issue that worsens when attempting to test moderation effects (e.g., between the intervention effect and marginalized social identity status). At the same time, there was some indication of section-level variability in baseline psychological variables, which would be expected to influence students' reactions to the intervention. For example, the range in mean belonging at baseline across sections is [4.63, 5.27]. This is a difference of 3.38 *SD* (using the *SD* of the aggregate belonging measure for each section). Note that I randomized within section and accounted for the class section differences in our analysis with fixed-effects and multilevel models. This approach allowed me to rule out group-level variance as a driver of intervention (null) effects, but it did not allow me to investigate potential intervention effect heterogeneity across classes (again, because I did not have sufficient power).

Although I could not inferentially test the presence of intervention effect heterogeneity, I could observe the descriptive patterns. These patterns suggested that there is section-level variability in intervention effects. For example, the contrast in GPA between intervention versus control participants ranged from -0.59 to 0.40 points. Among the 7 sections with 15 or more participants in each condition, the standard deviation of the section-level intervention effect (that is, the standardized difference in mean GPA between conditions) is 0.31 ($M = -0.10$). The pattern is similar for belonging.

Insights for the Next Phase of Research

The limitations noted above motivated a number of changes that I incorporated into the present study. The insights align with the domain in which each limitation was identified.

Sample Size. The present study substantially built upon the sample recruited in the pilot study by not only recruiting all students in the introductory engineering course, but also recruiting across all sections of the university's introductory calculus courses. It also amended the consent procedure to maximize the number of students who provided consent. For adequately-powered analyses of nested data (such as those in this study), it is suggested that there should be at least 30 Level 2 groups (here, class sections) with at least 30 individuals per group (Hox et al., 2018; Kreft, 1996). For the semester in which I conducted the present study, the introductory calculus courses spanned 41 sections and enrolled around 22 students per section. The engineering course spanned 15 sections and enrolled around 44 students per section. As described below, the final sample covered 56 sections with about 23 students per section. This sample size provided adequate power not only to test main and moderated student-level effects but also the cross-level interactions necessary to investigate contextual heterogeneity (see Hox et al., 2018, p. 216).

To retain as many students from the full sample as possible, the present study adopted a different procedure than that of the pilot for consenting students. In this approach, there was no baseline (Time 1) survey. Instead, all students in all 56 sections engaged with intervention or control materials as part of the course. After doing so, the students were then given the opportunity to allow or disallow the research team from using their data for the study. This approach has been used successfully in a prior project with a similar population at the same university (Williams et al., 2022). Whereas the pilot study's approach resulted in a 63% participation rate at Time 2 (during which the intervention/control materials were administered),

the new approach yielded a 75% participation rate. Relatedly, when the follow-up survey was administered, course instructors offered an incentive for completing it in the form of extra credit toward the students' course grade. This yielded an overall Time 1 to Time 2 response rate of 78%.

Intervention Processes. Based on the pilot study's results, I made two changes for the present study's intervention condition(s). First, I revised the strength-promoting intervention. Second, I added the threat-reducing social-belonging intervention to the design.

With regard to the strength-promoting intervention, the pilot study offered three key insights. First, acknowledging structural, identity-based inequality was not an effective intervention process. It may be the case that this particular format of intervention (e.g., brief, individual) is not appropriate for discussing the complex nature of systemic oppression (see Covarrubias & Laiduc, 2022, p. 743). Second, the intervention's specific strength-based focus on ability validation, while succeeding in increasing students' endorsement of their academic abilities, did not ultimately promote academic and well-being outcomes as expected. Third, the intervention did not adequately communicate strategies to students, even though writing about strategies was in fact related to achievement and well-being. To address these issues, the revised strength-promoting intervention deprioritized the discussion of structural inequality, empowered a diverse set of background-specific strengths, and connected strategies to strengths in an actionable way.

With regard to the threat-reducing intervention, I used the materials that were originally developed (but not tested) in the pilot study. This was possible due to the substantially increased sample size. The inclusion of this condition allowed for the hypothesis testing that theoretically guided this study at its inception.

Moderation by Student Subgroup. Rather than attempting to address the experiences of as many underrepresented student subgroups as possible, I decided to focus the revised intervention on one group in particular: underrepresented and racially-ethnically minoritized students (URM; Black, Hispanic, Native American, and/or Pacific Islander). I theorized that this would increase the potency of the intervention for URM students by being more relevant to their own experiences.

Control Condition. As noted above, there is some evidence from the pilot study that the control condition I used was not inert as intended. Rather, it may have operated as a type of threat-reducing belonging intervention. For the proposed study, I used a control condition that I have used successfully in prior work with a similar population of students (Williams et al., 2023). In this condition, students learned about and reflected upon various study strategies and skills they have developed or need to develop in order to be successful academically in college. The amount of reading and writing the condition required matched that required by the two intervention conditions, but there was no expected overlap in belonging-relevant psychological processes instigated by the interventions versus this control condition.

Contextual Heterogeneity. As noted above, the increased sample size allowed for adequately-powered tests of cross-level interactions, which were used analytically to investigate section-level heterogeneity of intervention effects. Additionally, in the present study I made specific predictions about the heterogeneity of intervention effects and structured the data collection to allow for the testing of those predictions. More specifically, I collected survey responses from students that measured their perceptions of the educational context (i.e., instructor and peers). Analytically, these perceptions were aggregated up to the section level to test their interaction with individual condition assignment and individual race.

Revised Strength-Promoting Intervention

The revised intervention follows the same general format of the other interventions: students read an introductory set of narrative text, review a series of anecdotes from upper-year students, and complete a written reflection on the messages contained in the text and anecdotes. Accordingly, the requisite amount of time was about the same (between 30 and 45 minutes). However, the materials now concentrate on the diverse set of strengths that individuals develop as a function of their backgrounds, with a particular focus on race/ethnicity. This approach aligns more with the growing literature on background-specific strengths, which has reported positive academic and well-being effects of making explicit the connection between individuals' backgrounds and their development of unique strengths (Bauer et al., 2021; Hernandez et al., 2021; Silverman et al., 2022). In particular, the revised intervention included four key components (for the logic model of the intervention, see Figure 1A):

1. Acknowledgement that there are potential deficit understandings of certain backgrounds in education
2. Description of background-specific strengths.
3. Examples of the benefits of background-specific strengths themselves
4. Examples of the relationship between background-specific strengths and beneficial strategies

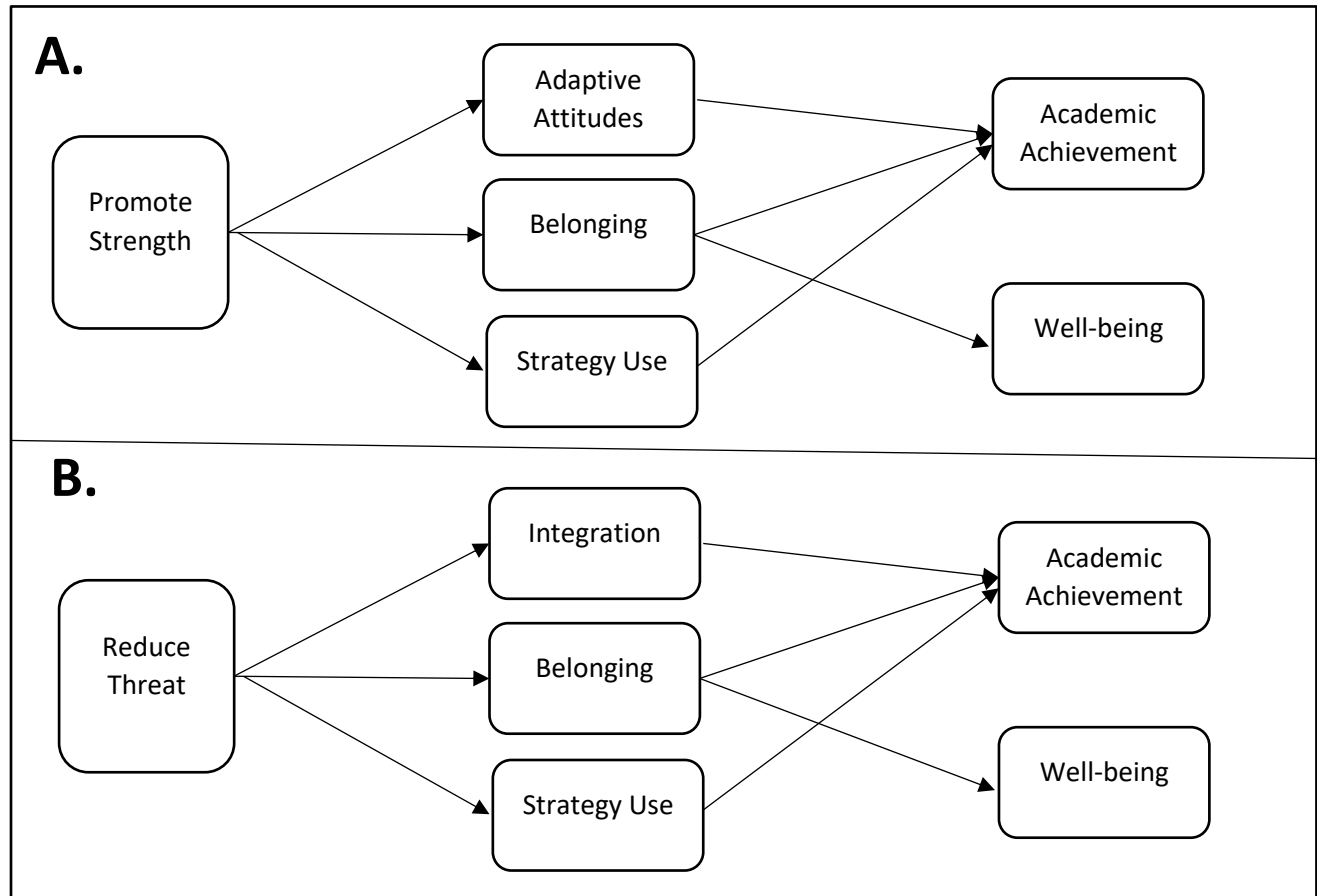
To update the intervention materials, I revisited the raw qualitative data from the original focus groups my team held with undergraduate engineering students. I produced anecdotes based on these data that describe how students have developed strengths as a direct function of their backgrounds. After developing the anecdotes, I solicited feedback on the anecdotes from both undergraduate STEM students and student-facing administrators in diversity offices at the

university. An example of one of the final anecdotes used in this study is “A strength I have developed due to my background would be determination. I grew up around Hispanic/Latinx workers with white vans filled with home remodeling tools and materials. My dad is one of these hardworking men who works rain or shine, whether it’s 20 or 100 degrees, outside or inside... The memories... are what motivate me to try my hardest at [school].”

With regard to the intervention’s planned focus on URM identities, my team and I decided to take a relatively light-touch approach. We made this decision based on findings from the pilot study, which suggested that drawing attention to ongoing structural inequalities faced by marginalized racial groups may increase URM students’ acknowledgement of inequality without necessarily empowering them. Another consideration was the fact that our study was going to recruit from all students in the introductory engineering and calculus courses (cf. Bauer et al., 2021; Hernandez et al., 2021), meaning that the majority of the sample would be students with non-URM identities. There is a large body of evidence suggesting the possibility of inducing reactance and/or backlash among dominant group members when faced with institutional messages of diversity and multiculturalism (Craig & Richeson, 2014; Dover et al., 2016; Wilkins & Kaiser, 2014; Kaiser et al., 2013; Plaut et al., 2011; Morrison et al., 2010) and/or messages of minoritized individuals describing experiences with discrimination (Dover et al., 2014; Kaiser, 2006; Kaiser et al., 2006; Kaiser & Miller, 2001; Kaiser & Miller, 2003). If the strength-promoting intervention prompted such reactions among individuals with more privileged racial-ethnic identities, it is possible that it would have yielded unintended negative consequences of the intervention for the very individuals whom it was designed to support.

Figure 1

Logic Models of the Strength-Promoting and Threat-Reducing Interventions



Note. Panel A is the logic model of the strength-promoting intervention. Panel B is the threat-reducing intervention. Adaptive Attitudes = adaptive academic attitudes (academic empowerment and the perception that difficulty is indicative of importance). Integration = academic and social integration (out-group closeness, academic comfort, and mentorship).

The writing prompts in the revised intervention were modeled off of the proof-of-concept demonstrations of the role of background-specific strengths in the literature (e.g., Hernandez et al., 2021). For example, the first prompt in the intervention begins with the following question: “What ‘background-specific strengths’ do you have that you have developed through your lived experiences?” It then offers examples of potential backgrounds that the students could draw upon and asks for the student to describe specific aspects of their strength(s), including how the strength(s) developed and how the strength(s) have helped and will continue to help the student

in school. This prompt is then followed by another prompt that asks students to describe background-specific strengths to a future student with the same background in order to promote further internalization of the intervention messages (e.g., “Now, please write a note to a future student from a similar background to yours about what background-specific strengths are, and why they matter...”).

Threat-Reducing Intervention

As described above, the threat-reducing intervention follows the approach taken by the social-belonging intervention, a well-researched social psychological belonging intervention (Walton & Brady, 2020). The core components of this intervention are the messages that challenges are shared by others, temporary in nature, and non-indicative of one’s ability to belong. They also include strategies to support students’ ability to deal with the challenges that they experience. The format (narrative text, student anecdotes, and written reflection) mirrors that of the strength-promoting intervention and requires a similar amount of time (30 to 45 minutes). For the logic model of the intervention as it is designed in our study, see Figure 1B.

The Present Study

Hypotheses

This study tested five primary hypotheses. Hypotheses 1, 2, and 3 were developed based on the literature that theorizes about and tests applications of background-specific strengths (Destin, 2019a; Destin, 2019b; Destin et al., 2022; Hernandez et al., 2021), difference-education (Stephens et al., 2014, 2015; Townsend et al., 2019), and social-belonging (Binning et al., 2020; Walton & Cohen, 2007, 2011; Yeager, Walton, et al., 2016; Williams et al., 2022).

Hypotheses 4 and 5 were developed based on the literature that investigates the interaction between students’ perceptions of the educational context and how those perceptions cultivate

adaptive psychological functioning (e.g., belonging and achievement motivation). This research has been conducted through contextual analyses of social-belonging interventions (Hammarlund et al., 2022; Walton & Yeager, 2020) and motivation-relevant identity processes (Browman & Destin, 2016; Silverman et al., 2022). The hypotheses are enumerated below.

1. Both the strength-promoting and threat-reducing interventions will improve the academic achievement and well-being of participants relative to the control condition. Effects will be stronger for students who carry marginalized racial-ethnic identities (i.e., Black, Hispanic, Native American, and/or Pacific Islander students) or are enrolled in lower-level calculus courses (i.e., beginner calculus I).
2. The strength-promoting intervention effects will be mediated by positive effects on participants' adaptive academic attitudes, belonging, and strategy use.
3. The threat-reducing intervention effects will be mediated by positive effects on participants' academic and social integration, belonging, and strategy use.
4. The strength-promoting intervention will be stronger in classes in which students perceive their instructor to have more of a pro-diversity attitude and greater recognition of students' identity-specific strengths.
5. The threat-reducing intervention will be stronger in classes in which students perceive their instructor to endorse more of a fixed versus growth mindset and their peers to be warmer and more competent.

Exploratory Analyses

Potential Negative Effects

Along with the primary hypotheses, the present study conducted exploratory analyses on the potential negative effects of the strength-promoting and threat-reducing interventions. A

recent review of threat-reducing interventions (namely, the social-belonging and difference-education interventions) conjectures that the social-belonging intervention promotes a “color-evasive” perspective that may lead recipients to be more likely to attribute societal inequalities to individual as opposed to structural forces (Covarrubias & Laiduc, 2022). Because I am using an adapted version of the social-belonging intervention in this study, I measured and investigated intervention effects on (individual) attributions for inequality.

Conversely, the difference-education intervention may promote a multicultural perspective that might lead recipients to greater endorsement of race essentialism, or the idea that the differences between individuals of different races are inherent to them as human beings (Covarrubias & Laiduc, 2022). Although my strength-promoting intervention is not a difference-education intervention, the explicit way in which it connects racial-ethnic identities to strengths may be subject to the same critique. As such, I measured and investigated intervention effects on race essentialism.

Effects of the Interventions for First-Generation Students

Although my primary hypotheses about the effects of the interventions centered on URM students, there is also good evidence that they may support first-generation (FG) students. As with URM students, FG students are underrepresented in STEM disciplines, and they are psychologically sensitive to similar types of threat and opportunity as URM students are.

Accordingly, the existing literature on both threat-reducing and strength-promoting interventions finds positive effects for samples including FG students. For example, a large ($N = 7,335$) scale-up of the social-belonging (i.e., threat-reducing) intervention in a context similar to that of this study found it to increase persistence of both URM and FG students (Yeager et al., 2016; for similar results, see Murphy et al., 2020). Similarly, an experimental strength-promoting

manipulation—in which a university professor was portrayed as valuing background-specific strengths—increased low-socioeconomic status (SES) college students' academic empowerment, adaptive interpretation of adversity, and persistence on a challenging academic task. Among these low-SES students ($N = 276$), 27% were first-generation (Silverman et al., 2022).

Given the similar (to URM students) psychological positioning of FG students, as well as the existing empirical evidence suggesting intervention benefits for lower-SES (including FG) students, I conducted exploratory analyses evaluating the interaction between condition and FG status. These analyses mirrored those conducted for URM students.

Method

Power and Minimum Detectable Effect Size

Because the present study had a ceiling on its possible sample size (i.e., all of the students in the engineering and calculus courses), I computed the minimum detectable effect size (MDES) of the study under a certain set of assumptions to communicate its power (Bloom, 1995). This value represents the smallest (standardized) intervention effect on an outcome that an experimental study has a good chance of detecting, if such an effect exists. By convention, I made the follow assumptions: alpha level is .05; test is two-tailed; power is 80% (this is the “good chance” noted in the MDES definition); and multiplier is 2.80.

Because the MDES depends on sample size, condition assignment, and rates of attrition and missingness in the data (among other factors), the MDES estimates are presented separately depending on the outcome variable under consideration. For space considerations, I have included the estimated MDES for only two of the primary outcomes: course grade and psychosocial well-being. I anticipate these to be similar, if not more conservative, MDES estimates relative to those of the other outcome variables.

Class Section Fixed Effects. Here I present the MDES estimates based on an assumption of class section fixed effects. That is, these are power estimates assuming that the variance in the outcome measures that can be attributed to differences between individual class sections is all noise and as such will simply be partialled out of the analyses.

The proportion of participants assigned to an intervention condition is 0.66, because in each section all students were randomly assigned to one of three conditions, two of which are intervention conditions. For the administratively-maintained course grade, the only source of attrition and missingness in the data was the proportion of students who did not consent to participate in the study. Based on a prior study with a similar design and scope to the present study, I assumed (conservatively) that consent-based attrition would be around 20% (in the prior study, it was 10%; Williams et al., 2023). Because the total possible number of students is roughly 1556, this assumption yields a Time 1 sample of 1245 participants. Based on the same prior study, I assumed the student-level covariates (race, gender, generation status) and section fixed effects would explain roughly 17% of the variance in course grade. Based on these assumed values, the MDES for course grade is 0.17 *SD* when comparing both intervention conditions to the control condition. In other words, if the two interventions together truly have an effect on course grade relative to the control condition, the smallest that effect can be in size in order to be detected with 80% power at an alpha level of .05 is 0.17 *SD*.

For the Time 2 survey outcomes, power is undermined slightly by higher levels of attrition and missingness and less-predictive covariates. Assuming an 80% consent rate and 15% attrition rate from Time 1 to Time 2 (the prior study's Time 1 to Time 2 attrition rate was 12%), there would be a 32% no-show rate (whereas with the administrative outcome, it was only 20%). These assumptions yield a sample of 1059 participants who would have responded at both Time

1 and Time 2. The student-level covariates of race, gender, and generation status are only assumed to explain about 3% of the psychosocial well-being measure (Williams et al., 2023). Based on these assumed values, the MDES for psychosocial well-being is 0.22 *SD* when comparing both intervention conditions to the control condition.

Class Section Random Effects. Here I present the MDES estimates based on an assumption of class section random effects, that is, assuming that there are heterogeneous intervention effects across class sections. In addition to the above assumptions and specifications, I made a conventional assumption that the rate of the intervention effect heterogeneity would be 0.10. This is the relative proportion of the variance of the intervention effect to the variance of the class section effect. Based on the existing literature, I assumed that student perceptions of context (their instructors and peers) would be able to explain at least 20% of the effect heterogeneity. Based on the aforementioned prior study (Williams et al., 2022), I assumed a section-level intraclass correlation coefficient of 0.01 for all outcomes. This is an indicator of the amount of outcome variance that is explained by differences between Level 2 units, which here are class sections. Based on these assumed values, the MDES is 0.17 *SD* for course grade and 0.22 *SD* for psychosocial well-being when comparing both intervention conditions to the control condition.

In sum, assuming the interventions have a true effect on achievement and/or well-being, the present study had 80% power to detect an effect around 0.20 *SD*. This is a reasonably-sized effect to expect, considering a meta-analysis of motivation interventions in education found the average standardized effect of social psychological belonging interventions to be 0.35 *SD* (95% CI [0.07, 0.63]; Lazowski & Hulleman, 2016).

Sample

Time 1. Using official course rosters, all students ($N_{students} = 1,724$) across five umbrella courses (Applied Calculus I; Applied Calculus II; Calculus I; Calculus II; Introduction to Engineering) were randomized within their section ($N_{sections} = 56$) to one of three conditions: strength-promoting intervention, threat-reducing intervention, and control. Of these students, 1,291 consented to have their data included in the study and were matched to administrative records (consent rate of 75%). They were distributed across condition such that 416 were in the strength-promoting intervention condition (32%), 419 were in the threat-reducing intervention condition (32%), and 456 were in the control condition (35%). Overall, the sample had the following demographic breakdown: 20% Asian, 8% Black or African American, 8% Hispanic, 49% White, 5% Multi-Race, 10% Other or Race and Ethnicity Unknown; 48% female; 15% first-generation. I created a URM category including underrepresented and racially-ethnically minoritized students (Black or African-American and Hispanic), which represents 18% of the sample. Students in the multiracial category were coded as URM if either of their identities was Black/African American or Hispanic.

Time 2. Among the 1291 consented students who were emailed a follow-up survey at the end of the fall semester, 1,009 responded (response rate of 78%). In a logit model regressing attrition ($0 = attrited$; $1 = did not attrit$) on the dichotomously-coded condition (strength-promoting, threat-reducing) and demographic (URM, female, first-generation) indicators, the only significant term was female, $b = 0.66$, $SE = 0.15$, $p < .001$. In other words, female students (response rate of 83%) were significantly more likely than male students (response rate of 74%) to respond to the follow-up survey. All other condition and demographic groups were roughly equally represented in the Time 2 sample.

The similarity of the actual sample sizes at Time 1 and Time 2 compared to the anticipated sample sizes used to calculate the MDES indicates that the present study had about 80% power to detect a true effect of about 0.20 *SD*. This represents a small-sized main effect (Cohen, 1988).

Procedure

In class during the third (math) or fourth (engineering) week of the semester, all students were given 45 minutes to engage with the materials of their assigned condition. At the end of the session, they were asked to provide their consent for the research team to use their data (survey responses and administrative data on demographics and grades) for the study. The responses of those who did not provide consent were not retained in the data set.

Survey data was collected at two timepoints. Time 1 occurred in the same session during which students engaged with condition materials and provided consent (or not). They were asked to respond to survey questions both before and after completing the intervention or control activities. Time 2 occurred about three months after Time 1 in the last week of the fall semester. Consented students were emailed a follow-up survey that took about 15 minutes to complete. Completion was incentivized with course credit from the students' section instructors.

At the end of the semester, administrative data was merged by the analytics office and the data set anonymized before being shared with the research team for analysis.

Measures

Unless indicated otherwise, all survey measures are Likert items on the following scale: 1 (*strongly disagree*), 2 (*moderately disagree*), 3 (*slightly disagree*), 4 (*slightly agree*), 5 (*moderately agree*), 6 (*strongly agree*). All participants saw all measures.

Time 1. During the 45-minute in-class session in which students engaged with their condition materials, there were a series of survey questions presented before and after the condition activities. These questions were used to conduct contextual heterogeneity analyses, evaluate the fidelity of the interventions, and check the manipulation of the interventions.

Pre-Condition. Prior to engaging with their condition materials, students responded to a series of two-item measures of their perceptions of the classroom context. In particular, they reported their perceptions that their instructor supports diversity (e.g., “My instructor in this class cares about all students equally, regardless of their diverse backgrounds.”; adapted from Browman & Destin, 2016; $r = .55$), endorses background-specific strengths (e.g., “My instructor in this class thinks that my unique background will help me succeed at [school].”; adapted from Silverman et al., 2022 and Hernandez et al., 2021; $r = .82$), and believes in the malleability of intelligence (e.g., “My instructor in this class seems to believe that students have a certain amount of intelligence, and they really can’t do much to change it.”; adapted from Muenks et al., 2020; $r = .64$). They also reported their perceptions of the extent to which their peers in the class are warm (“Most of the other students in this class are generally warm toward me.”) and competent (“Most of the other students in this class are generally competent academically.”; adapted from Binning et al., 2020; $r = .38$).

Post-Condition. After engaging with the condition materials, students responded to a few measures designed to assess intervention fidelity and check the manipulation. Intervention fidelity was measured with an attention check item included among the manipulation check items (described below) to identify any students who were not paying attention. The attention check item instructed participants to select a particular Likert response value (“Moderately disagree”) to indicate that they were paying attention to their responses.

There were four manipulation checks. First, to check the manipulation of the threat-reducing intervention, students were given a two-item measure of their endorsement of the central messages of the intervention (e.g., “At first, most people have concerns about belonging and fitting in at [school], but over time, those concerns tend to get better.”; adapted from Murphy et al., 2020; $r = .07$).

Second, also to check the manipulation of the threat-reducing intervention, students were given a six-item measure of their anticipated change in belonging (e.g., “At the end of your second year, to what extent do you think you will feel you belong at [school]?”; adapted from Yeager, Walton, et al., 2016; $\alpha = 0.94$). This measure is on a response scale ranging from 1 (*not at all*) to 7 (*an extreme amount*).

Third, to check the manipulation of the strength-promoting intervention, students were given a six-item measure of their personal endorsement of background-specific strengths (e.g., “I have gained certain strengths because of my background and lived experiences.”; adapted from Hernandez et al., 2021 and Silverman et al., 2022; $\alpha = 0.82$).

Fourth, to check the manipulation of both interventions, students were given a two-item measure of their perception that there are resources (i.e., strategies) available to them at their school to help them deal with social and academic adversity (“To what extent do you feel as though there are resources at [school] that you can take advantage of if you have (social/academic) difficulty?”; $r = .42$).

Time 2. During the last week of the semester, participants were asked to respond to a 15-minute follow-up survey. These questions were used to measure hypothesized intervention mediators and well-being outcomes. They also included the exploratory measures of potential negative consequences of each intervention.

Mediators. For the strength-promoting intervention, the hypothesized mediators are adaptive academic attitudes, belonging, and strategy use (see Figure 1A). The adaptive academic attitudes mediator is a composite index of academic empowerment and the perception that difficulty is indicative of importance. Academic empowerment was measured with an eight-item scale that covers perceived relative preparation (e.g., “I feel more academically prepared than other students at [school].”), academic efficacy (e.g., “I’m certain I can master the skills taught in my classes this upcoming year.”), and learner empowerment (e.g., “I can do things at [school] in a way that is right for me.”; Townsend et al., 2019). The perception of difficulty-as-importance was measured with two items (e.g., “If a school task is difficult, it means that it’s important for me.”; Oyserman et al., 2015; Silverman et al., 2022). The Cronbach’s alpha for all items included in the adaptive academic attitudes composite is 0.81.

Belonging was measured with a four-item scale (e.g., “I belong in Engineering/Mathematics at [school].”; Walton et al., 2015; $\alpha = 0.93$). The scale ranges from 1 (*strongly disagree*) to 7 (*strongly agree*).

Strategy use was measured with a scale that provides four examples of common help-seeking strategies, e.g., “Attended instructor/TA office hours.” Students were asked to indicate how often they had used each strategy, ranging from 1 (*Never*) to 5 (*7 or more times*).

For the threat-reducing intervention, the hypothesized mediators are academic and social integration, belonging, and strategy use (see Figure 1B). The academic and social integration mediator includes out-group closeness, academic comfort, and mentorship. Out-group closeness was measured by taking the average level of closeness that participants reported they feel toward all races/ethnicities other than their own (e.g., “How close do you feel to [race/ethnicity] students at [school]?”; Brannon & Lin, 2021). The response scale for this measure ranges from 0 (*very*

distant) to 10 (*very close*). Academic comfort was measured with two items (e.g., “I feel comfortable volunteering ideas in my classes at [school].”; adapted from Wilton et al., 2019; $r = .76$). The response scale for this measure ranges from 1 (*not at all true of me*) to 5 (*completely true of me*). Mentorship was measured with two items. The first item is “Thinking back on this past semester, have you developed a relationship with a mentor at [school] (formal or informal) that has been helpful to you and your academic and/or personal development?” (0 = *no*, 1 = *yes*). The response to this item is then multiplied by the follow-up item: “Think about the most important mentorship you have had this semester. How important has this mentorship been to you?” (1 = *not at all important*; 7 = *extremely important*). Those who did not report having a mentor received a value of zero for the composite.

Belonging was measured with the same four-item scale as that used for the strength-promoting intervention. Strategy use was measured with the same four-item scale as that used for the strength-promoting intervention.

Well-Being Outcomes. To assess participants’ well-being, I administered measures of psychosocial, mental, and physical well-being. Psychosocial well-being was measured with the eight-item Flourishing Scale (e.g., “I lead a purposeful and meaningful life.”; Diener et al., 2010; $\alpha = 0.93$).

Mental well-being was measured with three items from the Perceived Stress Scale (e.g., “In the last month, how often have you felt that you were unable to control the important things in your life?”; Cohen et al., 1983; $\alpha = 0.89$). The response scale for this measure ranges from 1 (*Never*) to 5 (*Very often*).

Physical well-being was measured with two items drawn from the National Longitudinal Survey of Freshmen (Brannon & Lin, 2021): “In the last 30 days, how often did a health problem

cause you to miss a day of school?” (1 = *never*; 2 = *once*; 3 = *just a few times*; 4 = *about once a week*; 5 = *almost every day*; 6 = *every day*) and “In general, how is your health?” (1 = *excellent*; 2 = *very good*; 3 = *good*; 4 = *fair*; 5 = *poor*).

Negative Consequences. Individual attributions for inequality was measured by the extent to which students endorse status-legitimizing beliefs (Levin et al., 1998; Wilkins et al., 2013; $\alpha = 0.65$). I adapted the scale to be slightly shorter and about the university specifically, rather than the United States in general. There are a total of six items across three subscales: system legitimacy (e.g., “[School] is a just system where differences in academic achievement between racial-ethnic groups reflect actual group differences.”), system permeability (e.g., “[School] is an open school where individuals of any race can get ahead.”), and Protestant work ethic (e.g., “Most people who don’t get ahead at [school] should not blame the school; they really have only themselves to blame.”). The response scale ranges from 1 (*strongly disagree*) to 7 (*strongly agree*).

Race essentialism was measured with six items from the Race Conceptions Scale (Williams & Eberhardt, 2008; $\alpha = 0.77$), which evaluates the degree to which individuals believe that race is a biological fact that does not change. An example item is, “Racial groups are primarily determined by biology.”

End-of-Semester. At the end of the semester, participants’ survey responses were merged onto administrative data that contains information on their demographic and background characteristics and achievement outcomes. The demographic and background data include race, gender, and generation status (i.e., students from families where neither parent/legal guardian has completed a four-year college degree). The academic achievement outcome is the grade that students earned in the STEM class in which they received the intervention.

Due to a highly disruptive school tragedy that occurred near the end of the fall semester in which this study was conducted, all students at the university were allowed to change their grading basis *after* they received their course grade. Instead of receiving a letter grade, they could convert their grades to credit (CR), general credit (GC) or no credit (NC). The conversion formula was as follows: Grades from 73 to 100 were converted to CR, grades from 60 to 72 were converted to GC, and failing grades were converted to NC. Accordingly, I assessed the effects of the intervention on both students' pre-conversion grades (the letter grade they received in the course, regardless of their decision of whether or not to convert the grade) and post-conversion grades (the grade that appeared on students' transcripts, which was either a letter grade or CR, GC, or NC). To analyze post-conversion grades, I assigned each student a numerical value as follows: If they received a letter grade, they were assigned a number on the traditional four-point scale (e.g., A = 4.0, B = 3.0, etc.). If they received a grade of CR, they were assigned a value of 3.3, because the minimum score threshold for CR is 73 and the maximum is 100; the middle value is 86.5; 86.5 (rounding up to 87) is associated with a grade of B+, which would have received a 3.3 on the traditional scale. If they received a grade of GC, they were assigned a value of 1, because the minimum score threshold for GC is 60 and the maximum is 72; the middle value is 66; 66 is associated with a grade of D, which would have received a 1 on the traditional scale. If they received a grade of NC, they were assigned a value of 0, because the minimum score threshold is 0 and the maximum is 59; the middle value is 29.5; 29.5 is associated with a grade of F, which would have received a 0 on the traditional scale.

Analytic Plan

Because the participants in this study are students clustered in sections, their individual data points cannot be assumed to be independent of one another, which is a necessary

assumption for traditional ordinary least squares linear modeling. To account for this non-independence, I analyzed the data with a series of multilevel models. These analyses were conducted using the lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), and lavaan (Rosseel, 2012) packages in R (R Core Team, 2020).

For all models, the same set of student-level (i.e., Level 1) covariates were included in order to increase the precision of the intervention effect estimate: race (0 = *non-URM [White or Asian]*; 1 = *URM [Black, Hispanic, Native American, and/or Pacific Islander]*), gender (0 = *male*; 1 = *female*), and generation status (0 = *continuing-generation*; 1 = *first-generation*, i.e., neither parent has a four-year college degree). A binary indicator of the type of STEM course was included at the section level (i.e., Level 2; 0 = *mathematics*; 1 = *engineering*).

Direct Effects

To estimate the direct effects of each intervention on the hypothesized (achievement and well-being) and exploratory (attributions for inequality and race essentialism) outcome variables, I fit a series of two-level multilevel models with random intercepts for each class section (see Equation 1). Separate models were fit for each outcome: course grade, psychosocial, mental, and physical well-being, attributions for inequality, and race essentialism. Dichotomously-coded terms for each intervention (0 = *control*; 1 = *intervention*) were included simultaneously. The estimated coefficient of each intervention term represents the direct effect of the respective intervention on the outcome.

$$\begin{aligned} Y_{is} &= \beta_{0s} + \beta_{1s}(PromoteStrength_{is}) + \beta_{2s}(ReduceThreat_{is}) + \alpha(X_{is}) + e_{is} \\ \beta_{0s} &= \gamma_{00} + u_{0s} \end{aligned} \quad (1)$$

In Equation 1, i represents students and s represents sections. Y_{is} is the value of the outcome variable for student i in section s . $PromoteStrength_{is}$ is the intervention contrast for the strength-promoting intervention, and $ReduceThreat_{is}$ is the intervention contrast for the threat-

reducing intervention. X_{is} is a vector of covariates, including race (0 = *non-URM* [*White or Asian*]; 1 = *URM* [*Black, Hispanic, Native American, and/or Pacific Islander*]), gender (0 = *male*; 1 = *female*), generation status (0 = *continuing-generation*; 1 = *first-generation*), and type of STEM course (0 = *mathematics*; 1 = *engineering*). The e_{is} term is the student-level error (i.e., the student-specific deviation from the within-section mean). The γ_{00} term is the weighted mean of the Y outcome across sections. The u_{0s} term is the section-level error (i.e., the section-specific deviation from the between-section mean). The α coefficients are modeled as fixed across sections (omitted in full form to save space).

The β_{1s} coefficient represents the estimated average treatment effect of the strength-promoting intervention, and the β_{2s} coefficient represents the estimated average treatment effect of the threat-reducing intervention, assuming the effects are constant across sections. It is worth noting that, somewhat surprisingly, there were several instances in which both interventions exhibited similar effects on particular outcomes (e.g., mentorship among URM students). In cases such as these, I modeled both intervention conditions as a single intervention contrast to increase the precision of the estimated effect.

In addition to the main effects of each intervention, I fit two sets of student-level interaction models. The first set used the full sample and tested the interaction term between the condition contrasts and race (0 = *non-URM*; 1 = *URM*). The second set used only the students in the math classes and tested the interaction term between the condition contrast and class type (0 = *lower level*; 1 = *upper level*). This interaction was an attempt to replicate the intervention x class type interaction that I found in a prior evaluation of a threat-reducing intervention in introductory calculus courses at the same university as that in the present study. In the prior study, I found that the intervention boosted grades for students in lower- but not upper-level

classes (Williams et al., 2023). In all interaction models, I followed the same procedure as in the main effects models: If both intervention interactions yielded similar effects, I combined them into a single intervention contrast and then refit the model with a single intervention x URM/class type term.

I also conducted exploratory moderation analyses using another student subgroup that is underrepresented in STEM disciplines: first-generation students. As with race, each intervention contrast was interacted with the dichotomously-coded indicator for generation status (0 = *continuing-generation*; 1 = *first-generation*).

To investigate contextual heterogeneity, I fit each model with a random slope for each condition contrast (strength-promoting alone, threat-reducing intervention alone, and both interventions together) and used a likelihood ratio test (LRT) to determine if the random slope significantly improved the fit of the random intercept model (which coerced the condition effect to be fixed across sections). If the random slope did improve the fit, then I tested the hypothesized contextual moderators of intervention effects by aggregating the student perception measures up to the section level and interacting them with the student-level condition contrast to predict the grade outcome (because it was our primary outcome measure). I also tested the three-way interaction among section-level perceptions, student-level condition, and student-level race (or, in an exploratory fashion, first-generation status).

Mediated Effects

If either (or both) of the two interventions demonstrated an effect on any of the hypothesized outcome variables and any hypothesized mediator, I tested the indirect effect of the intervention through the mediator. To do so, I fit a path model using full information maximum likelihood estimation, with standard errors clustered at the section level.

Results

In the following results, unless specified otherwise, all beta coefficients are in *SD* units of the outcome variable in question. Similarly, unless specified otherwise (e.g., due to issues with model convergence), all models include the following dummy-coded covariates: race (0 = *non-URM*; 1 = *URM*), gender (0 = *male*; 1 = *female*), first-generation status (0 = *continuing-generation*; 1 = *first-generation*), and STEM course (0 = *mathematics*; 1 = *engineering*).

Overview

As will be seen, the predicted effects of the interventions were nonsignificant on many of the outcome measures. There were no significant main effects of either intervention on any outcomes, nor did the hypothesized condition x URM status interaction exhibit effects on most outcomes. That said, my hypotheses concerning positive effects on some psychosocial measures and heterogeneity of intervention effects across class sections were supported (see Table 2 for a summary). In particular, I found that both interventions promoted mentorship and reduced stress among URM students. I also found that the interventions exhibited heterogeneous effects across class sections, and some of this variance was explained by students' perceptions of their instructor's growth mindset. Exploratory analyses revealed additional positive effects on the grades of URM participants in particular contexts, as well as positive effects on the grades of first-generation students.

Randomization Check

To evaluate the success of random assignment, I tested for differences in the mean level of each pre-condition survey composite (i.e., the four perceptions of context variables) across the three conditions. I also tested for differences in the distribution of each demographic group (race, gender, and generation status) across conditions. For the survey variables, I fit OLS models for

each composite with condition and section fixed effects as predictor variables and evaluated the significance of the F-test for the condition term. There were no significant effects of condition on any of the composites, indicating successful randomization.

For the demographic variables, I fit binomial generalized linear models for each dummy-coded indicator of group membership (URM, female, first-generation) with condition and section fixed effects as predictor variables. I then evaluated the significance of the likelihood ratio test for the condition term. The test of the condition term was non-significant for race and gender, but it was significant for generation status, $X^2(2) = 6.45, p = .040$. This significant effect indicates that the proportion of first-generation students in each condition is not fully balanced: The percentage of first-generation participants in the control and threat-reducing intervention condition was 17%, whereas in the strength-promoting intervention condition, it was 12%. Note that all future analyses control for generation status, so this difference should not have substantially influenced the results.

Attention and Manipulation Checks

Attention Check

Among all participants, 122 (9%) failed the attention check on the Time 1 survey. Failing this check (0 = *did not fail*; 1 = *did fail*) did not vary by condition, demographic characteristic, or any condition x underrepresented student subgroup interaction (URM, female, first-generation). Restricting the sample to participants who did not fail the attention check does not substantively change the pattern of results, so the remaining results are reported with the data from all participants.

Manipulation Check

Strength-Promoting Intervention. Contrary to hypotheses, there were no significant effects of the strength-promoting intervention on participants' endorsement of background-specific strengths (BSS; $b^* = 0.004$, $SE = 0.07$, $p = .954$), perceptions of social resources ($b^* = 0.03$, $SE = 0.07$, $p = .628$), or perceptions of academic resources ($b^* = 0.06$, $SE = 0.07$, $p = .388$). The null effects did not vary by URM status. As noted below, however, there was evidence that the strength-promoting intervention increased endorsement of background-specific strengths for first-generation students.

Threat-Reducing Intervention. Consistent with hypotheses, there was a significant main effect of the threat-reducing intervention on participants' endorsement of social-belonging messages, such that those in the threat-reducing intervention condition reported greater endorsement of the social-belonging messages than did those who were in the control condition, $b^* = 0.30$, $SE = 0.07$, $p < .001$. Similarly, there was a marginally-significant main effect of the threat-reducing intervention on participants' anticipated change in belonging, such that those in the threat-reducing intervention condition reported the anticipation of more growth in their sense of belonging than did those who were in the control condition, $b^* = 0.12$, $SE = 0.07$, $p = .081$. There was no significant effect of the intervention on perceptions of social resources ($b^* = 0.09$, $SE = 0.07$, $p = .202$), but there was a marginally-significant effect on perceptions of academic resources ($b^* = 0.12$, $SE = 0.07$, $p = .097$). Participants in the intervention condition ($M_{Adj} = 5.74$, 95% CI [5.63, 5.85]) reported higher average levels of feeling as though there are resources at their university that they can take advantage of if they face academic difficulty compared to those in the control condition ($M_{Adj} = 5.61$, 95% CI [5.50, 5.72]). None of these effects varied by URM status.

Hypothesis 1

Strength-Promoting Intervention

Main Effects. There were no significant main effects of the strength-promoting intervention on any of the academic achievement or well-being outcome variables, all p 's $> .100$.

Intervention x URM Status. There were no significant interaction effects of the strength-promoting intervention x URM on any of the academic achievement or well-being outcome variables, all p 's $> .100$.

Intervention x Class Level. Because the class level distinction only exists among students in math classes (lower-level calculus vs. upper-level calculus), these results apply only to the math sample: $n = 823$ for the grade outcomes and $n = 691$ for the well-being outcomes.

Academic Achievement. There was no significant interaction effect of the strength-promoting intervention x class level on the pre-conversion course grade ($b^* = -0.26$, $SE = 0.19$, $p = .177$), although there was a marginally significant intervention x class level effect on the post-conversion course grade ($b^* = -0.38$, $SE = 0.19$, $p = .051$).

As described below, however, the pattern of effects of both intervention conditions suggests that the intervention effects may be best analyzed by using a joint intervention indicator. When doing so, the two-way intervention x class level interaction is qualified by a significant three-way interaction among (joint) intervention, class level, and URM status. More details are provided below.

Well-Being. There were no significant interaction effects of the strength-promoting intervention x class level on any of the well-being outcome variables, all p 's $> .340$.

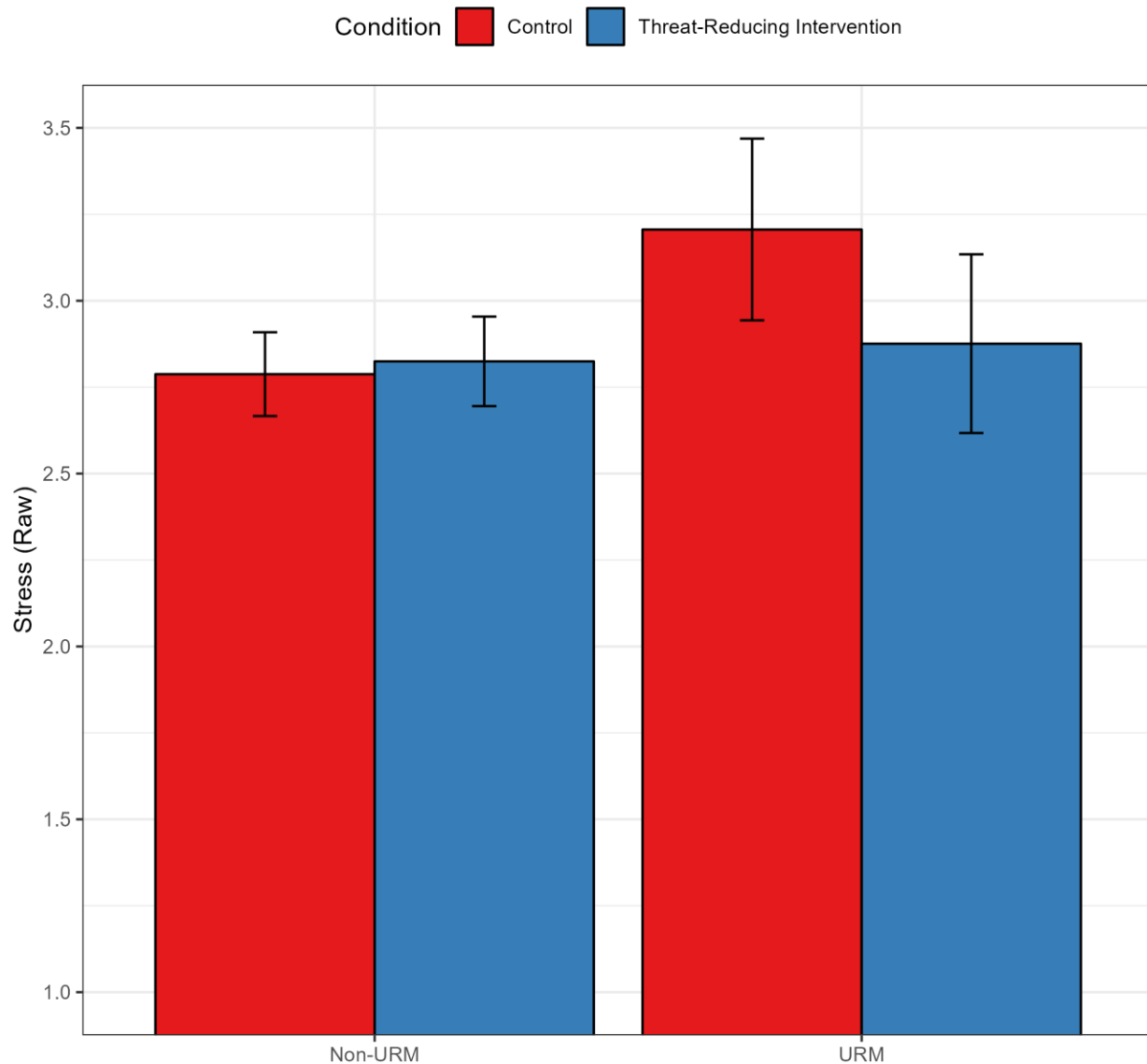
Threat-Reducing Intervention

Main Effects. There were no significant main effects of the threat-reducing intervention on any of the academic achievement or well-being outcome variables, all p 's $> .120$.

Intervention x URM Status. There were no significant interaction effects of the threat-reducing intervention x URM on any of the academic achievement or well-being outcome variables, all p 's $> .560$, with the exception of stress. The interaction effect of intervention x URM on stress was marginally significant, $b^* = -0.38$, $SE = 0.21$, $p = .065$. As seen in Figure 2, URM participants in the intervention condition ($M = 2.88$, $SD = 0.92$) reported a lower average level of stress than URM participants in the control condition ($M = 3.21$, $SD = 0.97$; Cohen's $d = -0.35$). Notably, the strength-promoting intervention x URM interaction was in the same direction ($b^* = -0.26$, $SE = 0.20$, $p = .183$) and as such the joint intervention x URM interaction was also marginally significant, $b^* = -0.32$, $SE = 0.17$, $p = .070$.

Intervention x Class Level. As noted above, these results apply only to the math sample: $n = 823$ for the grade outcomes and $n = 691$ for the well-being outcomes.

Academic Achievement. There were significant intervention x class level interactions on both pre- ($b = -0.40$, $se = 0.20$, $p = .044$) and post-conversion ($b = -0.45$, $se = 0.20$, $p = .024$) course grades, such that intervention participants in lower-level courses (pre-conversion grade: $M = 3.19$, $SD = 0.89$) scored lower than their control counterparts (pre-conversion grade: $M = 3.36$, $SD = 0.79$; $d = -0.20$). For visual depiction of the interaction, see Figure 3. Although not significant, the strength-promoting intervention x class level interaction term exhibited a similar pattern: $b^* = -0.26$, $SE = 0.19$, $p = .177$. Due to the size and direction of both interaction terms, I tested the joint intervention x class level interaction, which yielded a marginally significant effect for pre-conversion grade ($b^* = -0.32$, $SE = 0.17$, $p = .053$) and a significant effect for post-conversion grade ($b^* = -0.41$, $SE = 0.17$, $p = .016$). These effects were in the same direction as those of the threat-reducing and strength-promoting interventions x class level interactions when each term was modeled individually.

Figure 2*Time 2 Stress by Condition and URM Status*

However, this two-way joint intervention interaction was qualified by a significant three-way interaction of (joint) intervention x class level x URM status on pre-conversion grade: $b = -0.92$, $se = 0.41$, $p = .025$. The same three-way interaction on post-conversion grade was not significant, but it was in the same direction: $b = -0.54$, $se = 0.41$, $p = .191$. As seen in Figure 4, the intervention had no significant effect on grades, except for URM students in the upper-level

classes. Among those in the upper-level classes, the standardized difference in pre-conversion course grade between URM ($M = 2.39$, $SD = 1.46$, $n = 22$) and non-URM ($M = 3.29$, $SD = 0.82$, $n = 120$) students in the control condition was $d = -0.96$, whereas in either intervention condition, it was $d = -0.37$. This represents a 59% reduction in the achievement gap in upper-level courses.

Well-Being. There were no significant interaction effects of the threat-reducing intervention x class level on any of the well-being outcome variables, all p 's $> .550$.

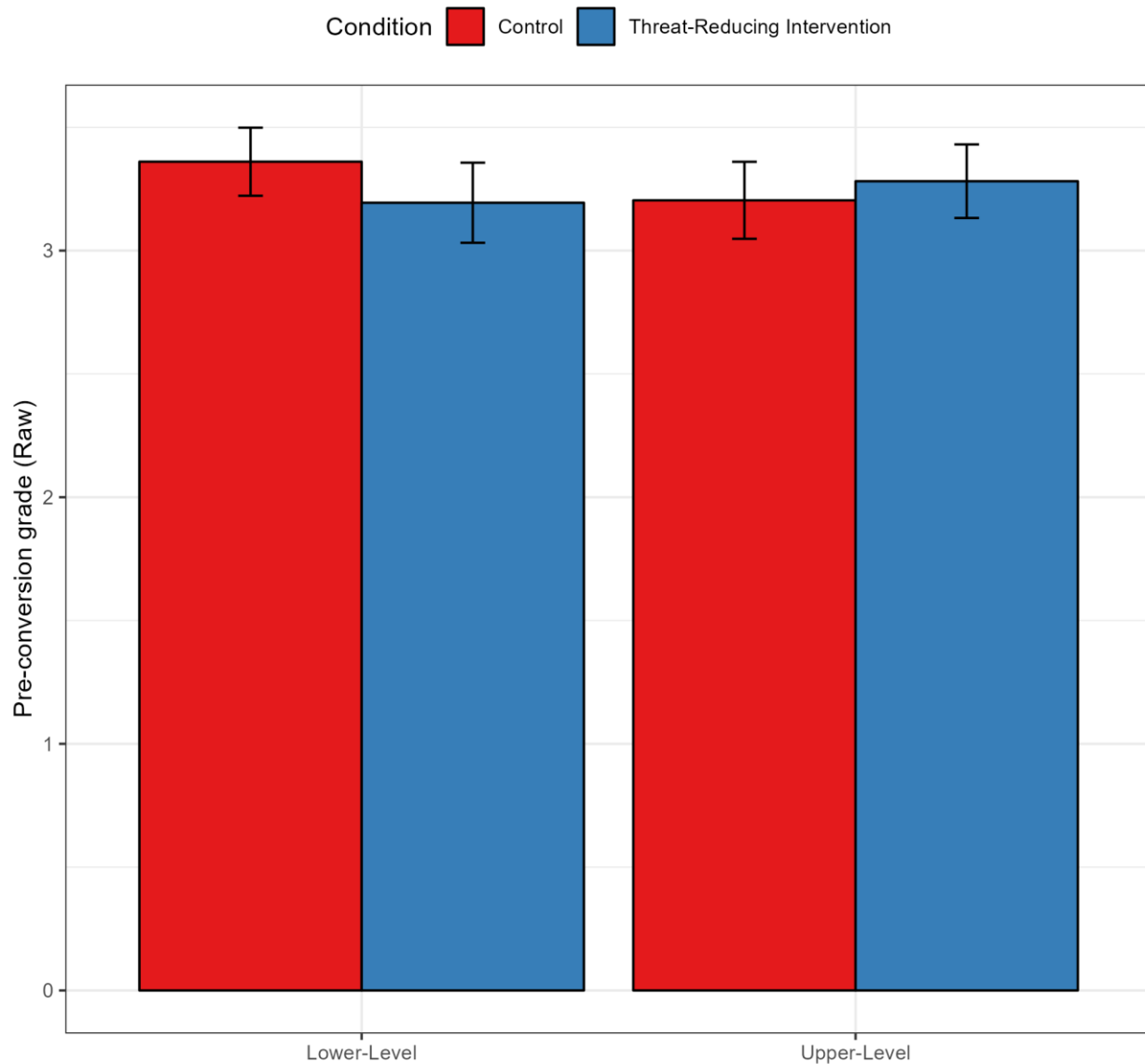
Hypothesis 2

There were no significant main or moderated (intervention x URM) effects of the strength-promoting intervention on academic achievement or well-being outcomes, so mediation was not tested. However, because both this and the threat-reducing intervention were designed to be belonging interventions, I did test intervention effects on the hypothesized belonging mediator variable. Contrary to hypotheses, there was a negative main effect of the strength-promoting intervention on belonging ($b^* = -0.20$, $SE = 0.08$, $p = .009$), such that those in the intervention condition ($M = 4.93$, $SD = 1.51$) reported lower levels of belonging than those in the control condition ($M = 5.22$, $SD = 1.51$; $d = -0.19$). This effect did not vary by URM status.

To better understand this negative effect, I tested the effect of the strength-promoting intervention on outgroup closeness (which was originally theorized to be a mediator of the threat-reducing intervention's effects). I found that the strength-promoting intervention also had a negative effect on outgroup closeness ($b^* = -0.18$, $SE = 0.08$, $p = .033$), such that those in the intervention condition ($M = 5.64$, $SD = 2.06$) reported lower levels of outgroup closeness than those in the control condition ($M = 6.00$, $SD = 2.12$; $d = -0.17$).

Figure 3

Pre-Conversion Course Grade by Condition and Class Level



Hypothesis 3

There were no significant main effects of the threat-reducing intervention on academic achievement or well-being outcomes, so main effect mediation was not tested. As with the strength-promoting intervention, I did test effects of the threat-reducing intervention on the hypothesized belonging mediator variable. The threat-reducing intervention did not have a

significant effect on belonging ($b^* = -0.13$, $SE = 0.08$, $p = .109$), nor was there moderation by URM status.

Because there was a negative effect of the strength-promoting intervention on outgroup closeness, I also tested this effect for the threat-reducing intervention. The threat-reducing intervention did not have a significant effect on outgroup closeness ($b^* = 0.001$, $SE = 0.08$, $p = .989$), nor was there moderation by URM status.

Mentorship

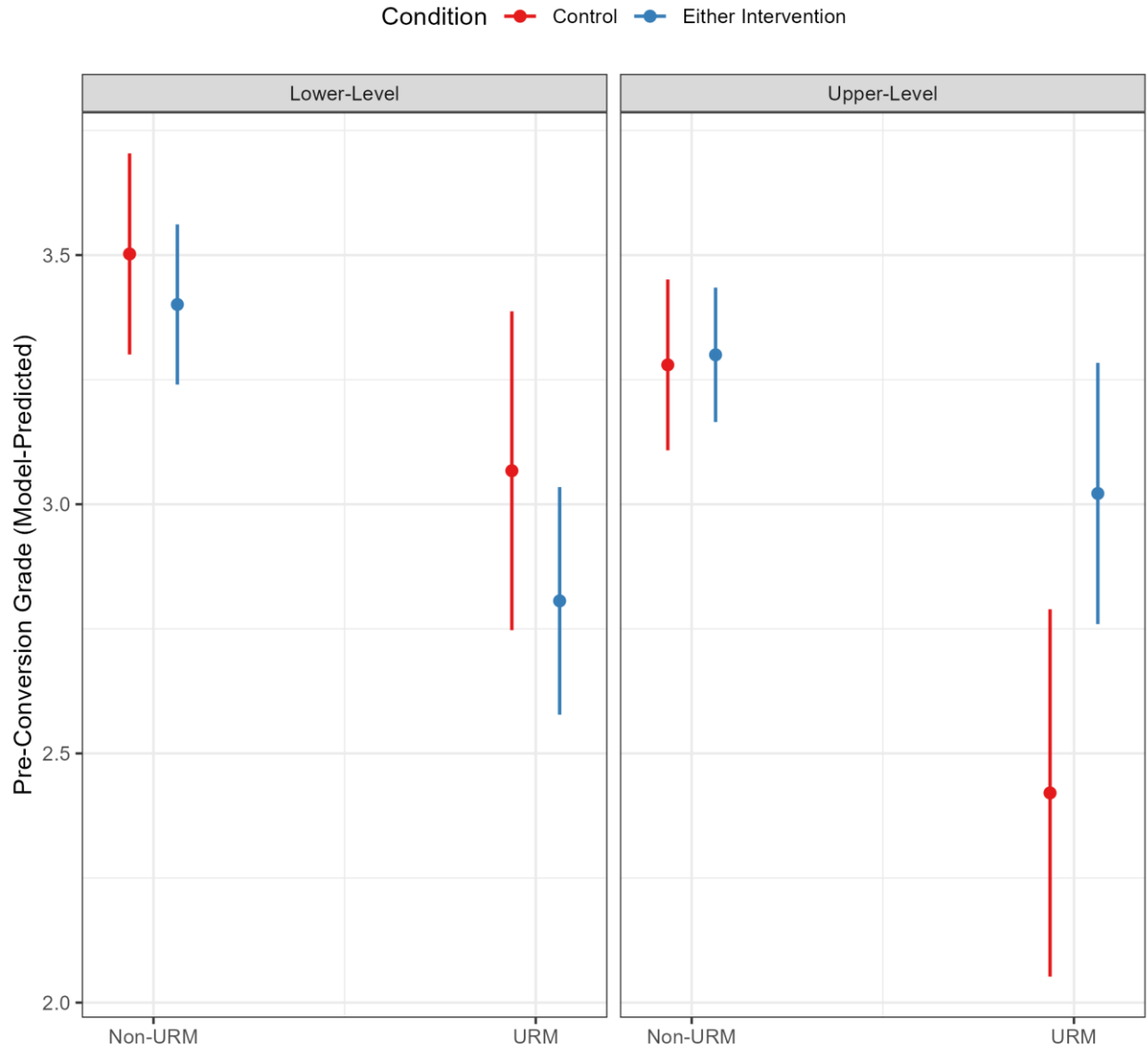
There was a significant threat-reducing intervention x URM effect on the hypothesized mentorship mediator variable ($b^* = 0.71$, $SE = 0.21$, $p < .001$), such that URM participants in the intervention condition ($M = 3.75$, $SD = 2.58$) reported higher average levels of mentorship than URM participants in the control condition ($M = 2.11$, $SD = 2.54$; $d = 0.42$). Notably, the strength-promoting intervention also yielded a significant interaction with URM status ($b^* = 0.66$, $SE = 0.20$, $p = .001$), and as such the joint treatment x URM interaction was also significant, $b^* = 0.67$, $SE = 0.18$, $p < .001$ (either treatment vs control among URM participants: $d = 0.54$). For a depiction of the interaction, see Figure 5.

Moderated Mediation

Because there were significant intervention x URM effects on stress (a well-being outcome variable) and mentorship (a hypothesized mediator of threat-reducing intervention effects), I tested a moderated mediation pathway from intervention to stress through mentorship. URM status was modeled as the moderator of the pathway from intervention to stress. The indirect effect estimate was not significant, $b = -0.03$, $SE = 0.02$, $p = .206$.

Figure 4

Pre-Conversion Course Grade by Condition, Math Class, and URM Status



Hypotheses 4 and 5

Contextual Heterogeneity of Intervention Effects

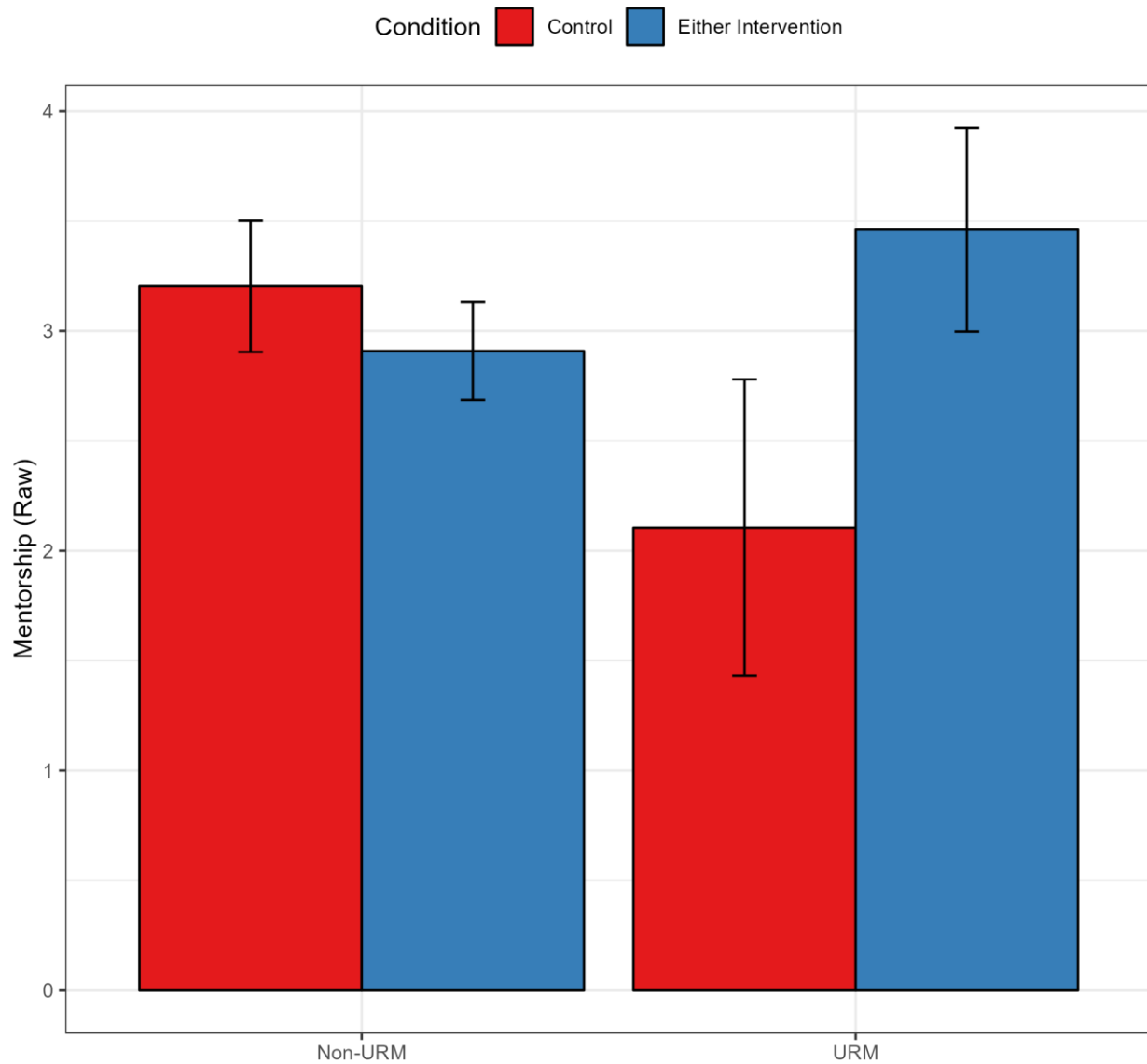
In order to explore heterogeneous intervention effects as a function of certain contextual features, there must be adequate variance in the outcome variable under consideration. One way to assess this variance is via intraclass correlation (ICC), which indexes the amount of between-

group (here, course section) variance that exists in the outcome. The ICCs of pre- and post-conversion grade were, respectively, 0.20 and 0.14. These values indicate that 20% of the pre-conversion grade variance and 14% of the post-conversion grade variance is between sections. To test if a meaningful amount of this variance in grade is due to intervention effect heterogeneity across sections, I fit two models per intervention contrast and compared them via likelihood ratio test (LRT). The first model, random intercepts, treats all effects as fixed across sections, but allows each section's mean level of the outcome variable (pre- or post-conversion grade) to vary. The second model, random slopes, not only allows each section's mean outcome to vary, but also allows the slope of the intervention contrast to vary across section. Functionally, this means that the intervention is free to have different estimated effects for each section. If the LRT is significant, it indicates that the random slopes model fits the data better than the random intercepts model. If this is the case, then section-level variables may be included as moderators of this varying intervention effect to help explain why the intervention effect is different across sections.

After running these comparisons on the pre-conversion grade outcome, I found that allowing either the strength-promoting or threat-reducing intervention effects to vary individually did not significantly improve the fit of the base random intercepts model. However, modeling the joint intervention contrast as random yielded a significantly better fit than its more restrictive random intercepts specification, $X^2(2) = 10.01, p = .007$. The same comparisons run on the post-conversion grade outcome did not yield any random slopes models that fit the data better than the random intercepts models. The discrepancy may be due to the relatively lower variance in post-conversion grade (ICC = 0.14) than pre-conversion grade (ICC = 0.20).

Figure 5

Time 2 Mentorship by Condition and URM Status



Because of this (and to remain consistent with earlier analyses), I report effects for both pre- and post-conversion grades for the following contextual heterogeneity analyses. Note, too, that Hypotheses 4 and 5 have been collapsed because there is no evidence that either intervention should be individually modeled as random across sections. Instead, I tested all hypothesized contextual moderators as interactions with the joint intervention contrast.

Perceptions of Instructor's Attitudes Toward Diversity

There was no significant interaction effect of section-level perceptions of instructor's attitudes toward diversity x intervention on pre- or post-conversion grades, both p 's $> .120$. Note that these models excluded the standard covariates because the models did not converge when the covariates were included.

Perceptions of Instructor's Endorsement of Background-Specific Strengths

There was no significant interaction effect of section-level perceptions of instructor's endorsement of background-specific strengths x intervention on pre- or post-conversion grades, both p 's $> .440$.

Perceptions of Instructor's Growth Mindset

There was no significant interaction effect of intervention x section-level perceptions of instructor's growth mindset on pre-conversion grade (including covariates), $b^* = 0.37$, $SE = 0.22$, $p = .107$. When excluding covariates, the effect was marginally significant, $b^* = 0.41$, $SE = 0.21$, $p = .055$. There was, however, a significant interaction effect on post-conversion grade, $b^* = 0.50$, $SE = 0.20$, $p = .016$. This model excluded covariates because it did not converge when they were included. As seen in Figure 6, the shape was such that in sections with instructors who had a higher level of perceived growth mindset, participants experienced a positive intervention effect on their grade, whereas in sections with instructors who had a lower level of perceived growth mindset, participants did not experience an intervention effect. Notably, when treated as a student-level (Level 1) variable, the interaction effect of perceptions of instructor's growth mindset x intervention was non-significant, $b^* = -0.01$, $SE = 0.06$, $p = .837$. In other words, the interaction effect was only observed when participants' perceptions of their instructor's growth mindset was aggregated up to the section level.

Perceptions of Peers' Warmth and Competence

There were no significant interaction effects of section-level perceptions of peers' warmth and competence x intervention on pre- or post-conversion grades, both p 's > .660.

Exploratory Analyses

Potential Unintended Consequences

Race Essentialism. As discussed earlier, some have argued that strength-promoting interventions might produce unintended negative effects, such as race essentialism, which is the idea that the differences between individuals of different races are inherent to them as human beings (Covarrubias & Laiduc, 2022). I found no evidence of such a negative effect. There were no significant main or moderated (intervention x URM) effects of the strength-promoting intervention on race essentialism, both p 's > .130.

Individual Attributions for Inequality. As discussed earlier, some have argued that threat-reducing interventions might also produce unintended negative effects, such as a greater endorsement of individual as opposed to structural reasons for societal inequality (Covarrubias & Laiduc, 2022). I found no evidence of such a negative effect. There were no significant main or moderated (intervention x URM) effects of the threat-reducing intervention on individual (as opposed to structural) attributions for inequality, both p 's > .350.

Moderation by Generation Status

The strength-promoting intervention appears to have had positive effects for first-generation (FG) students. First, there was a significant strength-promoting intervention x FG interaction on the strength-promoting intervention manipulation check, endorsement of background-specific strengths, $b^* = 0.57$, $SE = 0.20$, $p = .005$. There was also a significant strength-promoting intervention x FG interaction on both pre- ($b^* = 0.63$, $SE = 0.18$, $p < .001$)

and post-conversion ($b^* = 0.58$, $SE = 0.19$, $p = .002$) grades. As seen in Figure 7, FG participants in the intervention condition (pre-conversion grade: $M = 3.58$, $SD = 0.79$) earned higher grades than FG participants in the control condition (pre-conversion grade: $M = 3.11$, $SD = 1.07$; $d = 0.49$). There was no evidence of moderated mediation through BSS endorsement to pre- or post-conversion grade, both p 's $> .610$.

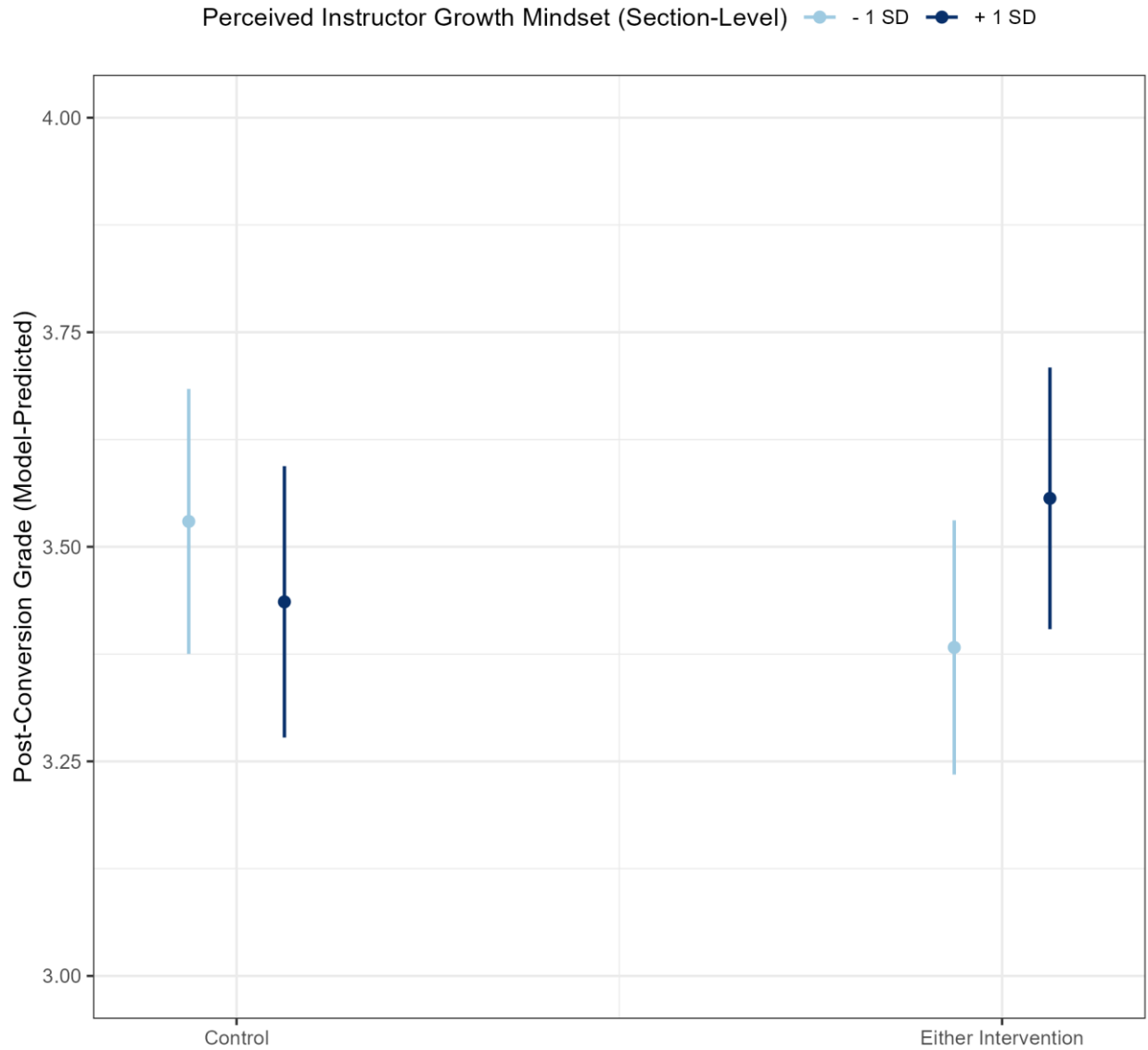
Exploratory analyses also revealed that the combined intervention was especially effective among first-generation students in class sections taught by instructors who were rated as having more of a growth mindset. That is, there was a significant three-way cross-level interaction among (joint) intervention x perceptions of instructor's growth mindset x FG status, $b^* = 1.17$, $SE = 0.47$, $p = .014$. As seen in Figure 8, in sections with instructors who had a higher level of perceived growth mindset, FG participants experienced a significant positive intervention effect on their grade, whereas in sections with instructors who had a lower level of perceived growth mindset, FG participants did not experience an intervention effect. This interaction was marginally significant when using pre-conversion grade as the outcome variable, $b^* = 0.81$, $SE = 0.46$, $p = .080$. The three-way interaction tested with perceptions of instructor's growth mindset modeled as a student-level (Level 1) variable (and post-conversion grade as the outcome) yielded a marginally significant effect, $b^* = 0.28$, $SE = 0.15$, $p = .069$. When using pre-conversion grade as the outcome variable, the effect was non-significant, $b^* = 0.13$, $SE = 0.15$, $p = .385$.

Discussion

In a randomized field experiment, I tested the effects of strength-promoting and threat-reducing social psychological interventions on well-being and achievement outcomes in a sample of undergraduate students enrolled in introductory STEM courses.

Figure 6

Post-Conversion Course Grade by Condition and Section-Level Perceptions of Instructor Growth Mindset



Some of the initial hypotheses were not confirmed; for example, there were no significant main effects of either intervention on any of the outcome variables.

Overall, however, pre-registered and exploratory analyses revealed positive effects of the interventions. The threat-reducing intervention reduced stress among URM participants and,

along with the strength-promoting intervention, increased mentorship among URM participants relative to those in the control condition. Also as hypothesized, I found that the interventions exhibited heterogeneous grade benefits across individual STEM course sections. The intervention effect variance across sections was partly explained by students' perceptions of their instructor's growth mindset, although in the opposite direction than what was hypothesized. That is, among courses with instructors who were rated as having more of a growth mindset, the intervention had a stronger positive effect on course grades.

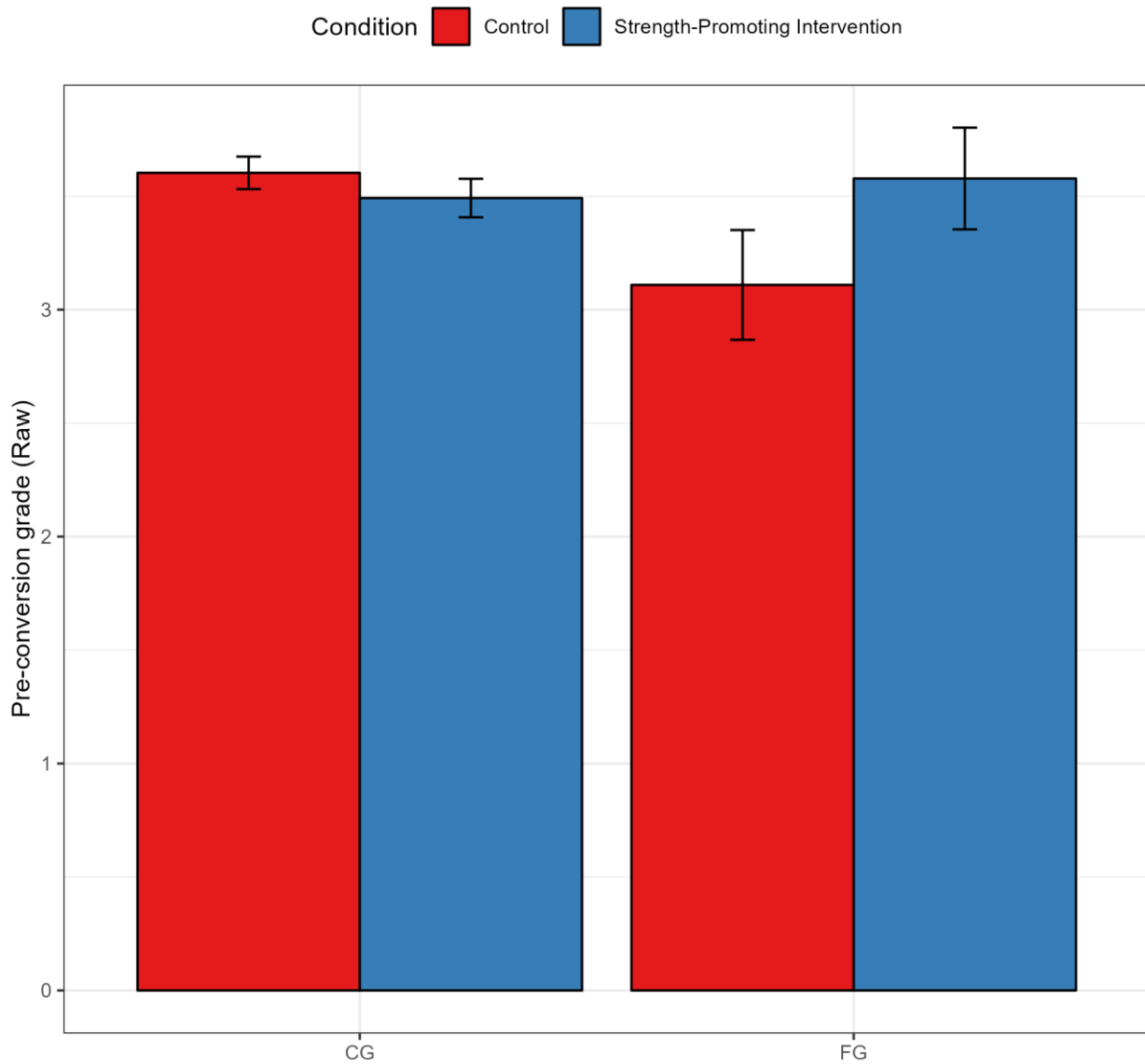
Follow-up exploratory analyses revealed additional positive effects of the interventions on the grades of certain student subgroups. Among participants in the upper-level calculus courses, the interventions reduced the gap between URM and non-URM students.

Among first-generation students, the strength-promoting intervention increased grades. This positive effect accrued primarily in classes in which instructors were perceived to have higher levels of growth mindset. This achievement effect of the strength-promoting intervention is notable, because thus far (to my knowledge) prior studies of background-specific strength-promoting manipulations have not been found to have a robust effect on grades (Bauer et al., 2021; Hernandez et al., 2021; Silverman et al., 2022).

Despite these effects, many of my hypotheses were not supported. I did not find evidence for either intervention having main effects on any outcome variable, nor did I find evidence for the hypothesized intervention x URM status interaction on any outcome (with the exception of stress). Because there were no such effects on the hypothesized outcome variables, I did not report tests of intervention effects on the hypothesized mediators, although most of these mediators were not impacted by either intervention.

Figure 7

Pre-Conversion Course Grade by Condition and Generation Status



A notable exception, however, is belonging. As described in the Results, there was a negative main effect of the strength-promoting intervention on belonging such that participants in the intervention condition reported a mean level of belonging that was 0.19 *SD* below that of participants in the control condition.

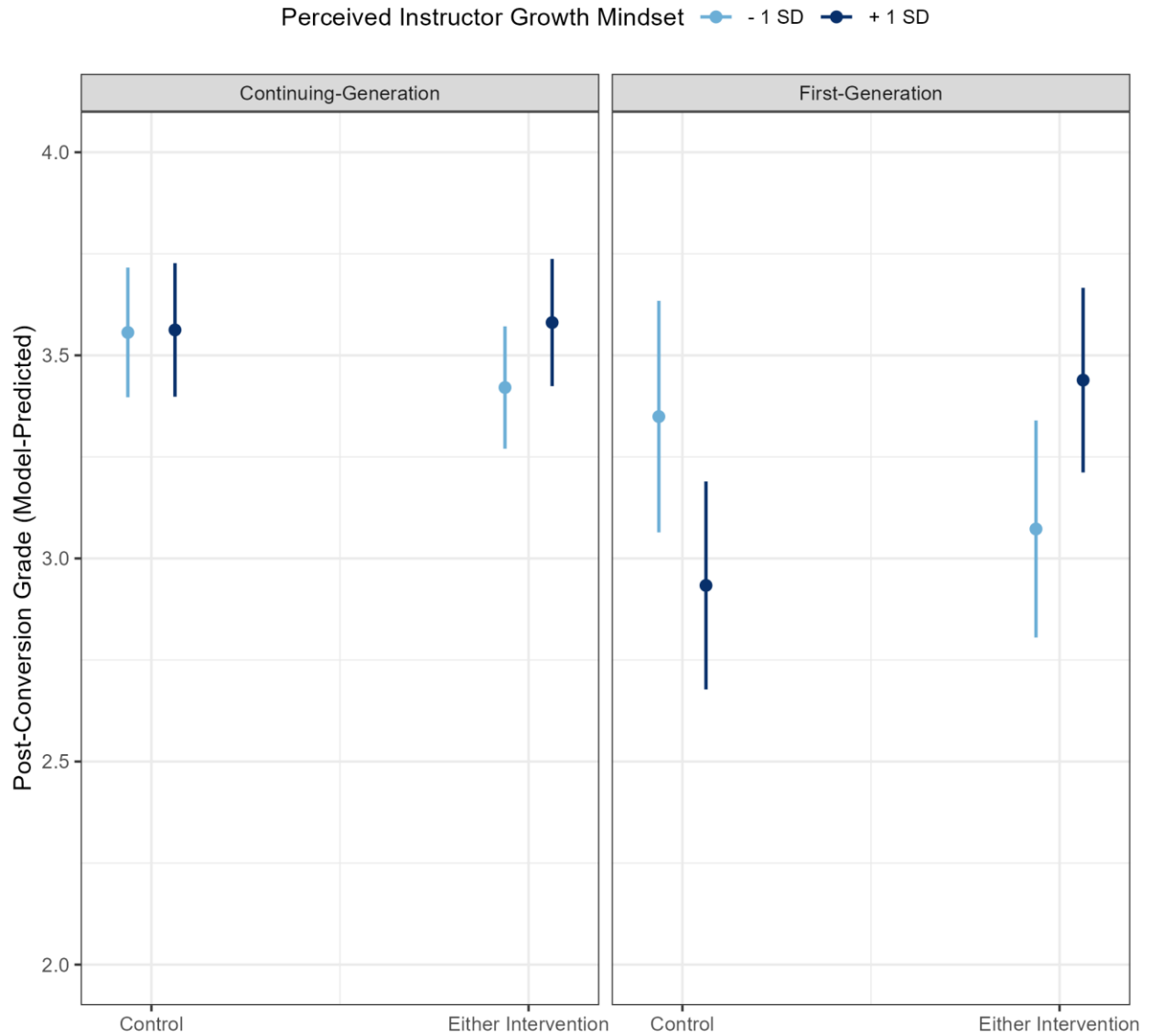
Table 2
Summary of Primary Moderation Effects

Condition	Moderator(s)	Outcome	Effect Estimate
Student-Level			
Both	URM	Mentorship	0.67 [0.32, 1.03]
Both	URM	Stress	-0.32 [-0.66, 0.02]
Both	URM x Class Level	Pre-Conversion Grades	-0.92 [-1.71, -0.11]
Strength-Promoting	FG	Pre-Conversion Grades	0.63 [0.28, 0.98]
Section-Level			
Both	Perceived Instructor Growth Mindset	Post-Conversion Grades	0.50 [0.11, 0.89]
Both	Perceived Instructor Growth Mindset x Student-Level FG	Post-Conversion Grades	1.17 [0.24, 2.10]

Note. Both = both the threat-reducing and strength-promoting intervention effect estimates were of a similar magnitude and in the same direction, so they were modeled as a single intervention contrast. URM = underrepresented and racially-ethnically minoritized. FG = first-generation. Effect estimates are standardized beta coefficients. Brackets are 95% confidence intervals of the coefficients. All models are multilevel models (class section is Level 2). Student-Level models include the following covariates: Race (0 = non-URM; 1 = URM), generation status (0 = continuing-generation; 1 = first-generation), gender (0 = male; 1 = female), and subject (except for URM x Class Level; 0 = math; 1 = engineering). Section-Level models did not include covariates because they did not converge when including the covariates.

Figure 8

Post-Conversion Course Grade by Condition, Section-Level Perceptions of Instructor Growth Mindset, and Generation Status



Why might the strength-promoting intervention have had a negative effect on belonging?

It is possible that prompting students to reflect on their individual backgrounds—especially those that may be otherwise stigmatized—led them to think of themselves as more different than those around them, even if such reflection is framed positively (i.e., in a strengths format). Providing

some support for this possibility, I found that the strength-promoting intervention also had a negative effect on outgroup closeness, such that participants in the intervention condition reported a mean level of outgroup closeness that was 0.17 *SD* below that of participants in the control condition.

Another unexpected finding was the seemingly negative effect of the threat-reducing intervention on grades in the lower-level calculus courses. There are two reasons that this effect should probably not be interpreted as indicating harm from the threat-reducing intervention. The first reason is that the significant two-way interaction (intervention x class level) seems to be driven by a boost in grades for the control condition participants in the lower-level classes rather than a drop in grades for the intervention condition participants in the lower-level courses (see Figure 3). As noted in the Results section, the control participants in the lower-level courses earned a mean pre-conversion grade of 3.36 (*SD* = 0.79), and the threat-reducing intervention participants earned a mean pre-conversion grade of 3.19 (*SD* = 0.89). In the academic year prior to the year in which the present study was conducted, I tested a similar threat-reducing intervention (against a similar study skills control condition) in the same mathematics courses as those included in the present study. In the prior data set, the mean pre-conversion grade of control participants in the lower-level courses was 2.89 (*SD* = 0.88), whereas that of threat-reducing intervention participants in those courses was 3.26 (*SD* = 0.73). Put more clearly, the average grade earned by lower-level participants in the present study's control condition ($M = 3.36$) was notably higher than that of lower-level participants in the prior study's control condition ($M = 2.89$). On the other hand, the average grade earned by lower-level participants in the present study's threat-reducing intervention condition ($M = 3.19$) was nearly identical to that of lower-level participants in the prior study's intervention condition ($M = 3.26$).

The second reason is that the significant two-way interaction of (joint) intervention x class level was qualified by a significant three-way interaction of intervention x class level x URM status, as noted in the Results section. As seen in Figure 4, this effect suggests that being assigned to either intervention actually had a positive effect on grades for URM participants in upper-level courses. That is, the achievement gap between URM and non-URM participants in the upper-level classes was reduced by 59%. It should be noted that this effect is in tension with a prior study of mine (Williams et al., 2023), in which I found that a threat-reducing intervention improved grades among students in lower-level (but not upper-level) calculus courses relative to a study skills control condition. The effect was not moderated by URM status. Despite testing similar interventions against similar control conditions, there were a number of differences between the prior study and present study, including intervention format (group discussion-based vs entirely individual) and school context (virtual vs in-person learning). Future research should conduct confirmatory tests to establish the validity of these effects (given the exploratory nature of the analyses) and, if reproduced, explore the contextual features that may explain the divergent results.

Turning now to the contextual heterogeneity analyses, most of my hypotheses about the section-level factors that might explain intervention effect heterogeneity were not supported. However, I did find evidence that students' perceptions of their instructor's growth mindset—measured at the section level—moderated intervention effectiveness. In other words, in class sections in which the instructor was rated, on average, as having more of a growth mindset, the intervention had a stronger positive effect on grades. This effect is interesting for at least four reasons. First, the direction of this effect is in the opposite direction than what I hypothesized and has been found in prior literature with a similar threat-reducing belonging intervention

(Hammarlund et al., 2022). For threat-reducing interventions such as the social-belonging intervention and values-affirmation intervention, effectiveness is expected (and has been found; e.g., Hanselman et al., 2014; Walton et al., 2023) to be strongest in contexts that present both psychological threat and psychological opportunity. That is, these types of interventions ought to be the most effective in settings in which (at least some) individuals hold maladaptive mindsets (due to contextual influences, such as pejorative stereotypes about academic ability) yet simultaneously allow for more adaptive mindsets to be constructed (Walton & Yeager, 2020). Given this logic, it is possible that (perceived) instructor mindset was operating differently in the present study than it was in the Hammarlund et al. (2022) study. Perhaps, in this study, section-level perceptions of instructor growth mindset were an indicator of psychological opportunity (in which the adaptive messages from the interventions could flourish; see Yeager et al., 2022), whereas in the study conducted by Hammarlund and colleagues (2022), instructor mindset was an indicator of psychological threat that permeated students' own mindsets (and then was assuaged by their intervention). This possibility is further supported by the fact that measurement of instructor mindset differed between the present study and the Hammarlund et al. (2022) study: I measured students' perceptions of their instructor's mindset, whereas Hammarlund and colleagues directly measured instructors' own mindsets in relation to their students.

Second, students' assessments of their instructor's growth mindset were made in either the third (mathematics courses) or fourth (engineering courses) week of classes. This suggests that students were able to detect meaningful information about their instructor's psychology in fewer than ten class meetings. This information may be communicated by the instructor in a number of (potentially subtle) ways, e.g., by offering plentiful opportunities for feedback (Kroeper et al., 2022). By the same logic, it is possible that the other hypothesized section-level

moderators (perceptions of instructor's attitudes toward diversity; perceptions of instructor's endorsement of background-specific strengths (BSS); and perceptions of peers' warmth and competence) of the interventions were not found to moderate because such perceptions could not be accurately developed within such a short time frame. Providing some support of this possibility, the perceptions of instructor growth mindset variable exhibits relatively less within-section variance (and more between-section variance) than the perceptions of instructor's attitudes toward diversity variable and perceptions of peers' warmth and competence variable. More specifically, the intraclass correlation coefficients indicate that 7% of the perceived instructor mindset variance is between-section, whereas 4% of the perceived instructor attitudes toward diversity and 2% of the perceptions of peers' warmth and competence are between-section. The pattern does not, however, hold for perceptions of instructor's endorsement of BSS, which has 8% of its variance between-section.

Third, there was no significant moderating effect of the perceptions of instructor growth mindset variable when it was modeled at the student-level (as opposed to the section-level). This suggests that, as hypothesized, the measure of students' perceptions of their instructor's growth mindset represents a context, rather than individual, effect. However, it may also be a statistical artifact: It is a well-known phenomenon that aggregating people's individual reports up to the group level reduces variance and improves accuracy (Surowiecki, 2004). As would be expected, the standard deviation of the student-level perceptions of instructor mindset variable ($SD = 0.94$) is more than three times the size of the standard deviation of the section-level perceptions of instructor mindset variable ($SD = 0.31$).

Fourth, the cross-level two-way interaction was further qualified by a three-way interaction among (student-level) condition, (section-level) perceptions of instructor growth

mindset, and (student-level) first-generation status. This is notable because URM students were the subgroup hypothesized to benefit the most from the interventions tested in this study. In the next section, I will consider potential reasons for the moderating effect of generation status, as opposed to URM status, on the intervention's effects on grades.

Effects of the Interventions for First-Generation Students

An important question to consider is why the strength-based intervention increased grades for first-generation (FG) participants, especially in contrast to its null effects on grades for URM participants. One possibility is that URM participants did not, on average, take up the key intervention messages. Whereas the strength-based intervention had a positive effect on the BSS manipulation check for FG participants, it did not have a significant effect for URM participants (although the estimate was in the expected direction, $b^* = 0.16$, $SE = 0.18$, $p = .385$). This may be due to the relatively light-touch approach my team and I took when addressing racial-ethnic identities in the strength-promoting intervention materials. For example, in the introductory text of the intervention (which described the concept of BSS), “race/ethnicity” is only explicitly mentioned twice, as a parenthetical example of the different types of backgrounds. Although “socioeconomic status” is also only specifically mentioned twice in the same text, students may have been more likely to associate the word “background” with their SES-informed upbringing (which occurred in the past) rather than their race/ethnicity (which is an ongoing feature of their identity). Moreover, among the five stories presented in the strength-promoting intervention condition, two focused solely on race/ethnicity (one of which was a non-URM identity: Asian-American), one focused solely on SES, and one mentioned both ethnicity (Hispanic/Latinx) and SES (working-class). Taken together, these materials may not have communicated to participants the intended message that race/ethnicity can be considered an identity that inherently produces

strengths. My team and I followed this light-touch approach largely due to our sample (in addition to the intervention process findings from the pilot study): Because we were recruiting all students across all of the introductory engineering and calculus courses, we had some concern that being too heavy-handed with the race/ethnicity language may produce undesired backfire effects. These effects could have been reactance or backlash from the non-URM participants and/or demoralization among the URM participants (such as what we saw in the pilot study of our original validation intervention). An important next step in analyzing this data set is to investigate the written reflections of participants to ascertain which identities they were most likely to write about after engaging with the strength-promoting intervention materials.

Another possibility for the divergent strength-promoting intervention effects for FG relative to URM participants is that the context may not have been sufficiently supportive of URM students to allow for the intervention's key messages to propagate forward into behavior. The higher education landscape (especially STEM fields) is entrenched in structural racism (Barber et al., 2020; McGee, 2020), which may overpower a targeted psychological-reframing intervention, such as the strength-promoting intervention. In other words, it is possible that the university context in which I ran this study did not have sufficient psychological opportunities for the intervention messages to take root in the minds of URM participants, undermining the processes necessary for a brief intervention of this sort to yield long-term positive effects (see Walton & Yeager, 2020).

Another contribution of this study is its direct investigation of two potential unintended negative consequences of social psychological interventions. As described above, I did not find evidence for any main or moderated (intervention x URM) effects on individual attributions for inequality (threat-reducing intervention) or race essentialism (strength-promoting intervention).

Although these results do not provide definitive proof that these types of interventions do not yield any unintended negative consequences, they do offer the first direct test (to my knowledge) of two outcomes that have been proposed as theoretically likely negative effects (Covarrubias & Laiduc, 2022). That said, because I did not observe widespread effects of the two interventions on the hypothesized outcomes, it is possible that the unintended negative effects were not observed due to the interventions' relatively low levels of impact. In partial support of this possibility, the strength-promoting intervention—which had a positive effect on grades for FG participants—exhibited a marginally-significant interaction with FG status on race essentialism, $b^* = 0.43$, $SE = 0.23$, $p = .064$. FG participants in the strength-promoting condition ($M = 4.66$, $SD = 1.25$) reported a higher average level of race essentialism than FG participants in the control condition ($M = 4.30$, $SD = 0.79$; $d = 0.37$). Because it is marginally significant and exploratory, this effect should not be overinterpreted. It does, however, suggest that when social psychological interventions such as the strength-promoting intervention improve “objective” indicators of academic achievement, they may also have undesirable effects on other aspects of individuals' belief systems. Future research should investigate this possibility more thoroughly.

Limitations

The results (significant and null) of this study must be considered in light of some limitations. First, two highly disruptive and racialized incidents occurred at the university during the semester in which this study was conducted. The first incident was a hate crime, which students were emailed about the week of intervention implementation. The second incident was a fatal act of violence, which occurred shortly before this study's Time 2 survey was administered. Because the incidents occurred after some (the former incident) or all (the latter incident) participants had been randomized into condition, it is possible that they interacted with our

intervention materials differentially across conditions. Given our data, we cannot directly test this possibility. Rather, caution must be taken when interpreting and considering the generalizability of the results.

Second, although the study was powered to detect main effect sizes around 0.20 *SD*, some of the exploratory analyses that I conducted resulted in relatively small cell sizes and as such should be considered underpowered. For example, the three-way interaction among (joint) intervention, class level, and URM status resulted in one cell of only 22 participants (URM participants in upper-level calculus classes who were assigned to the control condition). This and similar effects should be interpreted cautiously due to their exploratory nature and relative lack of power.

Third, this study is unable to fully explain the differential effects of the tested interventions in relation to other interventions with similar theoretical bases. For example, I have found that a threat-reducing belonging intervention improved grades for students in lower-level calculus courses in a prior study (Williams et al., 2023), whereas the threat-reducing intervention tested in this study had no effects in the same lower-level calculus courses (and potentially some positive effects among URM students in the upper-level courses). Despite some similarities between these studies, there are also many differences in context that likely explain these different patterns of effects. As of yet, however, the specific ways in which this occurs is still unknown. Although it is promising that social psychological interventions such as those designed and tested in this study have been found to be effective in a wide array of contexts (Walton & Wilson, 2018), there is still much to understand about the role of those contexts in moderating the interventions' (non) effects (Bryan et al., 2021).

Conclusion

This study provided the first test of a context-specific, racial-ethnic identity-focused strength-promoting intervention on a range of well-being and achievement outcomes among students in higher level STEM contexts, including potential unintended negative consequences. It also directly compared the impact of the strength-promoting intervention to that of a more traditional threat-reducing intervention. For both interventions, I made and directly tested novel hypotheses about the mechanisms that may explain intervention effects. Finally, this study was designed to test intervention effect heterogeneity across individual STEM class sections. To investigate this heterogeneity, I made and directly tested novel hypotheses about which features of the sections may moderate the interventions' effectiveness.

Confirmatory analyses indicated that the threat-reducing intervention decreased URM participants' stress near the end of the semester, and both the threat-reducing and strength-promoting intervention boosted URM participants' reports of mentorship near the end of the semester. Moreover, the interventions were found to vary in their effect on grades across class sections, partly as a function of section-level student perceptions of their instructor's growth mindset. More specifically, in courses that had instructors who were rated as having more of a growth mindset, the joint intervention effect on grades was more positive than in courses that had instructors who were rated as having less of a growth mindset.

Exploratory analyses provided some evidence that the interventions improved the grades of URM students in upper-level calculus courses and FG students in classes in which instructors were perceived to have higher levels of growth mindset.

Taken together, these results suggest that brief strength-promoting and threat-reducing interventions delivered directly to students have the potential to improve certain well-being and achievement outcomes among students from historically underrepresented groups in higher

education. They may be more effective in contexts in which students perceive their instructor to endorse the belief that intelligence is malleable and all students are able to perform well academically. More research is needed to understand how strength-based interventions can be best tailored to the experiences of underrepresented racial-ethnic groups.

Considered in light of the wider literature, there is growing evidence that both approaches to intervention can be effective across a diverse array of contexts (Walton & Wilson, 2018; Destin, 2019a). And while the field is beginning to consider and investigate effect heterogeneity more directly (Bryan et al., 2021; Walton et al., 2023; Walton & Yeager, 2020), there is still much more research necessary to determine which contexts are the most (and least) appropriate for each approach. In addition to contextual heterogeneity, future research should also continue to investigate the mechanisms of threat-reducing and strength-promoting interventions in order to better understand how they exhibit effects when such effects are observed. By considering the role of strength-promotion and threat-reduction strategies in tandem with the contextual determinants and mechanisms of their effectiveness, the field of social psychology will be better equipped to help redesign educational settings to be more equitable for students who have been historically underrepresented and underserved in those settings.

References

- Barber, P. H., Hayes, T. B., Johnson, T. L., Márquez-Magaña, L., & 10,234 signatories. (2020). Systemic racism in higher education. *Science*, *369*(6510), 1440–1441. <https://doi.org/10.1126/science.abd7140>
- Bates, D. M., Mäechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <http://doi.org/10.18637/jss.v067.i01>
- Binning, K. R., Kaufmann, N., McGreevy, E. M., Fotuhi, O., Chen, S., Marshman, E., Kalendar, Z. Y., Limeri, L., Betancur, L., & Singh, C. (2020). Changing social contexts to foster equity in college science courses: An ecological-belonging intervention. *Psychological Science*, *31*(9), 1059–1070. <https://doi.org/10.1177/0956797620929984>
- Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review*, *19*(5), 547–556. <https://doi.org/10.1177/0193841X9501900504>
- Borman, G. D., Grigg, J., Rozek, C. S., Hanselman, P., & Dewey, N. A. (2018). Self-affirmation effects are produced by school context, student engagement with the intervention, and time: Lessons from a district-wide implementation. *Psychological Science*, *29*(11), 1773–1784. <https://doi.org/10.1177/0956797618784016>
- Brannon, T. N., & Lin, A. (2021). “Pride and prejudice” pathways to belonging: Implications for inclusive diversity practices within mainstream institutions. *American Psychologist*, *76*(3), 488–501. <https://doi.org/10.1037/amp0000643>
- Brody, G. H., Yu, T., Chen, E., Miller, G. E., Kogan, S. M., & Beach, S. R. (2013). Is resilience

- only skin deep? Rural African Americans' socioeconomic status–related risk and competence in preadolescence and psychological adjustment and allostatic load at age 19. *Psychological Science*, 24(7), 1285–1293. <https://doi.org/10.1177/0956797612471954>
- Browman, A. S., & Destin, M. (2016). The effects of a warm or chilly climate toward socioeconomic diversity on academic motivation and self-concept. *Personality and Social Psychology Bulletin*, 42(2), 172–187. <https://doi.org/10.1177/0146167215619379>
- Bryk, A. S., Gomez, L. M, Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.
- Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, 65, 333–371. <https://doi.org/10.1146/annurev-psych-010213-115137>
- Covarrubias, R., & Laiduc, G. (2022). Complicating college-transition stories: Strengths and challenges of approaches to diversity in wise-story interventions. *Perspectives on Psychological Science*, 17(3), 732–751. <https://doi.org/10.1177/17456916211006068>
- Craig, M. A., & Richeson, J. A. (2014). On the precipice of a “majority-minority” America: Perceived status threat from the racial demographic shift affects White Americans' political ideology. *Psychological Science*, 25(6), 1189–1197. <https://doi.org/10.1177/0956797614527113>
- del Carmen Salazar, M. (2013). A humanizing pedagogy: Reinventing the principles and practice of education as a journey toward liberation. *Review of Research in Education*, 37(1), 121–148. <https://doi.org/10.3102/0091732X12464032>
- Destin, M., & Debrosse, R. (2017). Upward social mobility and identity. *Current Opinion in Psychology*, 18, 99–104. <https://doi.org/10.1016/j.copsyc.2017.08.006>

- Destin, M., Silverman, D. M., & Rogers, L. O. (2022). Expanding the social psychological study of educators through humanizing principles. *Social and Personality Psychology Compass*, *16*(6), e12668. <https://doi.org/10.1111/spc3.12668>
- Destin, M. (2019a). A path to advance research on identity and socioeconomic opportunity. *American Psychologist*, *74*(9), 1071–1079. <https://doi.org/10.1037/amp0000514>
- Destin, M. (2019b). Socioeconomic mobility, identity, and health: Experiences that influence immunology and implications for intervention. *American Psychologist*, *74*(2), 207–217. <https://doi.org/10.1037/amp0000297>
- DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*(3), 189–228. <https://doi.org/10.1214/ss/1032280214>
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*(2), 143–156. <https://doi.org/10.1007/s11205-009-9493-y>
- Dover, T. L., Major, B., & Kaiser, C. R. (2014). Diversity initiatives, status, and system justifying beliefs: When and how diversity efforts de-legitimize discrimination claims. *Group Processes & Intergroup Relations*, *17*(4), 485–493. <https://doi.org/10.1177/1368430213502560>
- Dover, T. L., Major, B., & Kaiser, C. R. (2016). Members of high-status groups are threatened by pro-diversity organizational messages. *Journal of Experimental Social Psychology*, *62*, 58–67. <https://doi.org/10.1016/j.jesp.2015.10.006>
- Ellis, B. J., Bianchi, J., Griskevicius, V., & Frankenhuis, W. E. (2017). Beyond risk and

- protective factors: An adaptation-based approach to resilience. *Perspectives on Psychological Science*, 12(4), 561–587. <https://doi.org/10.1177/1745691617693054>
- Engle, J. & Tinto, V. (2008). Moving beyond access: College success for low-income, first-generation students. The Pell Institute. <https://files.eric.ed.gov/fulltext/ED504448.pdf>
- Evans, G. W., & Kim, P. (2012). Childhood poverty and young adults' allostatic load: The mediating role of childhood cumulative risk exposure. *Psychological Science*, 23(9), 979–983. <https://doi.org/10.1177/0956797612441218>
- Fayer, S., Lacey, A., & Watson, A. (2017). *STEM occupations: Past, present, and future*. U.S. Bureau of Labor Statistics. <https://stats.bls.gov/spotlight/2017/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future/pdf/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future.pdf>
- Garcia, J., & Cohen, G. L. (2013). A social psychological approach to educational intervention. In E. Shafir (Ed.), *Behavioral foundations of policy* (pp. 329–350). Princeton University Press.
- Giffi, C., Wellener, P., Dollar, B., Ashton, H., Monck, L., & Moutray, C. (2018). *The jobs are here, but where are the people? Key findings from the 2018 Deloitte and The Manufacturing Institute skills gap and future of work study*. Deloitte. <https://www2.deloitte.com/us/en/insights/industry/manufacturing/manufacturing-skills-gap-study.html>
- Gopalan, M., & Brady, S. T. (2020). College students' sense of belonging: A national perspective. *Educational Researcher*, 49(2), 134–137. <https://doi.org/10.3102/0013189X19897622>
- Hammarlund, S. P., Scott, C., Binning, K. R., & Cotner, S. (2022). Context matters: How an

- ecological-belonging intervention can reduce inequities in STEM. *BioScience*, 72(4), 387–396. <https://doi.org/10.1093/biosci/biab146>
- Hanselman, P., Bruch, S. K., Gamoran, A., & Borman, G. D. (2014). Threat in context: School moderation of the impact of social identity threat on racial/ethnic achievement gaps. *Sociology of Education*, 87(2), 106–124. <https://doi.org/10.1177/0038040714525970>
- 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. <http://dx.doi.org/10.1146/annurev-psych122216-011725>
- Hernandez, I. A., Silverman, D. M., & Destin, M. (2021). From deficit to benefit: Highlighting lower-SES students' background-specific strengths reinforces their academic persistence. *Journal of Experimental Social Psychology*, 92, 104080. <https://doi.org/10.1016/j.jesp.2020.104080>
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.
- Kaiser, C. R., & Miller, C. T. (2001). Stop complaining! The social costs of making attributions to discrimination. *Personality and Social Psychology Bulletin*, 27(2), 254–263. <https://doi.org/10.1177/0146167201272010>
- Kaiser, C. R., & Miller, C. T. (2003). Derogating the victim: The interpersonal consequences of blaming events on discrimination. *Group Processes & Intergroup Relations*, 6(3), 227–237. <https://doi.org/10.1177/13684302030063001>
- Kaiser, C. R., Dyrenforth, P. S., & Hagiwara, N. (2006). Why are attributions to discrimination

- interpersonally costly? A test of system-and group-justifying motivations. *Personality and Social Psychology Bulletin*, 32(11), 1423–1536.
<https://doi.org/10.1177/0146167206291475>
- Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R. (2013). Presumed fair: Ironic effects of organizational diversity structures. *Journal of Personality and Social Psychology*, 104(3), 504–519. <https://doi.org/10.1037/a0030838>
- Kaiser, C. R. (2006). Dominant ideology threat and the interpersonal consequences of attributions to discrimination. In S. Levin & C. Van Laar (Eds.) *Stigma and group inequality* (pp. 59–78). Psychology Press. <https://doi.org/10.4324/9781410617057>
- Kreft, I. G. G. (1996). *Are multilevel techniques necessary? An overview, including simulation studies*. [Unpublished manuscript]. California State University, Los Angeles.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613.
<https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13) 1–26.
<http://doi.org/10.18637/jss.v082.i13>
- Langdon, D., McKittrick, G., Beede, D., Khan, B., & Doms, M. (2011). *STEM: Good jobs now and for the future*. U.S. Department of Commerce, Economics and Statistics Administration. https://permanent.fdlp.gov/gpo93914/stemfinaljuly14_1.pdf
- Layous, K., Davis, E. M., Garcia, J., Purdie-Vaughns, V., Cook, J. E., & Cohen, G. L. (2017).

- Feeling left out, but affirmed: Protecting against the negative effects of low belonging in college. *Journal of Experimental Social Psychology*, *69*, 227–231.
<https://doi.org/10.1016/J.JESP.2016.09.008>
- Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. *Review of Educational Research*, *86*(2), 602–640.
<https://doi.org/10.3102/0034654315617832>
- Levin, S., Sidanius, J., Rabinowitz, J. L., & Federico, C. (1998). Ethnic identity, legitimizing ideologies, and social status: A matter of ideological asymmetry. *Political Psychology*, *19*(2), 373–404. <https://doi.org/10.1111/0162-895X.00109>
- Lipson, S. K., Zhou, S., Wagner, B., Beck, K., & Eisenberg, D. (2016). Major differences: variations in undergraduate and graduate student mental health and treatment utilization across academic disciplines. *Journal of College Student Psychotherapy*, *30*(1), 23–41.
<https://doi.org/10.1080/87568225.2016.1105657>
- McGee, E. O., Griffith, D. M., & Houston, S. L. (2019). “I know I have to work twice as hard and hope that makes me good enough”: Exploring the stress and strain of Black doctoral students in engineering and computing. *Teachers College Record*, *121*(4), 1–38.
<https://doi.org/10.1177/016146811912100407>
- McGee, E. O. (2020). Interrogating structural racism in STEM higher education. *Educational Researcher*, *49*(9), <https://doi.org/10.3102/0013189X20972718>
- Morrison, K. R., Plaut, V. C., & Ybarra, O. (2010). Predicting whether multiculturalism positively or negatively influences White Americans’ intergroup attitudes: The role of ethnic identification. *Personality and Social Psychology Bulletin*, *36*(12), 1648–1661.
<https://doi.org/10.1177/0146167210386118>

- Muenks, K., Canning, E. A., LaCrosse, J., Green, D. J., Zirkel, S., Garcia, J. A., & Murphy, M. C. (2020). Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset beliefs predict their psychological vulnerability, engagement, and performance in class. *Journal of Experimental Psychology: General*, *149*(11), 2119–2144. <https://doi.org/10.1037/xge0000763>
- Murphy, M. C., Gopalan, M., Carter, E. R., Emerson, K. T., Bottoms, B. L., & Walton, G. M. (2020). A customized belonging intervention improves retention of socially disadvantaged students at a broad-access university. *Science Advances*, *6*(29), eaba4677. <https://doi.org/10.1126/sciadv.aba4677>
- Museus, S. D., & Quayle, S. J. (2009). Toward an intercultural perspective of racial and ethnic minority college student persistence. *The Review of Higher Education*, *33*(1), 67–94. <https://doi.org/10.1353/rhe.0.0107>
- National Center for Education Statistics (2021). *Table 322.30. Bachelor's degrees conferred by postsecondary institutions, by race/ethnicity and field of study: 2018-19 and 2019-20*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/digest/d21/tables/dt21_322.30.asp
- National Center for Education Statistics (2022a). *Educational attainment of young adults*. Condition of Education. U.S. Department of Education, Institute of Education Sciences. <https://nces.ed.gov/programs/coe/indicator/caa>.
- National Center for Education Statistics (2022b). *Table 505.10. Number, percentage distribution, unemployment rates, and median earnings of 25- to 29-year-old bachelor's degree holders and percentage of degree holders among all 25- to 29-year-olds, by field of study and science, technology, engineering, or mathematics (STEM) status of field: 2010*

- and 2019. U.S. Department of Education, Institute of Education Sciences.
https://nces.ed.gov/programs/digest/d20/tables/dt20_505.10.asp
- Oyserman, D., Destin, M., & Novin, S. (2015). The context-sensitive future self: Possible selves motivate in context, not otherwise. *Self and Identity*, 14(2), 173–188.
<https://doi.org/10.1080/15298868.2014.965733>
- Plaut, V. C., Garnett, F. G., Buffardi, L. E., & Sanchez-Burks, J. (2011). “What about me?” Perceptions of exclusion and Whites' reactions to multiculturalism. *Journal of Personality and Social Psychology*, 101(2), 337–353. <https://doi.org/10.1037/a0022832>
- President's Council of Advisors on Science and Technology, Executive Office of the President. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*.
<https://files.eric.ed.gov/fulltext/ED541511.pdf>
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Razzouk, R., & Shute, V. (2012). What is design thinking and why is it important? *Review of Educational Research*, 82(3), 330–348. <http://doi.org/10.3102/0034654312457429>
- Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91–116). Russell Sage Foundation.
- Reardon, S. F. (2013). The widening income achievement gap. *Educational Leadership*, 70(8), 10–16.
- Linares, L. I. R., & Muñoz, S. M. (2011). Revisiting validation theory: Theoretical foundations,

- applications, and extensions. *Enrollment Management Journal*, 5(2), 12–33.
- Rendón, L. I. (1994). Validating culturally diverse students: Toward a new model of learning and student development. *Innovative Higher Education*, 19(1), 33–51.
<https://doi.org/10.1007/BF01191156>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sherman, D. K., Lokhande, M., Muller, T., & Cohen, G. L. (2021). Self-affirmation interventions. In G. M. Walton & A. J. Crum (Eds.), *Handbook of wise interventions: How social psychology can help people change* (pp. 63–99). The Guilford Press.
- Shnabel, N., Purdie-Vaughns, V., Cook, J. E., Garcia, J., & Cohen, G. L. (2013). Demystifying values-affirmation interventions: writing about social belonging is a key to buffering against identity threat. *Personality & Social Psychology Bulletin*, 39(5), 663–676.
<https://doi.org/10.1177/0146167213480816>
- Silverman, D. M., Hernandez, I. A., & Destin, M. (2022). Educators' beliefs about students' socioeconomic backgrounds as a pathway for supporting motivation. *Personality and Social Psychology Bulletin*. <https://doi.org/10.1177/01461672211061945>
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Stephens, N. M., Hamedani, M. G., & Destin, M. (2014). Closing the social-class achievement gap: A difference-education intervention improves first-generation students' academic performance and all students' college transition. *Psychological Science*, 25(4), 943–953.
<https://doi.org/10.1177/0956797613518349>

- Stephens, N. M., Townsend, S. S., Hamedani, M. G., Destin, M., & Manzo, V. (2015). A difference-education intervention equips first-generation college students to thrive in the face of stressful college situations. *Psychological Science, 26*(10), 1556–1566.
<https://doi.org/10.1177/0956797615593501>
- Walton, G. M., & Brady, S. T. (2017). The many questions of belonging. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (pp. 272–293). The Guilford Press.
- Walton, G. M., & Brady, S. T. (2020). The social-belonging intervention. In G. M. Walton & A. J. Crum (Eds.), *Handbook of wise interventions: How social psychology can help people change* (pp. 36–62). The Guilford Press.
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*(1), 82–96.
<https://doi.org/10.1037/0022-3514.92.1.82>
- Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*(6023), 1447–1451.
<https://doi.org/10.1126/science.1198364>
- Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review, 125*(5), 617–655.
<https://doi.org/10.1037/REV0000115>
- 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*(3), 219–226. <https://doi.org/10.1177/0963721420904453>
- 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*(2), 468–485.
<http://dx.doi.org/10.1037/a0037461>
- Walton, G. M., Murphy, M. C., Logel, C., Yeager, D. S., & The College Transition Collaborative (2017). *The social-belonging intervention: A guide for use and customization*.
- Wilkins, C. L., & Kaiser, C. R. (2014). Racial progress as threat to the status hierarchy: Implications for perceptions of anti-White bias. *Psychological Science*, *25*(2), 439–446.
<https://doi.org/10.1177/0956797613508412>
- Wilkins, C. L., Wellman, J. D., & Kaiser, C. R. (2013). Status legitimizing beliefs predict positivity toward Whites who claim anti-White bias. *Journal of Experimental Social Psychology*, *49*(6), 1114–1119. <https://doi.org/10.1016/j.jesp.2013.05.017>
- Williams, M. J., & Eberhardt, J. L. (2008). Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology*, *94*(6), 1033–1047.
<https://doi.org/10.1037/0022-3514.94.6.1033>
- Williams, C. L., Wilson, T. D., Henderson, A., & Walton, G. M. (2023). *An online, discussion based belonging intervention improves outcomes for students in introductory calculus courses*. [Manuscript in preparation]. Department of Psychology, University of Virginia.
- Wilson, T. D. (2006). The power of social psychological interventions. *Science*, *313*(5791), 1251–1252. <http://dx.doi.org/10.1126/science.113301>
- Wilson, T. D. (2011). *Redirect: Changing the stories we live by*. New York, NY: Little, Brown and Company.
- Wilton, L. S., Apfelbaum, E. P., & Good, J. J. (2019). Valuing differences and reinforcing them:

- Multiculturalism increases race essentialism. *Social Psychological and Personality Science*, *10*(5), 681–689. <https://doi.org/10.1177/1948550618780728>
- Xue, Y., & Larson, R. C. (2015). *STEM crisis or STEM surplus? Yes and yes*. U.S. Bureau of Labor Statistics. <https://doi.org/10.21916/mlr.2015.14>
- Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*, 267–301.
<http://dx.doi.org/10.3102/0034654311405999>
- Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., Lee, H. Y., O'Brien, J., Flint, K., Roberts, A., Trott, J., Greene, D., Walton, G. M., & Dweck, C. S. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, *108*(3), 374–391. <https://doi.org/10.1037/edu0000098>
- Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., Kamentz, D., Ritter, G., Duckworth, A. L., Urstein, R., Gomez, E. M., Markus, H. R., Cohen, G. L., & Dweck, C. S. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(24), E3341–E3348. <https://doi.org/10.1073/PNAS.1524360113>
- Yosso, T. J. (2016). Whose culture has capital? A critical race theory discussion of community cultural wealth. In A. D. Dixson, C. K. R. Anderson, & J. K. Donnor, *Critical race theory in education: All God's children got a song* (2nd ed., pp. 113–136). Routledge.
<https://doi.org/10.4324/9781315709796>