*Empirical Research Paper*

Personality and Social Psychology Bulletin
1–22

DOI: 10.1177/01461672241265954
journals.sagepub.com/home/pspb
Sage

# Positive Feedback as a Lever to Boost Students' STEM Outcomes

**Lora E. Park[1], Deborah E. Ward[2], Deborah Moore-Russo[3], Brian Rickard[4], Valerie Vessels[1], and Joseph Hundley[1]**

**Abstract**

Although many college students intend to major in Science, Technology, Engineering, and Mathematics (STEM), dropout from these fields is high, especially among members of historically underrepresented groups, such as women and racial-ethnic minorities. We propose a minimal, yet potentially powerful intervention to broaden participation in STEM: giving positive feedback to students in STEM. Studies 1 and 2 found that giving positive feedback is less normative in math (vs. English) courses, and instructors' feedback-giving practices and students' experiences mirror these norms. However, students who received positive (vs. only objective) feedback on introductory-level college calculus exams showed greater belonging and self-efficacy in math, which predicted better STEM outcomes (i.e., increased interest in STEM and higher final math course grades, respectively, Study 3). These findings were especially strong for racial-ethnic minority students. Giving positive (vs. only objective) feedback is thus a potentially transformative tool that boosts student outcomes, especially for underrepresented groups.

**Keywords**

feedback, STEM, belonging, self-efficacy, performance, motivation

Received February 28, 2024; revision accepted June 16, 2024

Although many college students begin as Science, Technology, Engineering, and Mathematics (STEM) majors, their experiences in gateway STEM courses pose a major barrier to their continued motivation and interest in these fields (Ellis et al., 2016). In particular, instructors play a key role in shaping students' initial STEM experiences. As Seymour and Hewitt (1997) noted, ". . . the educational experience and the culture of the discipline . . . make a much greater contribution to STEM attrition than the individual inadequacies of students or the appeal of other majors" (p. 392). Indeed, former STEM majors report losing interest and dropping out of these fields due to a "chilly climate" characterized by intimidating classroom environments, poor quality of instruction, and impersonal teaching style (Seymour & Hewitt, 1997; Vogt, 2008).

According to the cues hypothesis, people look to features of their environment to determine who or what is valued and accepted in that context (Murphy et al., 2007). Since college is a time of transition, students often feel uncertain of their academic abilities and belonging in campus communities. Due to their marginalized group membership, underrepresented minorities (URMs), such as women and racial-ethnic minorities, may experience heightened concerns about their academic aptitude and belonging. Thus, in STEM contexts, students—especially URMs—may be highly attuned to cues that convey whether or not they belong and can succeed in that setting.

In the present research, we examine *positive feedback* as a minimal but potentially powerful cue to improve student outcomes, especially for URMs in STEM fields. Indeed, college students report higher self-efficacy, effort, and critical thinking in their engineering courses when instructors are supportive, approachable, and accessible (Vogt, 2008). Thus, giving positive feedback to students in STEM may be an underutilized method to boost students' math self-efficacy, belonging, and intended study habits in these courses, with consequences for their interest and performance in STEM settings.

We first investigated norms and practices related to feedback provision in college math and English courses from the perspective of instructors (Study 1) and students (Study 2). We then conducted a large-scale intervention study in which

[1]University at Buffalo, The State University of New York, NY, USA
[2]Saint Joseph's University, Philadelphia, PA, USA
[3]The University of Oklahoma, Norman, USA
[4]University of Arkansas, Little Rock, USA

**Corresponding Author:**
Lora E. Park, University at Buffalo, 344 Park Hall, Buffalo, NY 14260, USA.
Email: lorapark@buffalo.edu

**Correction (August 2024):** Article updated to correct data errors in Tables 4 and 5.

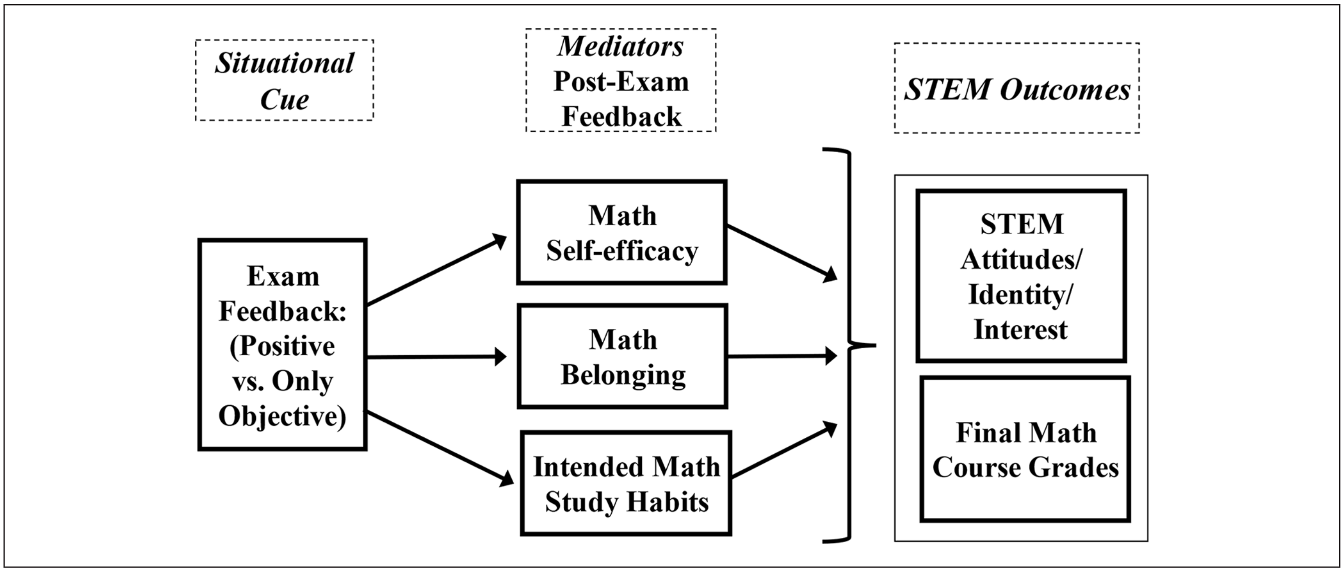


**Figure 1.** Proposed Model of Effects of Positive Versus Only Objective Feedback on Students' STEM-Related Outcomes.

students across two universities received either positive feedback (conveying a job well done) or only objective feedback (their score only) on exams in their introductory calculus course. We predicted that positive feedback would increase students' self-efficacy, belonging, and/or intended study habits in their math course, which would boost their attitudes/identity/interest in STEM and final course grades. If positive feedback improves students' outcomes, then this brief intervention could have far-reaching effects by broadening participation in STEM.

## Effects of Positive Feedback

As perceived gatekeepers, STEM instructors can impact students through the feedback they give. Feedback refers to "information provided by an agent . . . regarding aspects of one's performance" (Hattie & Timperley, 2007, p. 81). In the current research, we propose that positive feedback is a malleable situational cue that may boost students' math self-efficacy, belonging, intended study habits and in turn, their STEM interest and performance (Figure 1).

Positive feedback (e.g., comments conveying a job well done) is thought to motivate goal pursuit by increasing expectations to achieve desired outcomes (Bandura, 1982). Given that introductory college STEM courses often convey a competitive, intimidating, "weed-out" mentality (Mervis, 2011), receiving positive feedback may be especially beneficial for boosting students' self-efficacy, belonging, and intentions to adopt proactive study habits in their STEM courses.

The literature, however, reveals mixed effects of positive feedback (Butler, 1987, 1988; see Hattie & Timperley, 2007; Henderlong & Lepper, 2002; Kluger & DeNisi, 1996, for reviews). Some studies reveal negative consequences of positive feedback, such as increased pressure to achieve external standards of worth, heightened self-consciousness, ego-involvement, and reduced motivation and performance in the face of difficulty (Baumeister et al., 1990; Butler, 1987, 1988; Kamins & Dweck, 1999). For example, children who were verbally praised for their performance on a task performed worse on a subsequent task following failure (Dweck, 1975). Children who were praised for their intelligence (vs. effort) on a lab task also adopted more performance goals and fixed mindsets about intelligence and showed decreased task enjoyment, performance, and persistence following failure (Mueller & Dweck, 1998). Thus, positive feedback may be ineffective for learning due to a disproportionate focus on the person versus the task (Hattie & Timperley, 2007). In addition, some meta-analyses suggest a weak relationship between praise and student performance (Kluger & DeNisi, 1996).

Despite such findings, other studies suggest that positive feedback is beneficial. For example, positive verbal feedback increased self-reported task interest (vs. receiving no feedback, neutral feedback, or material rewards; Deci, 1971; Harackiewicz, 1979; see Cameron & Pierce, 1994; Deci et al., 1999 for reviews). Positive feedback also increased students' motivation and intended engagement, especially for URMs such as women in STEM contexts (Park et al., 2018; Park, O'Brien et al., 2023). These latter studies, however, were limited to controlled laboratory settings involving hypothetical scenarios—in which college students imagined receiving positive verbal feedback from instructors (Park, O'Brien et al., 2023)—or received positive written feedback on a standardized math test from a research assistant in the laboratory with no real-world consequences (Park et al., 2018). In addition, this past work did not test for possible racial-ethnic differences in response to receiving positive feedback.

## Impact of Positive Feedback on URMs

Owing to negative stereotypes and ongoing prejudice and discrimination, URMs may experience heightened concerns about negative evaluation based on their group membership (Roberts & Rizzo, 2021; Spencer et al., 1999; Steele & Aronson, 1995; Swim et al., 1995). Although such findings might suggest that positive feedback from authority figures could be especially beneficial for URMs, other work finds that positive feedback has differential effects based on group membership. For example, when Black students received positive interpersonal feedback from a White student and their race was visible (vs. not visible), they reported *lower* self-esteem; presumably, Black students attributed the positive feedback to their race, rather than to something personal about them (Crocker et al., 1991). However, when women received positive feedback from a male student evaluator, the visibility of their gender did not matter (Crocker et al., 1991); presumably, women were less mistrusting of positive feedback from men than Blacks were of receiving positive feedback from Whites.

Similarly, another set of studies found that Latina students who were suspicious of Whites' positive behavior toward minorities rated positive feedback from a White evaluator as being more insincere and showed increased stress, uncertainty, and lowered self-esteem after receiving this feedback (Major et al., 2016). Indeed, White evaluators tend to be more lenient and give positive feedback more often to Blacks, especially when subjective evaluations are involved (Harber, 1998).

Based on these and other findings (e.g., Park et al., 2018; Park, O'Brien et al., 2023), we expected women to benefit from receiving positive feedback in STEM contexts. However, due to mixed findings for the effects of positive feedback on racial-minorities, we did not have strong a priori predictions about whether positive feedback would be more beneficial for URMs. We did expect that, overall, students who received positive feedback would show increased math self-efficacy, belonging, and/or intentions to adopt proactive study habits, which in turn, would boost their attitudes/identity/interest in STEM and final math grades.

## Mechanisms Linking Positive Feedback With Academic Outcomes

### *Self-Efficacy*

Self-efficacy refers to a belief that one has the ability to succeed and achieve desired outcomes; indeed, those with higher self-efficacy are more motivated, perform better, and persist longer than those with lower self-efficacy (Bandura, 1982; Henderlong & Lepper, 2002). Self-efficacy in STEM is especially important during the first 2 years of college when students—especially URMs—are at risk of dropping out of introductory STEM courses such as calculus (Cromley et al., 2016). When individuals have low self-efficacy, they procrastinate, put less effort into tasks, and disengage when encountering obstacles or setbacks (Lent et al., 1994). Thus, giving positive feedback to students in college STEM courses might be one way to increase self-efficacy in math and subsequently, their interest and performance in STEM.

### *Belonging*

Belonging refers to feeling accepted as a member of a particular community (Cheryan et al., 2009; Good et al., 2012; Murphy et al., 2007). Belonging may be important for all students, but especially for URMs who doubt their academic abilities and inclusion due to experiences of racism, sexism, and other forms of bias. For example, a longitudinal study of disadvantaged students (i.e., racial-ethnic minority and first-generation college students) found that a social-belonging intervention increased social and academic fit (e.g., academic belonging, self-efficacy) at the end of the second year of college. Increased fit then predicted students' persistence through the third year of college (Murphy et al., 2020). When students lack belonging, they show poorer academic outcomes. For instance, women who felt low sense of belonging in a college calculus course performed worse and showed decreased interest in pursuing math in the future (Good et al., 2012). By contrast, women who imagined asking a question in a STEM research seminar and receiving a positive verbal response from the instructor (i.e., being told "Great question, I'm glad you brought that up!" vs. receiving a negative or neutral instructor response) reported greater belonging and self-efficacy, which increased their interest in joining the STEM lab (Park, O'Brien et al., 2023). Thus, belonging may be another key mechanism by which positive feedback boosts students' interest and performance in STEM, especially for URMs, such as women in STEM.

### *Intended Study Habits*

Study habits refer to how students manage their time, effort, and study environment to regulate their learning and make progress toward academic goals (Duncan & McKeachie, 2005; Pintrich & De Groot, 1990). We suggest that positive feedback is likely to motivate students to adopt proactive study habits to excel in their courses. Consistent with this idea, graduate students in an educational research methods course reported greater motivation to learn, spent more time doing homework, and performed better on exams when they received verbal praise from their professor (i.e., being told "Good job," "Very good," or "Great work" vs. being told "Thank you" for completing class assignments; Hancock, 2002). Based on such findings, we expected that students who received positive (vs. only objective) feedback on calculus exams would adopt better study habits (e.g., managing their study time, seeking academic help) and in turn, show better course performance and possibly, increased interest in STEM.

## Other Types of Feedback

In addition to positive feedback, other types of feedback may be relevant in STEM contexts. *Objective feedback* provides information about students' performance, but this approach may not be optimal for boosting motivation or interest. Indeed, research suggests that merely indicating whether an answer is correct or not has little impact on subsequent performance (Crooks, 1988). Another type of feedback—*corrective feedback*—involves giving specific information about what to improve and has been shown to be more effective than simply providing students with their grades (Hattie & Timperley, 2007).

Another type of feedback—*wise feedback*—involves giving critical, corrective feedback that invokes high standards and reassures students they can meet these standards (e.g., "I'm giving you this comment because I have very high expectations and I know that you can reach them" Cohen et al., 1999; Yeager et al., 2014). Past work has shown that wise feedback is more effective than unbuffered critical feedback or a combination of critical and positive feedback. For example, when Black students received wise feedback from White evaluators, they reported greater trust in their school and earned higher grades (Cohen et al., 1999; Yeager et al., 2014). Although wise feedback may be beneficial, giving such feedback may be effortful and time-consuming.

## Overview of Current Research

Students' experiences in their gateway STEM courses, such as introductory college calculus, play a crucial role in shaping their interest and success in these fields. Although giving positive feedback may not be normative in these courses, this type of feedback may be beneficial to students, especially URMs who may doubt their academic ability and belonging due to negative stereotypes or experiences of prejudice and discrimination. To investigate these ideas, we first identified the norms and practices of giving (Study 1) and receiving (Study 2) positive feedback (vs. other types of feedback) in math versus English fields. We specifically compared math and English because math and English/language arts are core subject areas in the United States educational system; language/arts is thought to be a "natural contrasting category" (Nosek et al., 2002, p. 46) with math/science; and higher education is typically organized around these distinctions (e.g., the Scholastic Aptitude Test and Graduate Records Examination—two of the most frequently used standardized tests—assess math and verbal skills, and students can earn a degree emphasizing either the arts or science—Bachelor of Arts vs. Bachelor of Science).

In Study 3, we examined real-world consequences of giving students positive feedback or only objective feedback on introductory college calculus exams. We focused on positive feedback, given previous work already establishing the benefits of wise feedback (Cohen et al., 1999; Yeager et al., 2014). We expected positive (vs. objective) feedback to increase students' self-efficacy, belonging, and intended study habits in their math course, which might positively impact course outcomes (i.e., STEM attitudes/identity/interest, final course grades). Because we did not have strong a priori hypotheses about whether self-efficacy, belonging, or intended study habits would predict the outcomes, we examined all three as parallel mediators. We additionally tested whether positive feedback would be more beneficial for members of certain groups (i.e., URMs) versus others.

## Transparency and Openness

For all studies reported in this manuscript, we report how we determined sample size, any data exclusions, manipulations, and measures. All data, analysis codes, and research materials are available at https://osf.io/jb97z/?view_only=ccf8eda06de1482ba8ebc47023359666 These studies were not preregistered. Data for all studies were analyzed with SPSS (version 28) and *R* (version 4.3).

## Study 1: Feedback Culture in Math and English Courses (Instructors' Reports)

### Participants and Procedure

College math and English instructors ($N$ = 205, 50.2% men, 46.3% women, 2.0% non-binary, 1.5% prefer not to answer, $M_{age}$ = 39.77, $SD$ = 12.41) completed an online survey about the norms and practices of giving different types of feedback in their courses (see Table 1 for demographics). No participants were excluded from analyses. For Studies 1 and 2, sample size was based on an a priori power analysis using the G*Power program (Faul et al., 2007), which indicated that approximately 200 participants would be needed to detect a moderate effect size with 80% power using a $t$-test with an alpha of .05. To recruit participants, we emailed the chairpersons of 35 mathematics and 33 English departments at 26 public research universities across the United States and asked them to forward an email containing the study description and survey link to instructors in their departments who taught introductory-level courses. The email explained that researchers were interested in how instructors give feedback in their introductory undergraduate math (or English) courses. Instructors received a $20 e-gift card for participating.

### Materials

Participants reported their demographics (e.g., age, ethnicity, gender, department affiliation, years as instructor-of-record) and listed the name of a course they recently taught in their department.[1] They reported when they last taught this course, course level (i.e., introductory, intermediate, advanced), course format (i.e., in-person, online, hybrid,

**Table 1.** Descriptive Statistics for Instructor Demographics (Study 1).

| Study | Ethnicity | Highest degree earned | Current position |
|---|---|---|---|
| **Study 1**<br>$N = 205$ | 76% White<br>11% Asian/Asian American<br>5% Mixed race<br>4% Latinx/Hispanic<br>3% Black<br>1% Middle Eastern/North African | 50% PhD<br>43% Masters degree<br>5% Bachelor's degree<br>2% Professional degree<br>(MD, JD, DDS, etc.) | 21% Teaching faculty (clinical/renewable term faculty)<br>20% Adjunct instructor<br>16% Lecturer<br>15% Other (grad student/TA)<br>8% Assistant professor (tenure-track)<br>8% Associate professor (with tenure)<br>6% Full professor (with tenure)<br>6% Post-doc |

| | Sample | Gender | Number of years taught undergraduate courses | Course level | Number of students enrolled in section(s) | Last time taught this undergraduate course | Course format |
|---|---|---|---|---|---|---|---|
| **Math Course** | 88 instructors | 66% men<br>32% women<br>1% non-binary<br>1% unknown | Median: 7–8 years<br>Mode: More than 10 years | 68% intro-level<br>29% intermediate<br>3% advanced/upper-level | Median: 50 students<br>Mode: 30 students | 74% Fall 2020<br>17% Spring 2020<br>9% Fall 2019 or other | 60% online<br>27% hybrid/other<br>12% in-person |
| **Englis Course** | 117 instructors | 35% men<br>57% women<br>3% non-binary<br>2% unknown | Median: 7–8 years<br>Mode: More than 10 years | 82% intro-level<br>14% intermediate<br>4% advanced/upper-level | Median: 24 students<br>Mode: 24 students | 72% Fall 2020<br>13% Spring 2020<br>15% Fall 2019 or other | 58% online<br>26% hybrid/other<br>16% in-person |

other), and information about course composition (e.g., number of students enrolled, gender/ethnic distribution of students, types of assignments). Participants were instructed to focus on this course when responding to the next set of questions.

Participants reported how much time, thought, effort, and care they put into the feedback they gave from 1 = *very little/not at all/not a lot* to 7 = *very much/a lot/a great deal*. A factor analysis with principal axis factoring and Promax rotation revealed that the items loaded onto a single factor with an eigenvalue of 3.02 that explained 75.63% of the variance (4 items, $\alpha = .89$).

They next reported how often they gave students different kinds of feedback: *Objective Feedback* (i.e., indicating what students got right or wrong and how many points they earned); *Positive Feedback* (i.e., comments conveying a job well done, such as "Good job!"); *Corrective Feedback* (i.e., comments on what students could improve on, such as "Need to work on X"); and *Wise Feedback* (i.e., corrective feedback plus assuring students they can reach high standards, such as "Need to work on X—I'm giving you this feedback because I have high expectations and know that you can reach them").

Instructors reported how often they gave each type of feedback on exams and assignments using the scale: 1 = *never*, 2 = *rarely*, 3 = *occasionally*, 4 = *often*, and 5 = *very often;* items were averaged together to reflect the frequency of giving objective, $r(123) = .48$, $p < .001$; positive, $r(118) = .53$, $p < .001$; corrective, $r(117) = .46$, $p < .001$; and wise feedback, $r(116) = .72$, $p < .001$.[2] Finally, participants reported the norms in their department for instructors to give objective, positive, corrective, and wise feedback from 1 = *no expectation* to 5 = *very strong expectation of giving this type of feedback.*

## Results

### Instructors' Time and Effort in Giving Feedback

Results of an independent samples *t*-test showed that math course instructors reported putting less time, thought, effort, and care into giving feedback to students ($M = 5.72$, $SD = 1.18$) than English course instructors ($M = 6.57$, $SD = .47$), $t(108.13) = -6.37$, $p < .001$, $d = .95$).

### Frequency of Giving Different Types of Feedback

We conducted a mixed-factorial ANOVA with area of study (math vs. English) as the between-subjects variable and type of feedback (objective, positive, corrective, wise feedback) as the within-subjects variable. Results showed a significant main effect of Area of Study, $F(3, 606) = 75.13$, $p < .001$, $\eta_p^2 = .27$, qualified by an Area of Study × Feedback Type interaction, $F(3, 606) = 61.75$, $p < .001$, $\eta_p^2 = .23$. Table 2 reports descriptive statistics and pairwise comparisons with Bonferroni correction to adjust for multiple comparisons; Figure 2 (left panel) shows mean differences in frequency of giving different types of feedback across courses. Consistent with hypotheses, instructors reported giving positive feedback less often to students in math versus English courses. Instructors also gave less corrective, wise, and more objective feedback in math versus English courses.

### Departmental Norms of Giving Feedback

We next conducted a mixed ANOVA with area of study as the between-subjects variable and feedback type as the within-subjects variable. Results showed a significant main effect of Area of Study, $F(3, 606) = 71.59$, $p < .001$, $\eta_p^2 = .26$, qualified by an Area of Study × Feedback Type interaction, $F(3, 606) = 66.09$, $p < .001$, $\eta_p^2 = .25$ (see Table 2 and Figure 2, right panel). Supporting hypotheses, instructors of math (vs. English) courses reported less of a departmental norm to give positive feedback to students. They also said it was less normative to give corrective or wise feedback than to give only objective feedback in math (vs. English) courses.

Instructors of math courses also reported a stronger departmental norm to give only objective feedback compared to positive, corrective, or wise feedback; a stronger norm to give corrective versus positive feedback; and less of a norm to give wise feedback versus all other feedback types. By contrast, instructors of English courses reported a stronger departmental norm to give positive (vs. only objective) feedback and less of a norm to give wise (vs. positive or corrective) feedback.

## Discussion

Study 1 showed key differences in norms and practices of giving feedback to students in math and English courses. Instructors of math courses reported putting less time and effort into giving feedback than instructors of English courses. Math instructors also said it was more normative to give only objective feedback—and less normative to give positive, corrective, or wise feedback—compared to English instructors. Notably, instructors' perceptions of norms mirrored their feedback-giving practices, giving objective feedback more often (and positive, corrective, and wise feedback less often) in math versus English courses.

## Study 2: Feedback Culture in Math and English Courses (Students' Reports)

Having documented feedback norms and practices from the perspective of college math and English instructors, Study 2 examined students' perceptions of receiving feedback in these courses. Based on Study 1, we expected students would perceive math (vs. English) instructors to put less time and effort into giving feedback and would report receiving less positive (vs. only objective) feedback in their math versus English courses. We also expected women to say they would

**Table 2.** Descriptive Statistics and Results of Pairwise Comparisons Predicting Frequency and Perceived Departmental Norm of Instructors Giving Different Types of Feedback in Math versus English Courses (Study 1).

| | Means (*SDs*) | | |
|---|---|---|---|
| Frequency of giving feedback: | Math course | English course | Pairwise comparisons |
| **Only Objective Feedback** | 4.51 (.84) | 4.17 (.96) | $p = .010$, CI [.08, .59] |
| **Positive Feedback** | 2.94 (1.03) | 4.50 (.65) | $p < .001$, CI [−1.79, −1.32] |
| **Corrective Feedback** | 3.30 (.65) | 4.40 (.70) | $p < .001$, CI [−1.36, −.86] |
| **Wise Feedback** | 2.23 (1.20) | 3.94 (1.13) | $p < .001$, CI [−2.04, −1.39] |

| | Pairwise comparisons | |
|---|---|---|
| Frequency of giving feedback: | Math course | English course |
| **Only Objective vs. Positive** | $p < .001$, CI [1.23, 1.90] | $p = .017$, CI [−.62, −.04] |
| **Only Objective vs. Corrective** | $p < .001$, CI [.88, 1.54] | $p = .186$, CI [−.53, .05] |
| **Only Objective vs. Wise** | $p < .001$, CI [1.86, 2.69] | $p = .601$, CI [−.14, .59] |
| **Positive vs. Corrective** | $p = .003$, CI [−.62, −.09] | $p = 1.00$, CI [−.14, .32] |
| **Positive vs. Wise** | $p < .001$, CI [.39, 1.05] | $p < .001$, CI [.26, .84] |
| **Corrective vs. Wise** | $p < .001$, CI [.73, 1.40] | $p < .001$, CI [.17, .75] |
| **Only Objective vs. Positive** | $p < .001$, CI [1.23, 1.90] | $p = .017$, CI [−.62, −.04] |

Departmental norm of giving different types of feedback

| | Math course | English course | Pairwise comparisons |
|---|---|---|---|
| **Only Objective Feedback** | 4.48 (1.02) | 3.91 (1.22) | $p < .001$, CI [.25, .88] |
| **Positive Feedback** | 2.59 (1.09) | 4.27 (.83) | $p < .001$, CI [−1.94, −1.41] |
| **Corrective Feedback** | 3.35 (1.11) | 4.48 (.64) | $p < .001$, CI [−1.39, −.87] |
| **Wise Feedback** | 2.13 (1.12) | 3.78 (1.23) | $p < .001$, CI [−1.98, −1.32] |

| | Pairwise comparisons | |
|---|---|---|
| Departmental norm of giving different types of feedback | Math course | English course |
| **Only Objective vs. Positive** | $p < .001$, CI [1.49, 2.28] | $p = .042$, CI [−.70, −.01] |
| **Only Objective vs. Corrective** | $p < .001$, CI [.73, 1.52] | $p < .001$, CI [−.91, −.23] |
| **Only Objective vs. Wise** | $p < .001$, CI [1.89, 2.80] | $p = 1.00$, CI [−.26, .54] |
| **Positive vs. Corrective** | $p < .001$, CI [−1.05, −.47] | $p = .150$, CI [−.47, .04] |
| **Positive vs. Wise** | $p < .001$, CI [.16, .77] | $p < .001$, CI [.22, .76] |
| **Corrective vs. Wise** | $p < .001$, CI [.89, 1.56] | $p < .001$, CI [.42, 1.00] |

*Note*. For pairwise analyses, the Bonferroni correction was used to adjust for multiple comparisons. CI = 95% confidence interval of the mean difference.

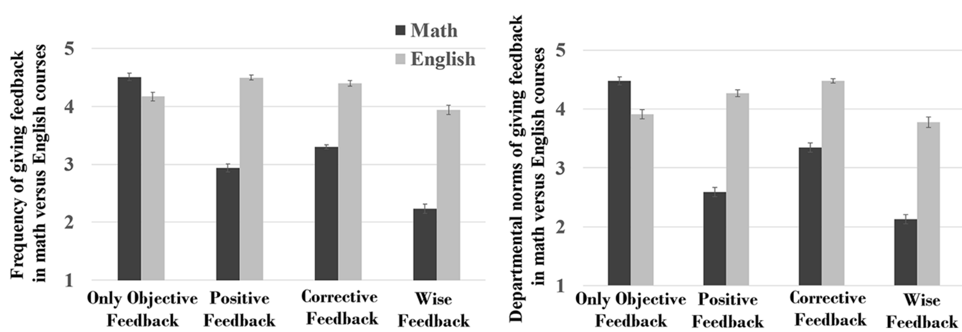


**Figure 2.** Study 1: Frequency (Left Panel) and Perceived Departmental Norms (Right Panel) of Instructors Giving Different Types of Feedback in Math Versus English Courses.
*Note.* Error bars represent the standard error around the mean of each condition.

**Table 3.** Descriptive Statistics for Undergraduate Student Demographics (Study 2).

| | Ethnicity | Year in school | Major |
|---|---|---|---|
| **Study 2** | 51% White | 20% freshmen | 92% STEM |
| *N* = 183 | 34% Asian/Asian American | 51% sophomores | 2% Arts/humanities |
| | 5% Mixed race | 17% juniors | 2% Social sciences |
| | 4% Latinx/Hispanic | 8% seniors | 4% Other |
| | 3% Black | 3% 5th year + | |
| | 2% Middle Eastern/North African | | |
| | 1% Native American/Alaskan Native | | |
| | Number of students in section | Grade earned in course | Course format |
| **Math course** | *Median* = 50 | *Median* = A– | 24% online |
| | *Mode* = 100 | *Mode* = A | 14% hybrid/other |
| | | | 61% in-person |
| **English course** | *Median* = 29 | *Median* = A– | 33% online |
| | *Mode* = 30 | *Mode* = A | 15% hybrid/other |
| | | | 52% in-person |

have benefited more from receiving positive (vs. objective) feedback from their male (vs. female) math instructor, based on past research showing that women feel greater self-efficacy and belonging when they receive positive feedback in math from a male (vs. female) authority figure (Park et al., 2018).

## *Participants and Procedure*

College students ($N$ = 183, 48.6% men, 49.2% women, 1.1% non-binary, 1.1% prefer not to answer, $M_{age}$ = 19.85, $SD$ = 2.06; see Table 3 for demographics) completed an online survey about their experiences receiving feedback in their college math and English courses. After reviewing the Mathematics Department websites of a number of public universities in the United States, we emailed the Director of Undergraduate Studies (or comparable position) and asked them to forward an email to undergraduates in their department with the study description and survey link.

Students were informed that researchers were interested in understanding how college students perceive their instructors and the feedback they typically receive in their introductory college math and English courses. Students were expected to have completed an introductory-level college calculus course and English course. Recruitment materials appear in the Methodology file. Students from 20 U.S. colleges and universities participated in the study in exchange for a $10 e-gift card.

## *Materials*

Participants first reported the name of an introductory-level math course they completed at their current institution. They then answered a series of questions related to this course. Participants reported the name of an introductory-level English course they completed at their current institution, followed by answering the same series of questions.

For each class, participants reported how much time, thought, effort, and care their instructor put into the feedback they gave from 1 = *very little/not a lot* to 7 = *very much/a lot*. A principal axis factor analysis with Promax rotation revealed that these items loaded onto a single factor with an eigenvalue of 3.20 that explained 80% of the variance (4 items, $\alpha$ = .92). For perceived effort in giving feedback in English courses, a factor analysis revealed a single factor ($\lambda$ = 3.45, 86% of variance; 4 items, $\alpha$ = .95). Using the same descriptions of feedback types as in Study 1, participants also reported how often they received objective, positive, corrective, and wise feedback on exams and assignments (using the same descriptions of these feedback types as in Study 1) using the scale 1 = *never*, 2 = *rarely*, 3 = *occasionally*, 4 = *often*, and 5 = *very often*. We averaged together participants' responses for each type of feedback to reflect the frequency of receiving objective, $r(171)_{math} = .55$, $p < .001$, $r(107)_{English} = .65$, $p < .001$; positive, $r(170)_{math} = .74$ $p < .001$, $r(107)_{English} = .70$, $p < .001$; corrective, $r(171)_{math} = .65$, $p < .001$, $r(107)_{English} = .62$, $p < .001$; and wise feedback, $r(170)_{math}=.86$, $p < .001$, $r(181)_{English}=.49$, $p < .001$.[3]

Next, participants reported how they would have felt if they had received objective, positive, corrective, and wise feedback. The items were: "Confident about my abilities in this course," "Like I belonged in this field," "Capable of doing well in this course," and "Motivated to do well in this course" from 1 = *not at all* to 5 = *very.* The four items were averaged to reflect the perceived impact of receiving objective ($\alpha_{math}$ = .90, $\alpha_{English}$ = .89), positive ($\alpha_{math}$ = .91, $\alpha_{English}$ = .88), corrective ($\alpha_{math}$ = .90, $\alpha_{English}$ = .89), and wise ($\alpha_{math}$ = .94, $\alpha_{English}$ = .88) feedback. Finally, participants reported their demographics (e.g., age, gender, ethnicity, year in school, major).

**Table 4.** Descriptive Statistics and Pairwise Comparisons Predicting Students' Overall Reports of Frequency and Anticipated Impact of Receiving Different Types of Feedback in Math and English Courses (Study 2).

| | Means (SDs) | | |
|---|---|---|---|
| Frequency of receiving: | Math course | English course | Pairwise comparisons |
| **Only Objective Feedback** | 4.04 (1.12) | 3.95 (1.24) | $p$ = 1.00, CI [−.28, .45] |
| **Positive Feedback** | 2.53 (1.27) | 3.68 (1.31) | $p$ < .001, CI [−1.58, −.72] |
| **Corrective Feedback** | 2.59 (1.26) | 3.50 (1.33) | $p$ < .001, CI [−1.34, −.49] |
| **Wise Feedback** | 1.75 (1.12) | 3.51 (1.57) | $p$ < .001, CI [−2.17, −1.36] |

| | Pairwise comparisons | |
|---|---|---|
| Frequency of receiving: | Math course | English course |
| **Only Objective vs. Positive** | $p$ < .001, CI [1.16, 1.86] | $p$ = .237, CI [−.05, .60] |
| **Only Objective vs. Corrective** | $p$ < .001, CI [1.10, 1.81] | $p$ < .001, CI [.13, .77] |
| **Only Objective vs. Wise** | $p$ < .001, CI [1.93, 2.64] | $p$ = .026, CI [.02, .84] |
| **Positive vs. Corrective** | $p$ = 1.00, CI [−.37, .25] | $p$ = 1.00, CI [−.09, .45] |
| **Positive vs. Wise** | $p$ < .001, CI [.52, 1.04] | $p$ = 1.00, CI [−.15, .48] |
| **Corrective vs. Wise** | $p$ < .001, CI [.57, 1.10] | $p$ = 1.00, CI [−.34, .31] |

| | Means (*SDs*) | | |
|---|---|---|---|
| Anticipated impact of receiving: | Math course | English course | Pairwise comparisons |
| **Only Objective Feedback** | 5.37 (1.25) | 5.21 (1.34) | $p$ = 1.00, CI [−.20, .52] |
| **Positive Feedback** | 5.91 (1.08) | 5.73 (1.20) | $p$ = 1.00, CI [−.09, .46] |
| **Corrective Feedback** | 5.46 (1.15) | 5.34 (1.23) | $p$ = 1.00, CI [−.18, .41] |
| **Wise Feedback** | 5.77 (1.25) | 5.67 (1.17) | $p$ = 1.00, CI [−.20, .41] |

| | Pairwise comparisons | |
|---|---|---|
| Anticipated impact of receiving: | Math course | English course |
| **Only Objective vs. Positive** | $p$ < .001, CI [−.77, −.31] | $p$ < .001, CI [−.76, −.27] |
| **Only Objective vs. Corrective** | $p$ = 1.00, CI [−.36, .18] | $p$ = 1.00, CI [−.36, .10] |
| **Only Objective vs. Wise** | $p$ = .002, CI [−.72, −.09] | $p$ < .001, CI [−.75, −.16] |
| **Positive vs. Corrective** | $p$ < .001, CI [.19, .71] | $p$ < .001, CI [.17, .60] |
| **Positive vs. Wise** | $p$ = 1.00, CI [−.11, .38] | $p$ = 1.00, CI [−.16, .28] |
| **Corrective vs. Wise** | $p$ = .009, CI [−.59, −.04] | $p$ < .001, CI [−.56, −.09] |

## Results

### *Perceived Time and Effort of Instructors Giving Feedback*

Results of a paired samples *t*-test showed that students viewed instructors to put less time and effort into giving feedback to students in math ($M$ = 4.59, $SD$ = 1.48) versus English courses ($M$ = 5.59, $SD$ = 1.54), $t(182) = -6.22$, $p < .001$, $d$ = .46).[4]

### *Frequency of Receiving Different Types of Feedback*

We next ran a repeated measures ANOVA with Feedback Type as a within-subjects variable which showed a significant main effect, $F(7, 1267) = 92.77$, $p < .001$, $\eta_p^2 = .34$. Descriptive statistics and pairwise comparisons with Bonferroni correction are shown in Table 4 and in Figure 3 (left panel). Students said they received less positive, corrective, and wise feedback in their math versus English courses. There was no significant difference in receiving only objective feedback in math (vs. English) courses.

Comparing feedback types within each course, students in math courses said they received less positive, corrective, and wise feedback compared to only objective feedback; they also received more positive and corrective feedback than wise feedback; and reported no difference in receiving positive versus corrective feedback in math. In English courses, students received more corrective and wise feedback then only objective feedback; no other significant differences emerged in receipt of different types of feedback in English courses.

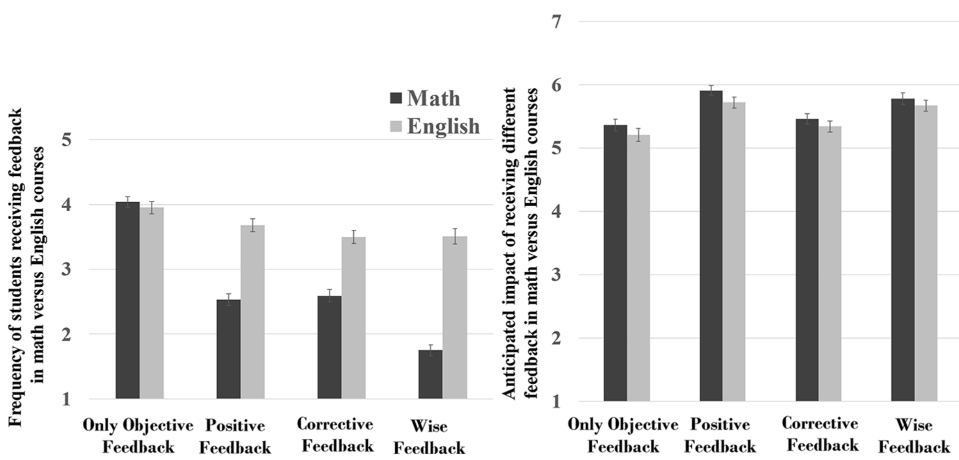


**Figure 3.** Study 2: Frequency of Students Receiving (Left Panel) and Anticipated Impact of Receiving (Right Panel) Different Types of Feedback From Instructors in Math and English Courses.
*Note.* Error bars represent the standard error around the mean of each condition.

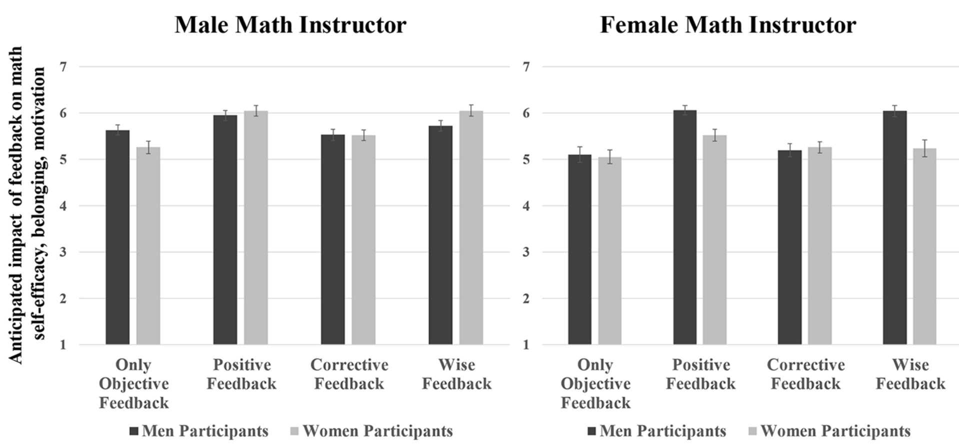


**Figure 4.** Study 2: Anticipated Impact of Receiving Different Types of Feedback From Male and Female Instructors in Math Courses as a Function of Participant Gender and Instructor Gender.
*Note.* Error bars represent the standard error around the mean of each condition.

## *Anticipated Impact of Receiving Different Types of Feedback*

Results of a repeated measures ANOVA showed a significant main effect of Feedback Type, $F(7, 1267) = 13.12, p < .001, \eta_p^2 = .07$. Descriptive statistics and pairwise comparisons are shown in Table 4 and in Figure 3 (right panel). There were no significant differences in students' reports of anticipated impact in math versus English courses if they were to have received only objective, positive, corrective, or wise feedback. However, if students were to have received positive feedback in their math course, they anticipated greater positive impact than if they had received only objective feedback or corrective feedback. The same pattern of findings emerged in English courses.

## *Moderation Analyses*

Next, we conducted a series of mixed-factorial ANOVAs with Participant Gender and Instructor Gender as the between-subjects variables and Feedback Type as the within-subjects variable to determine whether the frequency of receiving feedback and the perceived impact of receiving different types of feedback differed for women and men in math versus English courses, based on whether the instructor was male or female.[5]

**Table 5.** Descriptive Statistics and Results of Pairwise Comparisons Predicting Women's and Men's Anticipated Impact of Receiving Different Types of Feedback in Math Courses from Male and Female Instructors (Study 2).

| Women in math course | | | |
|---|---|---|---|
| | Means (*SDs*) | | |
| Anticipated impact of receiving: | Male instructor | Female instructor | Pairwise comparisons |
| **Only Objective Feedback** | 5.26 (1.31) | 5.05 (1.41) | $p$ = .460, CI [−.36, .79] |
| **Positive Feedback** | 6.05 (1.07) | 5.52 (1.17) | $p$ = .036, CI [.04, 1.02] |
| **Corrective Feedback** | 5.52 (1.13) | 5.26 (1.16) | $p$ = .332, CI [−.27, .80] |
| **Wise Feedback** | 6.05 (1.18) | 5.23 (1.73) | $p$ = .005, CI [.25, 1.39] |

| Women in math course | | |
|---|---|---|
| Pairwise Comparisons | | |
| | Male instructor | Female instructor |
| **Only Objective vs. Positive** | $p < .001$, CI [−1.02, −.55] | $p$ = .018, CI [−.86, −.08] |
| **Only Objective vs. Corrective** | $p$ = .074, CI [−.54, .03] | $p$ = .365, CI [−.67, .25] |
| **Only Objective vs. Wise** | $p < .001$, CI [−1.12, −.46] | $p$ = .505, CI [−.71, .35] |
| **Positive vs. Corrective** | $p < .001$, CI [.26, .79] | $p$ = .234, CI [−.17, .69] |
| **Positive vs. Wise** | $p$ = .977, CI [−.26, .25] | $p$ = .170, CI [−.13, .71] |
| **Corrective vs. Wise** | $p < .001$, CI [−.81, −.25] | $p$ = .897, CI [−.43, .49] |

| Men in math course | | | |
|---|---|---|---|
| | Means (*SDs*) | | |
| Anticipated impact of receiving: | Male instructor | Female instructor | Pairwise comparisons |
| **Only Objective Feedback** | 5.63 (1.03) | 5.10 (1.62) | $p$ = .170, CI [−.23, 1.29] |
| **Positive Feedback** | 5.95 (1.03) | 6.06 (1.01) | $p$ = .737, CI [−.76, .54] |
| **Corrective Feedback** | 5.53 (1.13) | 5.19 (1.34) | $p$ = .348, CI [−.37, 1.04] |
| **Wise Feedback** | 5.72 (1.07) | 6.04 (1.12) | $p$ = .404, CI [−1.07, .43] |

| Men in math course | | |
|---|---|---|
| Pairwise comparisons | | |
| | Male instructor | Female instructor |
| **Only Objective vs. Positive** | $p$ = .005, CI [−.54, −.10] | $p < .001$, CI [−1.52, −.40] |
| **Only Objective vs. Corrective** | $p$ = .404, CI [−.15, .37] | $p$ = .803, CI [−.74, .58] |
| **Only Objective vs. Wise** | $p$ = .555, CI [−.39, .21] | $p$ = .017, CI [−1.70, −.17] |
| **Positive vs. Corrective** | $p < .001$, CI [.18, .67] | $p$ = .006, CI [−.25, 1.50] |
| **Positive vs. Wise** | $p$ = .060, CI [−.01, .46] | $p$ = .945, CI [−.58, .62] |
| **Corrective vs. Wise** | $p$ = .128, CI [−.06, −.46] | $p$ = .011, CI [−1.51, −.20] |

*Note*. For pairwise analyses, the Bonferroni correction was used to adjust for multiple comparisons. CI = 95% confidence interval of the mean difference.

## *Frequency of Receiving Different Types of Feedback*

Results showed no significant two-way or three-way interactions among participant gender, instructor gender, or feedback type in predicting frequency of receiving different types of feedback in math courses, all *p*s > .29, or English courses, all *p*s > .31.

## *Anticipated Impact of Receiving Feedback in Math Courses*

Results showed a significant Participant Gender × Instructor Gender × Feedback Type interaction, $F(3, 525) = 4.53$, $p = .004$, $\eta_p^2 = .03$ (see Figure 4). Decomposing this three-way interaction revealed a significant Participant Gender × Feedback Type interaction for students who had

**Table 6.** Descriptive Statistics for Student Demographics (Study 3).

| Ethnicity | Year in school | Major | Semester enrolled in calculus course |
|---|---|---|---|
| 52% White | 76% freshmen | 96% STEM | 77% in Fall 2020 |
| 23% Asian/Asian American | 16% sophomores | 4% non-STEM | 23% in Spring |
| 10% Latinx/Hispanic | 6% juniors | | |
| 8% Black | 1% seniors | | |
| 7% Other ethnicities | 1% 5th year + | | |

a male math instructor, $F(3, 420) = 5.05$, $p = .002$, $\eta_p^2 = .04$, but not a female math instructor, $F(3, 105) = 1.59$, $p = .197$, $\eta_p^2 = .04$. Table 5 shows that women anticipated greater impact if they were to have received positive feedback (or wise feedback) from their male (vs. female) math instructor. For men, there were no significant differences in anticipated impact of feedback type from a male or female math instructor.

### *Anticipated Impact of Receiving Feedback in English Courses*

Results showed no significant two-way or three-way interactions in predicting anticipated impact of receiving feedback in English courses, all $ps > .062$.

## Discussion

Study 2 provided further evidence for the culture of feedback in college math courses. Consistent with Study 1, students reported perceiving that instructors put less time and effort into giving feedback in math (vs. English) courses and, when they did, they gave positive feedback less often than only objective feedback. Importantly, however, all students said they would have benefited from receiving positive feedback in their math (vs. English) courses. In particular, women said they would have felt more confident, a sense of belonging, and motivated in their math course if they had received positive feedback from their male math instructor. Men also said they would have benefited from receiving positive feedback in math, but instructors' gender did not matter. In English courses, women (vs. men) said they would have benefited from receiving positive or corrective feedback, although instructor gender did not matter as it did for math courses.

Together, these studies suggest a mismatch between what students say would be helpful versus the current norms and practices of feedback-giving in math courses. Both instructors and students report that positive feedback is underutilized in math courses, which could pose a barrier to students' interest and success in fields that require math, especially for women who may doubt their abilities and belonging in STEM.

## Study 3: Effects of Positive Feedback in College Calculus Courses

Study 3 examined the real-world consequences of giving positive (vs. objective) feedback to students in introductory college calculus courses. Based on our proposed conceptual model, we expected students who received positive (vs. objective) feedback on exams to show increased self-efficacy, belonging, and intentions to adopt proactive study habits (i.e., greater motivation to succeed). In turn, we expected one or more of these mediators to predict more favorable attitudes/identity/interest in STEM and higher final math course grades. In addition, we tested whether these effects were stronger among URM students in STEM (i.e., women, racial minorities).

### *Participants and Procedure*

We recruited 657 undergraduate students (62% men, 37% women, 1% non-binary) enrolled in introductory college calculus courses at two large public research universities ($N = 320$ at University 1, $N = 337$ at University 2) in the United States. Table 6 reports demographics. During the second week of the semester, instructors at both institutions sent an email to students and also made a verbal announcement offering an extra credit opportunity to participate in a research study. Students were told they would complete brief online surveys throughout the semester and that findings from the study could be used to improve the quality of instruction in college calculus courses. If students chose not to participate but still wanted to earn extra credit, they could complete an alternate writing assignment.[6]

To assign participants to feedback condition, we first determined which instructors agreed to be involved in the study at each institution and how many students were enrolled in each section of the calculus course(s) they were teaching that semester. Then, using a matched-pairs approach, we paired up similar size classes (i.e., recitation sections) and randomly assigned entire sections of students to receive either positive feedback ($n = 327$) or only objective feedback ($n = 330$) from their instructor on two exams during the semester. We sought to ensure that a relatively equal number of students were assigned to each feedback condition based on the student enrollments per section. Instructors were

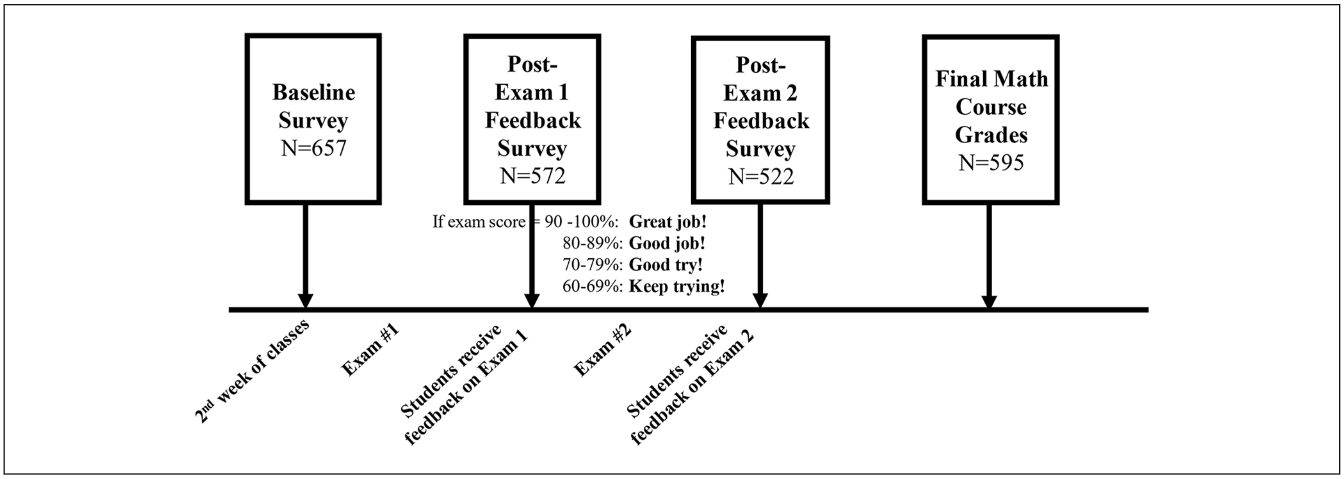


**Figure 5.** Study 3 Timeline of Data Collection Points.

aware of which feedback condition students were assigned to, as they were the ones giving the feedback. In a few cases, teaching assistants in the recitation section assisted with the provision of feedback, but students were always led to believe that the feedback came from the instructor.

All students in a particular section were assigned to the same feedback condition and kept in that same feedback condition throughout the semester so they received only objective feedback on both exams or positive feedback on both exams. Final course grades were based on assignments and exam performance throughout the semester based on a standard grading rubric. Figure 5 shows the timeline of the study.

At baseline, students reported demographics, their initial math self-efficacy, belonging in math, intended study habits in their math course, and attitudes/identity/interest in STEM. Approximately 3 to 4 weeks later, students took their first exam, and 3 to 6 weeks after that, the second exam.[7] Due to the COVID-19 pandemic, students took exams online/remotely. At University 1, students uploaded their exams to Gradescope, a software program that streamlines the grading process. At University 2, students took exams using the quiz feature of the Canvas Learning Management System while being proctored. Exams at both institutions involved a mixture of question types including short-answer and longer written responses.

Instructors then graded exams as they normally would. Students assigned to the *objective feedback condition* received their exam score (e.g., +88/100), which they could view via the online grading center of the learning management system at their institution (e.g., Blackboard at University 1, Canvas at University 2). Students assigned to the *positive feedback condition* also received their exam score, plus one typewritten comment they could view, by clicking on their exam score in the online grading center. The comment corresponded to the scale described below, which instructors copied and pasted into the online grading center.

At University 1, this comment went into a "Feedback to Learner" box that appeared as a bubble icon next to students' exam scores in the online grading center with a message tailored to different ranges of scores: 89.5% to 100% ("*Great job*!"); 79.5% to 89.4% ("*Good job*!"); 69.5% to 79.4% ("*Good try*!"); and 69.4% or below ("*Keep trying*!"). At University 2, this comment was given in a feedback box with a code. Students had to reply to the instructor's feedback with that same code so that instructors knew that the student had seen their exam grades and feedback. Students across both feedback conditions also received the following instructions: "Please complete this brief online survey, which is for the study you are participating in, for 1% extra credit in this class" with a link to the postexam feedback survey.

Students completed each postexam feedback survey (two in total) up to 1 week after receiving their graded exams. The postexam survey contained a manipulation check assessing perceptions of the feedback they received, shortened versions of math self-efficacy, belonging in math, intended study habits in their math course, and attitudes/identity/interest in STEM. Instructors also submitted students' final math grades as a measure of performance.

## *Materials*

Responses to the following measures were collected at the beginning of the semester (baseline) and following both exam 1 and exam 2. Measures were collected with response scales ranging from 1 (*strongly disagree/not at all true of me/very unlikely*) to 7 (*strongly agree/very true of me/very likely*), with instructions adapted to reflect current feelings.

*Demographics.* At baseline only, students reported their gender, age, race/ethnicity, year in school, major, current university, and which semester they were taking the calculus course (Fall or Spring semester).

**Table 7.** Zero-Order Correlations between Baseline and Post-exam 1 and Post-exam 2 Measures (Study 3).

| Study Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Baseline measures | | | | | | | | | |
| 1. Math self-efficacy | — | .61*** | .33*** | .56*** | .47*** | .17*** | .31*** | .22*** | .15*** |
| 2. Math belonging | .61*** | — | .38*** | .50*** | .66*** | .22*** | .26*** | .20*** | .12*** |
| 3. Intended math study habits | .33*** | .38*** | — | .23*** | .29*** | .56*** | .21*** | .11* | −.04 |
| Post-exam measures | | | | | | | | | |
| 4. Math self-efficacy | .62*** | .47*** | .26*** | — | .79*** | .24*** | .32*** | .54*** | .22*** |
| 5. Math belonging | .50*** | .68*** | .30*** | .73*** | — | .31*** | .35*** | .45*** | .14** |
| 6. Intended math study habits | .14*** | .19*** | .57*** | .21*** | .28*** | — | .27*** | .12* | −.09* |
| 7. STEM interest | .26*** | .26*** | .22*** | .26*** | .32*** | .25*** | — | .15*** | .02 |
| 8. Exam score | .20*** | .17*** | .19*** | .53*** | .43*** | .17*** | .15*** | — | .14** |
| 9. University affiliation | .15*** | .12** | −.04 | .28*** | .21*** | −.10* | −.01 | .17*** | — |

*Note.* University affiliation was coded as 1 = University 1, 0 = University 2. STEM interest = composite of attitudes/identity/interest in STEM fields. Correlations below the diagonal are between baseline and post-exam 1 measures; correlations above the diagonal are between baseline and post-exam 2 measures.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

*Math Self-Efficacy.* To measure math self-efficacy, students responded to seven items from the Motivated Strategies for Learning Questionnaire (MSLQ, Duncan & McKeachie, 2005), which was adapted to assess self-efficacy for learning and performance in their calculus course. Sample items were, "I'm confident I can do an excellent job on the assignments and tests in this course" and "I expect to do well in this class." Items were selected based on past work showing strong correlations between these items and final course grades (Duncan & McKeachie, 2005; $\alpha_{baseline} = .95$, $\alpha_{post\text{-}exam1} = .96$, $N = 572$; $\alpha_{post\text{-}exam2} = .97$, $N = 521$).

*Math Belonging.* Students were asked to think about the math course they were currently taking and responded to the prompt: "When I am in this math class . . ." followed by statements such as, "I feel that I belong to the math community" and "I feel accepted" (Good et al., 2012; 16 items, $\alpha_{baseline} = .90$, $\alpha_{post\text{-}exam1} = .92$, $N = 571$; $\alpha_{post\text{-}exam2} = .93$, $N = 570$).

*Math Study Habits.* Students reported their general habits (baseline) and intentions (postexam) to engage in effective study habits and seek help by responding to items from the MSLQ-Resource Management Strategies subscales (Duncan & McKeachie, 2005). Sample items were, "To prepare for the next exam, I plan to . . ." "make sure to keep up with the weekly readings and assignments for this course" and "Ask another student in this class for help if I can't understand the material in this course" (15 items, $\alpha = .83_{baseline}$, $\alpha_{post\text{-}exam1} = .85$, $N = 572$; $\alpha_{post\text{-}exam2} = .86$, $N = 519$).

*STEM Attitudes/Identity/Interest.* Students reported how much they liked, identified with, and were interested in pursuing STEM (Park et al., 2018; Stout et al., 2011). Sample items were, "How much do you like STEM (Science, Technology, Engineering, Math)?" "How much do you consider yourself to be a 'STEM' person?" and "How likely are you to pursue a degree or career in STEM?" from 1 = *not at all* to 7 = *very much* (6 items, $\alpha = .92_{baseline}$, $\alpha_{post\text{-}exam1} = .95$, $N = 572$; $\alpha_{post\text{-}exam2} = .95$, $N = 520$).

*Perceptions of Feedback.* Students rated how positive, encouraging, supportive, negative, condescending, and patronizing the feedback they received seemed from 1 = *not at all* to 7 = *very*. Negative items were reverse-scored and averaged together with the positive items to compute an overall score with higher numbers reflecting more favorable perceptions of the feedback ($\alpha_{post\text{-}exam1} = .80$, $N = 458$; $\alpha_{post\text{-}exam2} = .81$, $N = 462$).[8]

*Final Math Grades.* At the end of the semester, instructors reported students' ($N = 595$) final math grades using a percentage scale from 0% to 100% .

## Results

We first conducted zero-order correlations among the study variables (see Table 7 for postexam 1 and 2 correlations). We then examined students' perceptions of the feedback they received, followed by mediation analyses to see how receiving positive (vs. objective) feedback affected their math self-efficacy, belonging, intended study habits, and in turn, their STEM attitudes/identity/interest and final math grades. We then conducted moderated mediation analyses to determine whether students' gender and ethnicity differentially affected their responses to positive (vs. objective) feedback.

### Perceptions of Feedback

We conducted *t*-tests to examine participants' perceptions of the feedback they received on the two exams in their calculus course.[9] We had 80% power to detect a small effect size of

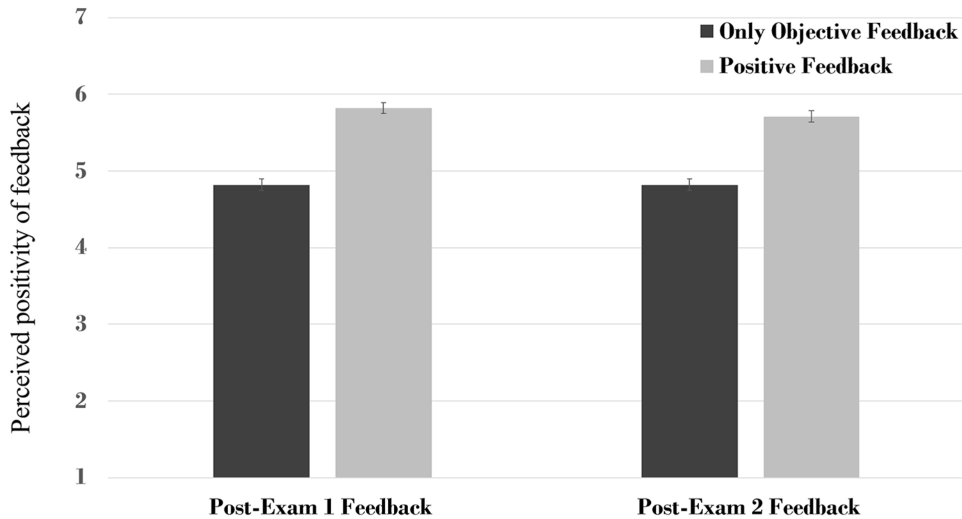


Figure 6. Study 3: Participants' ratings of the perceived positivity of feedback.

**Figure 6.** Study 3: Participants' Ratings of the Perceived Positivity of Feedback.

$d = .26$. Students who received positive feedback on exam 1 rated the feedback as more positive ($n = 255$, $M = 5.82$, $SD = 1.17$) than students who received only objective feedback on exam 1 ($n = 203$, $M = 4.82$, $SD = 1.09$), $t(456) = 9.35$, $p < .001$, $d = .88$, 95% confidence interval [CI] [.79]. After exam 2, students also rated the feedback as more positive when they received positive ($n = 249$, $M = 5.71$, $SD = 1.22$) versus only objective feedback ($n = 213$, $M = 4.82$, $SD = 1.05$), $t(460) = 8.41$, $p < .001$, 95% CI [.68, 1.10] (see Figure 6). Thus, the feedback manipulation was effective: students perceived positive (vs. only objective) feedback to be more supportive and encouraging, and less patronizing and condescending.[10]

## Mediation Analyses

Mediation analyses tested whether feedback condition (1 = *positive feedback*, 0 = *only objective feedback*) predicted the post-exam 1 and post-exam 2 mediators (i.e., math self-efficacy, belonging, intended study habits in math) and, in turn, the outcome variables (i.e., post-exam 1 and 2 STEM attitudes/identity/interest and final math grades, respectively). For all analyses, we controlled for variables that were significantly correlated with at least one mediator or outcome (i.e., baseline measures of the mediators, initial STEM interest, exam 1 or 2 scores, and university affiliation). These analyses were conducted using PROCESS version 4.3 for R.

## Post-exam 1

*STEM Interest.* We conducted mediation analyses using Hayes' (2018) PROCESS macro (model 4) examining the effects of receiving positive (vs. only objective) feedback on exam 1 on students' post-exam 1 STEM-related mediators and, in turn, their STEM attitudes/identity/interest (i.e., STEM interest). Results showed no significant direct or indirect effects of feedback condition in predicting post-exam 1 STEM interest (see Figure 7).

*Final Math Grades.* We conducted mediation analyses examining the effects of receiving positive (vs. only objective) feedback on exam 1 on students' post-exam 1 STEM-related mediators and, in turn, their final math grades. Results showed a significant effect of feedback condition on intended math study habits; those who received positive feedback on exam 1 reported greater intentions to engage in proactive study habits in their math course. Results of indirect tests revealed that greater intended study habits, in turn, were related to earning higher final math course grades (see Figure 8).

## Post-exam 2

*STEM Interest.* Results of mediation analyses examining the effects of receiving positive (vs. only objective) feedback on exam 2 on students' post-exam 2 STEM-related mediators and, in turn, their STEM interest showed a significant indirect effect of belonging. Participants who received positive (vs. only objective) feedback on exam 2 reported greater belonging in math, which predicted more favorable attitudes/identity/interest in STEM after exam 2 (see Figure 9).

*Final Math Course Grades.* Results of mediation analyses examining the effects of receiving positive (vs. only objective) feedback on exam 2 on students' post-exam 2 STEM-related mediators and, in turn, their final math grades showed a significant indirect effect of self-efficacy. Participants who

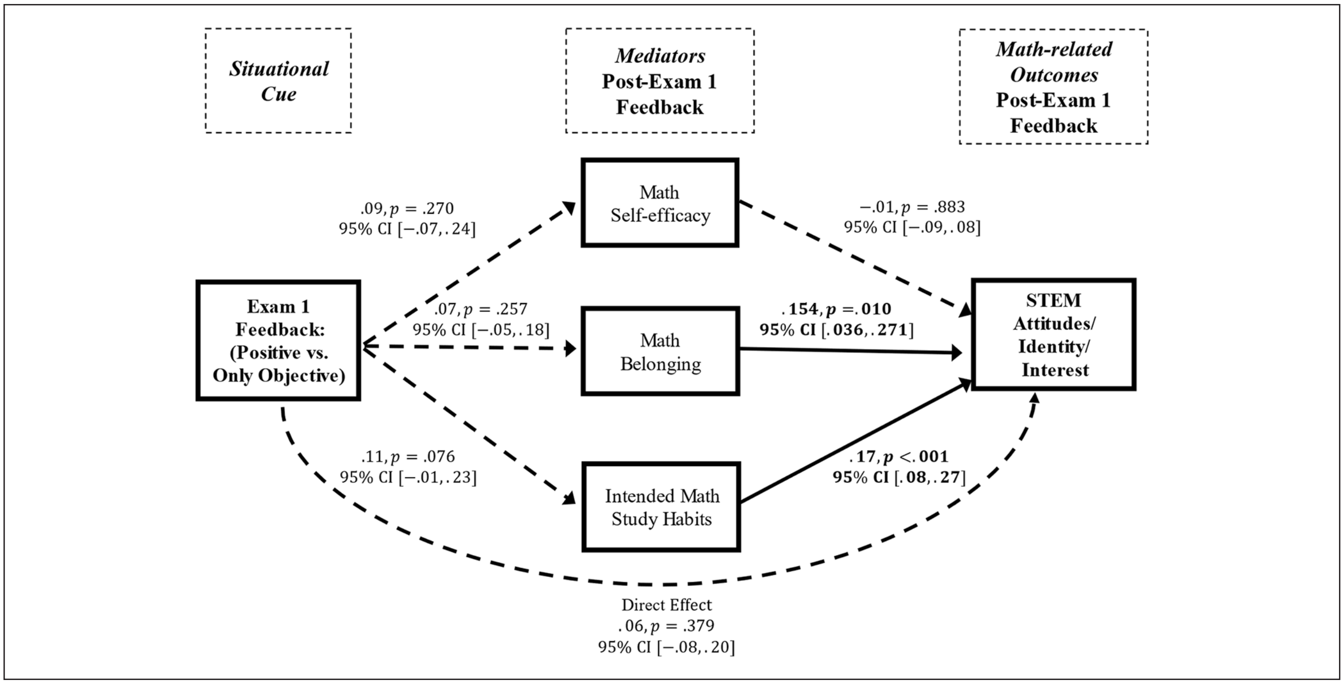


**Figure 7.** Effects of Post-exam 1 Feedback on Students' STEM Attitudes/Identity/Interest.
*Note.* Results show unstandardized coefficients. Solid lines and bolded numbers reflect significant paths; dotted lines reflect nonsignificant paths. Covariates: baseline measures of math-self-efficacy, belonging, intended math study habits, STEM attitudes/identity/interest, exam 1 scores, university affiliation. Numbers in brackets reflect 95% confidence intervals based on 5,000 bootstrapped samples.

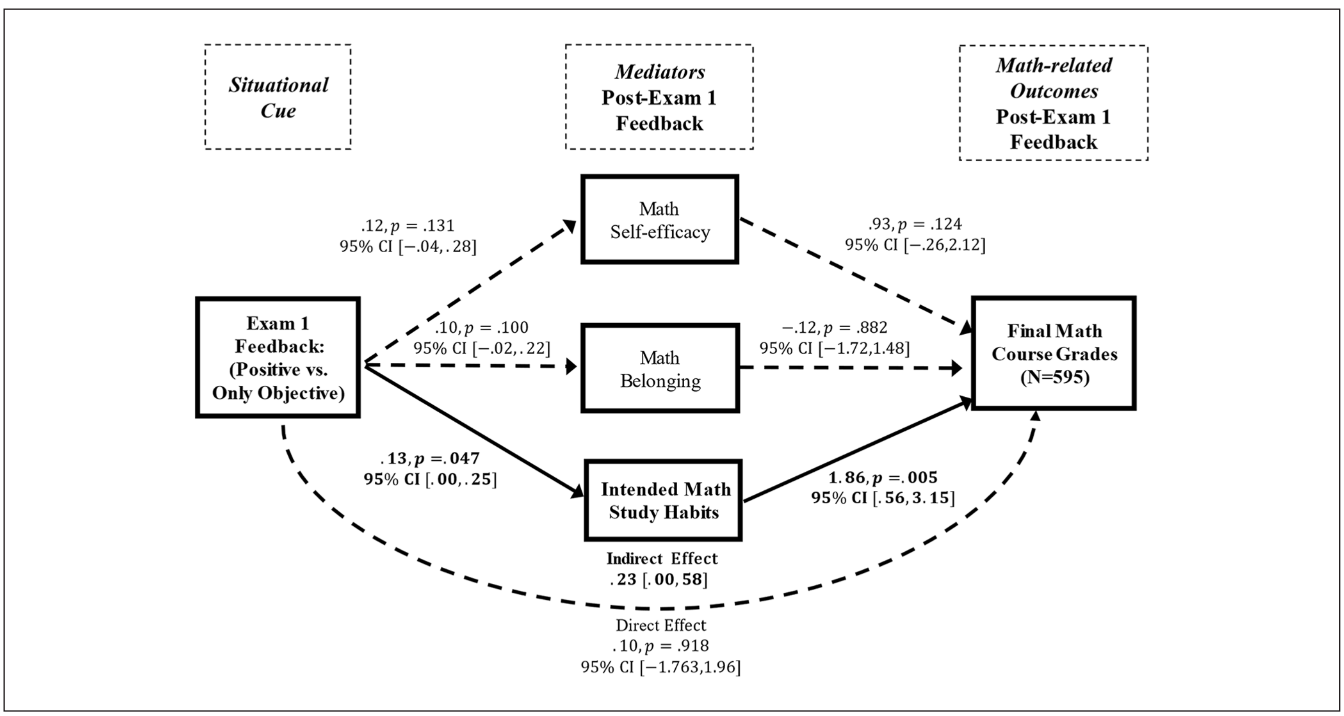


**Figure 8.** Effects of Post-exam 1 Feedback on Students' Final Math Course Grades.
*Note.* Results show unstandardized coefficients. Solid lines and bolded numbers reflect significant paths; dotted lines reflect nonsignificant paths. Covariates: baseline measures of math-self-efficacy, belonging, intended math study habits, STEM attitudes/identity/interest, exam 1 scores, university affiliation. Numbers in brackets reflect 95% confidence intervals based on 5,000 bootstrapped samples.

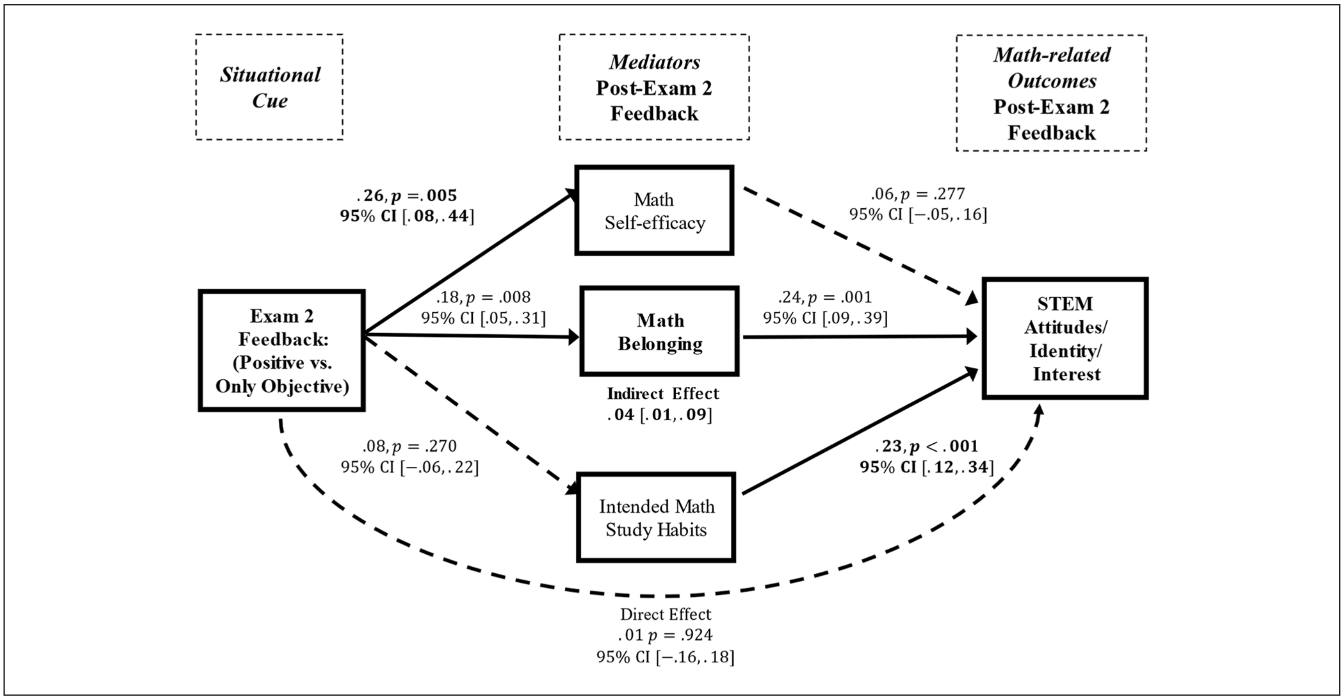


**Figure 9.** Effects of Post-exam 2 Feedback on Students' STEM Attitudes/Identity/Interest.
*Note.* Results show unstandardized coefficients. Solid lines and bolded numbers reflect significant paths; dotted lines reflect nonsignificant paths. Covariates: baseline measures of math-self-efficacy, belonging, intended math study habits, STEM attitudes/identity/interest, exam 2 scores, university affiliation. Numbers in brackets reflect 95% confidence intervals based on 5,000 bootstrapped samples.

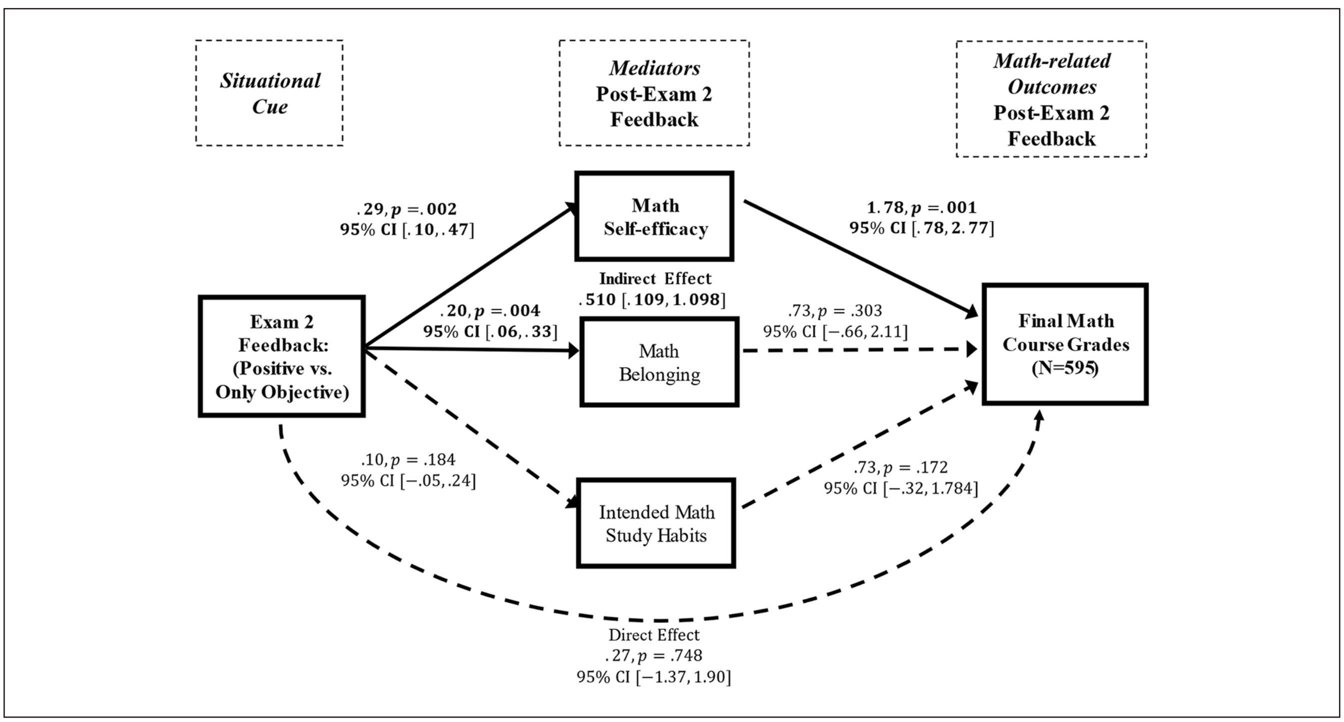


**Figure 10.** Effects of Post-exam 2 Feedback on Students' Final Math Course Grades.
*Note.* Results show unstandardized coefficients. Solid lines and bolded numbers reflect significant paths; dotted lines reflect nonsignificant paths. Covariates: baseline measures of math-self-efficacy, belonging, intended math study habits, STEM attitudes/identity/interest, exam 2 scores, university affiliation. Numbers in brackets reflect 95% confidence intervals based on 5,000 bootstrapped samples.

received positive feedback on exam 2 reported higher math self-efficacy, which predicted higher final math course grades. No other indirect effects were significant (see Figure 10).

## Moderated Mediation Analyses

To examine whether participants' gender or ethnicity interacted with feedback condition to predict the mediators and the outcomes, we conducted moderated mediation analyses using Hayes' (2018) PROCESS macro (model 8). Specifically, we tested participants' gender (0 = men, 1 = women; nonbinary not included in analyses) and ethnicity (0 = non-White or non-Asian, 1 = White or Asian).[11] With a sample size of $N = 657$, our dataset should have been able to detect an effect size of $f = .10$ for any 2 (positive vs. only objective feedback) × 2 (male vs. female participant gender or White/Asian vs. non-White/Asian) interaction on any mediator variable.

As before, analyses for the dependent measures collected from post-exam 1 and post-exam 2 surveys were conducted separately. We examined post-exam 1 and post-exam 2 mediators in predicting post-exam STEM interest and final math grades. Also, as before, for each analysis we controlled for variables that were significantly correlated with the outcomes. For post-exam 1 STEM interest and final math grades, we controlled for baseline mediators, baseline STEM interest, exam 1 scores, and university affiliation; for post-exam 2 STEM interest, we controlled for these variables and replaced exam 1 with exam 2 scores.

*Moderating Effects of Gender.* For both post-exam 1 and post-exam 2, the index of moderated mediation examining gender as a moderator was not significant in predicting STEM interest or final math course grades (see Table S1 in Supplemental Materials).[12]

*Moderating Effects of Ethnicity.* For post-exam 1, the index of moderated mediation examining ethnicity as a moderator was not significant in predicting STEM interest or final math course grades. However, for post-exam 2, the index of moderated mediation was significant in predicting STEM interest and final math grades. Specifically, URMs (i.e., non-White, non-Asian participants) who received positive (vs. only objective) feedback on exam 2 reported greater belonging in math, which predicted greater STEM interest. URM students who received positive (vs. only objective) feedback on exam 2 also reported higher math self-efficacy, which predicted higher final math course grades (see Table S2 in Supplemental Materials).

# Discussion

Overall, the results of Study 3 showed that students in introductory college calculus courses benefited from receiving positive feedback on exams. Specifically, students who received positive (vs. only objective) feedback from their math instructor on exam 1 showed increased intentions to adopt proactive study habits, which predicted higher final math course grades. Likewise, students who received positive (vs. only objective) feedback on exam 2 reported increased belonging in math and math self-efficacy, which predicted greater interest in STEM and higher final math course grades, respectively. Furthermore, the postexam 2 feedback results were especially strong for URMs. These findings emerged even after controlling for baseline levels of the mediators and initial STEM interest, actual exam performance, and institution. Together, these findings underscore the benefits of positive feedback and reveal key psychological mechanisms contributing to STEM interest and performance.

Although Study 2 found that women anticipated that positive feedback in their math course would lead greater self-efficacy, belonging, and motivation, this study found no gender differences—only racial-ethnic difference—in the effects of positive feedback. While there are important differences between these historically disadvantaged groups, what they share in common is negative ability-based stereotypes and uncertainty about belonging in academic settings. For example, women are stereotyped as being bad at math (Inzlicht & Ben-Zeev, 2000; Spencer et al., 1999) and doubt their belonging in STEM-related contexts (Cheryan et al., 2009; Murphy et al., 2007); similarly, Black students are stereotyped as being intellectually inferior (Steele & Aronson, 1995) and feel uncertain about their belonging in academic settings (Walton & Cohen, 2007). Given these common concerns, we might have expected positive feedback to be beneficial for students across these backgrounds. However, some research suggests that not all groups respond the same way to cues in the environment, and certain cues may be more impactful than others based on one's group membership (Apfelbaum et al., 2016). Indeed, in this study, positive feedback was beneficial for men and women alike, but especially for URMs.

Other studies suggest that membership within the same group can lead to different reactions to different environmental cues. For example, past research found that Black women—especially those who were highly conscious of the potential for being discriminated against based on their race and gender—reported greater trust and belonging in a science company when they saw a profile of a Black woman scientist compared to seeing a Black man scientist, a White woman scientist, or no profile (Pietri et al., 2018). In another study, Black women reported greater trust and belonging when they saw a profile of a Black man or woman professor in the School of Science and Engineering versus a White man or woman professor (Johnson et al., 2019). In sum, both the current findings and previous work suggest that interventions may not affect all groups equally, whether that be in people's responses to the same external cues or in group members' responses to different cues.

## General Discussion

The present research reveals a mismatch between the prevailing norms and practices of giving feedback in college math courses and the types of feedback that are beneficial for students. Studies 1 and 2 provided evidence—from both instructor and student perspectives—that positive feedback is underutilized in math (vs. English) courses, consistent with perceived norms in these departments. Instructors (and students) also reported that instructors put less time and effort into giving feedback to students in (vs. English) courses, suggesting that the mere practice of feedback-giving may not be emphasized in math (vs. English) settings. Furthermore, women in particular said that they would have felt greater confidence, belonging, and motivation if they had received positive feedback from their male math instructors. Positive feedback from math instructors was perceived by male students as beneficial, regardless of whether it came from a male or female instructor. Such findings are consistent with previous work showing that college women benefit most from receiving positive feedback from perceived gatekeepers in STEM (Park et al., 2018; Park, O'Brien et al., 2023).

Study 3 examined the real-world consequences of positive feedback and found that students in introductory college calculus courses who received positive feedback on their first exam reported greater intentions to adopt proactive study habits in their math course, which predicted earning higher final math grades. Receiving positive feedback on the second exam was related to greater sense of math belonging, which predicted greater interest in STEM, and to higher math self-efficacy, which predicted higher final math grades. These latter relationships were strongest among URM students. However, given that the findings for URM students varied across studies—with some of our studies showing moderating effects of gender (Study 2) and other studies showing moderating effects of ethnicity but not gender (Study 3)—further research is needed to test the robustness of these findings. Nevertheless, all students showed better STEM-related outcomes after receiving positive (vs. only objective) feedback, suggesting that positive feedback is beneficial for all students, regardless of their background.

### Strengths and Limitations

These studies support the cues hypothesis by demonstrating that even small cues in the environment can have a real-world impact. Past work found that situational cues in STEM environments, such as physical features (Cheryan et al., 2009), the presence of ingroup versus outgroup members (Dennehy & Dasgupta, 2017; Murphy et al., 2007), and social cues from authority figures (Canning et al., 2019; Park et al., 2018) predict outcomes such as belonging, motivation, and interest in STEM. In the present research, positive feedback—as a minimal situational cue—was especially effective for URMs, suggesting that cues conveying psychological safety may be particularly important for members of marginalized groups (Park, Naidu, et al., 2023; Park et al., under review). Furthermore, we identified key psychological mechanisms—belonging, self-efficacy, and intended study habits—linking positive feedback with STEM-related outcomes. Findings emerged even after controlling for initial math belonging, self-efficacy, STEM interest, and actual exam performance, underscoring the unique role of positive feedback.

Although positive feedback is likely to be beneficial, the specific departmental norms of giving positive feedback may differ depending on institution type or broader cultural context. For example, in individualistic cultures such as North America, receiving positive feedback is consistent with self-enhancement motives to feel good about oneself (Crocker & Park, 2004); in collectivistic cultures such as East Asia, critical feedback and self-improvement are prioritized (Heine et al., 2001). Thus, norms of giving positive feedback—and the magnitude of its effects—may vary depending on both local and broader cultural contexts.

It is also noteworthy that the current studies were conducted during the COVID-19 pandemic when schools and colleges across the United States were shut down and teaching was moved online. Originally, Study 3 was intended to take place in-person with the instructor giving handwritten feedback to students on exams. Given the move to a virtual format, students received feedback through their school's learning management system. This form of delivery may have potentially weakened the effects of the feedback students received, especially in terms of hypothesized effects of instructor's gender. For example, we expected women to benefit more from receiving positive feedback from a male (vs. female) instructor, but results of Study 3 showed no gender differences in the mediators, perhaps due to the impersonal nature of the feedback compared to receiving feedback in-person.

### Future Directions

In the present research, positive feedback was presented with a brief comment (e.g., "Good job!") depending on students' actual exam scores and was delivered through an online learning management system. While effects were observed using this approach, future work could examine whether positive feedback on other key assessments, or the combination of positive written and verbal feedback, could exert an even stronger or longer-lasting effect on students' self-perceptions and performance.

Future research could also examine other potential mediators. For example, positive feedback might shift students' perceptions of instructors as being more caring, approachable, and believing in students' ability to succeed, which may shape students' own perceptions of self-efficacy and belonging, boosting their subsequent interest and performance in STEM. Positive feedback might also increase students' perceptions that intelligence is malleable, leading to better performance. Consistent with this idea, after receiving positive feedback on exam 1, students reported greater intentions to engage in productive study habits, which predicted higher final math grades.

There may also be possible downsides to positive feedback. For example, comfort-oriented feedback—meant to reassure students that not everyone has what it takes to succeed in math—has been shown to increase perceptions that instructors hold a fixed mind-set about math ability, which decreases academic motivation (Rattan et al., 2012). Praise may also diminish motivation by increasing pressure to perform (Baumeister et al., 1990; Kamins & Dweck, 1999). Future research could therefore examine boundary conditions in which positive feedback may or may not be helpful.

## Conclusion

As society becomes increasingly reliant on STEM, research is needed to identify ways to recruit and retain a diverse pool of individuals to pursue degrees and careers in STEM. The current research shows a mismatch between the types of feedback that college math instructors typically give in their courses and what actually benefits students. Notably, this work expands our understanding of how minimal instructor behaviors—giving positive feedback in math courses—serves as a powerful situational cue to boost students' self-efficacy, belonging, and intended study habits in STEM, especially for URMs.

### Authors' Note

Portions of this research were presented at the 2022 meetings of the American Association for the Advancement of Science, Duck Social Cognition Conference, Society of Experimental Social Psychology, Society for Personality and Social Psychology, and the 2023 Joint Mathematics Meetings and European Association of Social Psychology Conference.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a National Science Foundation grant (grant no. DUE-2011321) awarded to the first author.

### ORCID iDs

Lora E. Park https://orcid.org/0000-0002-2779-3034

Deborah E. Ward https://orcid.org/0000-0002-2574-4861

### Supplemental Material

Supplemental material is available online with this article.

### Notes

1. Instructors who did not teach introductory courses were asked to list a different undergraduate course they taught in their department.
2. The response option 6 = *not applicable (I didn't have exams or assignments in this course*) was coded as missing. Varying degrees of freedom for the correlations reflect these missing values, which may be due to the fact that 37% of English instructors reported giving exams in their courses versus approximately 99% of math instructors.
3. The response option 6 = *not applicable (I didn't have exams or assignments in this course*) was coded as missing. Varying degrees of freedom for the correlations reflect these missing values.
4. For the paired samples *t*-test, we had 80% power to detect a small effect size of $dz = .20$ with alpha = .05.
5. We also tested whether students' ethnicity (coded as 0 = White or Asian; 85% of the sample, 1 = non-White, non-Asian; 15% of the sample) moderated the effects. There were no consistent, significant results when conducting these analyses in predicting frequency or anticipated impact of receiving different feedback types in math versus English courses.
6. No students opted to complete the alternate written assignment.
7. The number of exams administered varied across schools. However, because one school had only two midterm exams, we gave feedback and analyzed responses to feedback on the first two exams across schools.
8. Not all participants rated their perceptions of feedback, so the sample sizes are relatively lower for these measures than for the other measures in the postexam feedback surveys.
9. Sample sizes varied across exams 1 and 2 because not all students completed both post-exam surveys.
10. We also ran a series of ANOVAs to test whether participants reacted differently to positive feedback based on their exam performance (see Supplemental Materials for full results). Overall, there were no consistently meaningful differences in perceptions or responses to positive feedback based on the specific comment (i.e., "Great job!" vs. "Good job!" vs. "Good try!" vs. "Keep trying!") that students received.
11. PROCESS Model 8 tests whether a variable (e.g., race or gender in this study) moderates both the "a" path and the "direct" path (i.e., the effects of feedback condition on the proposed mediators and outcome), but not the "b" path from the mediator to the outcome.
12. The Feedback Condition × Participant Gender × Instructor Gender interaction was not significant in predicting the outcomes.

## References

Apfelbaum, E. P., Stephens, N. M., & Reagans, R. E. (2016). Beyond one-size-fits-all: Tailoring diversity approaches to the representation of social groups. *Journal of Personality and Social Psychology*, *111*(4), 547–566.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147.

Baumeister, R. F., Hutton, D. G., & Cairns, K. J. (1990). Negative effects of praise on skilled performance. *Basic and Applied Social Psychology*, *11*, 131–148.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology*, *79*, 474–482.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation of interest and performance. *British Journal of Educational Psychology*, *58*(1), 1–14. https://doi.org/10.1111/j.2044-8279.1988.tb00874.x

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, *64*, 363–423.

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, *5*(2), Article eaau4734.

Cheryan, S., Plaut, V. C., Davies, P., & Steele, C. M. (2009). Ambient belonging: How stereotypical environments impact gender participation in computer science. *Journal of Personality and Social Psychology*, *97*, 1045–1060.

Cohen, G., Ross, L., & Steele, C. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin*, *25*(10), 1302–1318.

Crocker, J., & Park, L. E. (2004). The costly pursuit of self-esteem. *Psychological Bulletin*, *130*(3), 392–414.

Crocker, J., Voelkl, K., Testa, M., & Major, B. (1991). Social stigma: The affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology*, *60*(2), 218–228. https://doi.org/10.1037/0022-3514.60.2.218

Cromley, J. G., Perez, T., & Kaplan, A. (2016). Undergraduate STEM achievement and retention: Cognitive, motivational, and institutional factors and solutions. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 4–11.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*, 438–481. https://doi.org/10.2307/1170281

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*, 105–115.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668.

Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(23), 5964–5969.

Duncan, T. G., & McKeachie, W. J. (2005). The making of motivated strategies for learning questionnaire. *Educational Psychologist*, *40*(2), 117–128.

Dweck, C. S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology*, *31*(4), 674–685.

Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PLOS ONE*, *11*(7), Article e0157447.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*, 700–717.

Hancock, D. R. (2002). Influencing graduate students' classroom achievement, homework habits and motivation to learn with verbal praise. *Educational Research*, *44*, 83–95.

Harackiewicz, J. M. (1979). The effects of reward contingency and performance feedback on intrinsic motivation. *Journal of Personality and Social Psychology*, *37*(8), 1352–1363.

Harber, K. D. (1998). Feedback to minorities: Evidence of a positive bias. *Journal of Personality and Social Psychology*, *74*(3), 622–628.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Heine, S., Kitayama, S., & Lehman, D. (2001). Cultural differences in self-evaluation: Japanese readily accept negative self-relevant information. *Journal of Cross-Cultural Psychology*, *32*, 434–443.

Henderlong, J., & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin*, *128*(5), 774–795.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*(5), 365–371.

Johnson, I. R., Pietri, E. S., Fullilove, F., & Mowrer, S. (2019). Exploring identity-safety cues and allyship among black women students in STEM environments. *Psychology of Women Quarterly*, *43*(2), 131–150.

Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology*, *35*(3), 835–847.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historic review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284.

Lent, R., Brown, S., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*, 79–122.

Major, B., Kunstman, J. W., Malta, B. D., Sawyer, P. J., Townsend, S. S. M., & Mendes, W. B. (2016). Suspicion of motives predicts minorities' responses to positive feedback in interracial interactions. *Journal of Experimental Social Psychology*, *62*, 75–88.

Mervis, J. (2011). Undergraduate science. Weed-out courses hamper diversity. *Science*, *334*(6061), 1333.

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology*, *75*(1), 33–52.

Murphy, M. C., Gopalan, M., Carter, E., Emerson, K., Bottoms, B., & Walton, G. M. (2020). A customized belonging intervention improves retention of socially disadvantaged students at a broad-access university. *Science Advances*, *6*(29), Article eaba4677.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*, 879–885.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math=male, me=female, therefore math not=me. *Journal of Personality and Social Psychology*, *83*, 44–59.

Park, L. E., Kondrak, C., Ward, D. E., & Streamer, L. (2018). Positive feedback from male authority figures boosts women's math outcomes. *Personality and Social Psychology Bulletin*, *44*, 359–383.

Park, L. E., Naidu, E., Lemay, E. P., Canning, E. A., Ward, D. E., Panlilio, Z. A., & Vessels, V. (2023). Social evaluative threat across individual, relational, and collective selves. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 68, pp. 139–199). Academic Press.

Park, L. E., O'Brien, C., Italiano, A., Ward, D. E., & Panlilio, Z. (2023). "That's a great question!" Instructors' positive responses to students' questions improves STEM-related outcomes. *Self and Identity*, *22*(6), 849–895.

Park, L. E., Ward, D. E., Naragon-Gainey, K., Canning, E. A., Koefler, N., Panlilio, Z., Vessels, V., & Pascuzzi, G. (under review). *Measuring perceived safety versus threat in the environment: The STEP scale*.

Pietri, E. S., Johnson, I. R., & Ozgumus, E. (2018). One size may not fit all: Exploring how the intersection of race and gender and stigma consciousness predict effective identity-safe cues for black women. *Journal of Experimental Social Psychology*, *74*, 291–306.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40.

Rattan, A., Good, C., & Dweck, C. S. (2012). "It's ok—Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology*, *48*(3), 731–737.

Roberts, S. O., & Rizzo, M. T. (2021). The psychology of American racism. *American Psychologist*, *76*(3), 475–487.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Westview Press.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, *100*, 255–270.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199–214.

Vogt, C. M. (2008). Faculty as a critical juncture in student retention and performance in engineering programs. *Journal of Engineering Education*, *97*, 27–36.

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

Yeager, D. S., Purdie-Vaughns, V., Garcia, J., Apfel, N., Brzustoski, P., Master, A., Hessert, W. T., Williams, M. E., & Cohen, G. L. (2014). Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide. *Journal of Experimental Psychology: General*, *143*(2), 804–824.