EXAMINING THE LEAKY PIPELINE: DO WOMEN AND MEN DIFFER IN THEIR ATTRIBUTIONS FOR SUCCESS AND FAILURE IN STEM CONTEXTS?

By

HAYLEY NICHOLLE SVENSSON

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

Shana Cole

And approved by

____

____

____

New Brunswick, New Jersey

January 2023

ABSTRACT OF THE THESIS

Examining the Leaky Pipeline: Do Women and Men Differ in Their Attributions for Success and Failure in STEM Contexts?

by HAYLEY NICHOLLE SVENSSON

Thesis Director:

Shana Cole

Women leave science, technology, engineering, and mathematics (STEM) fields at much higher rates than men. One reason for this gender disparity may be differences in the way women respond to STEM-related feedback relative to men. Some past research suggests women tend to make stable/internal attributions for their failures and unstable/external attributions for their successes, whereas men do the opposite. However, research in recent years has failed to replicate this finding that was once well-established in the literature. To reconcile these conflicting findings, in two studies I examined the relationship between gender and causal attributions for success and failure in STEM fields. In Study 1, undergraduate STEM students at Rutgers University read one of two hypothetical scenarios in which they either earned an A on or failed an exam in one of their STEM courses. In Study 2, participants completed an actual STEM test and received randomized success or failure feedback about their performance on the test. In both studies, participants then completed a series of measures capturing the reasons they may have succeeded or failed. Across both studies, there was little evidence for gender differences in the types of attributions STEM students made for success and failure. Although the

notion that women tend to make maladaptive attributions has been long held in psychology, it may not be as prevalent now as it once was.

Acknowledgement

First and foremost, I would like to thank my advisor and thesis committee chair, Shana Cole, for her role in this project. Shana, your unwavering guidance and support the past two and a half years have fostered my growth both as a researcher and person. You are an incredible role model, and I could not be more grateful to work with and know you.

I would also like to thank the members of my thesis committee, Laurie Rudman and Diana T. Sanchez, for their support, invaluable feedback, and expertise, which were essential to the execution of this project. Additionally, I would like to thank the research assistants in the Regulation, Action, and Motivated Perception (RAMP) Lab who assisted with various tasks related to this project.

To my parents, Suzanne and Mark, and sister, Aubrey, you are the reason I am where I am today. Thank you for your unending support throughout graduate school and beyond.

To my grandparents, Ernest, Janet, William, and Carolynne, your wisdom and guidance are always with me; I work hard every day to make you proud. Thank you also to my extended family and close friends who have encouraged me every step of the way.

Logan, thank you for the sacrifices you have made to support me as I pursue my dreams (and for previewing all my surveys). Your constant love and encouragement made all this possible.

Table of Contents

List of Tables

List of Figures

Examining the Leaky Pipeline: Do Women and Men Differ in their Attributions for Success and Failure in STEM Contexts?

Women are vastly underrepresented in many science, technology, engineering, and mathematics (STEM) fields. For example, women comprise a mere 15% of the engineering workforce and only 25% of computer-based occupations, a number which has actually decreased by 7% in the last 30 years (Fry et al., 2021). Moreover, women leave the STEM workforce at much higher rates than men. One recent study found that 50% of women abandoned their technology careers by the age of 35, a rate that is 45% higher than men's (Accenture, 2020). In light of these statistics, scholars have used the "leaky pipeline" metaphor to describe how women leave the math- and science-career path and ultimately become minorities in the field.

Prior research has found that gender disparities in STEM fields are due, in part, to unique challenges women (vs. men) encounter. From an early age, girls receive cues from parents, teachers, peers, and the media that dissuade them from pursuing interests in math and science (Saucerman & Vasquez, 2014). In one study conducted at a science museum with children as young as one to three years old, researchers found evidence that parents interacted with their sons and daughters differently when discussing science exhibits (Crowley et al., 2001). Although both boys and girls initiated an equivalent number of interactions with their parents, parents were three times as likely to provide a scientific explanation in their conversations with their sons than in their conversations with their daughters. Such findings suggest gender biases in scientific learning and literacy may occur years before children encounter science lessons in educational settings.

Unfortunately, biases toward women and girls interested in STEM are not ameliorated with age. Even when they overcome early obstacles to science literacy and become STEM majors, women frequently report threatening academic environments, gender biases, and a lack of perceived support as barriers to their STEM education (Casad et al., 2019; Eaton et al., 2020). Additionally, research suggests that as young as elementary school, people hold stereotypes that link science and math with masculinity (Archer et al., 2010; Cvencek et al., 2011; Hand et al., 2017). Such associations are related to women's self-concepts; identifying as a woman with a math-intensive major was associated with a slower tendency to associate math with the self because they implicitly associated math with men (Nosek et al., 2002). Moreover, gender-science stereotypes can influence important long-term outcomes; for women, stronger gender-science stereotypes were related to less identification with science and weaker science career aspirations (Cundiff et al., 2013). As is evident in prior research across the lifespan, women face myriad obstacles in STEM environments that perpetuate their underrepresentation in the field.

In addition to, and likely as a result of, educational barriers and pervasive gender-science stereotypes, some evidence suggests women respond to STEM-related setbacks differently than men. For example, recent research found that women who failed a difficult STEM pre-requisite course (i.e., Calculus I) were significantly less likely to graduate with a degree in STEM than men who failed the same course (Sanabria & Penner, 2017). In my own research examining the effects of perceived disruption due to the COVID-19 pandemic on math-related outcomes such as objective numeracy (i.e., math ability), interest in future math courses, and grades during the pandemic, high-ability women were more likely to disengage from further STEM courses when they faced unexpected hardships (Svensson et al., 2022). Gendered responses to STEM-related setbacks may be contributing to women leaving the STEM pipeline at higher rates and thus maintaining the STEM gender gap. Because setbacks and failures are often inevitable occurrences during the pursuit of difficult STEM goals, exploring women's responses to STEM-related feedback is important for understanding the cognitive precursors to STEM disengagement.

Causal Attributions

One reason why women may be more likely to disengage following negative STEM-related feedback is because of differences in the causal attributions they make for their performance. Causal attributions refer to the explanatory reasons people give for an outcome or a behavior. Most classic and contemporary research examining gender differences in causal attributions focuses on two dimensions: stability and locus of control (i.e., internality and externality; Weiner, 1974). Stability refers to the belief that an outcome or behavior is likely to occur again. Internality refers to the belief that an outcome is because of one's own abilities or traits, as opposed to factors outside of the self like the situation or environment. Much past research has found women/girls are likely to make maladaptive "self-derogatory" attributions in which they make stable/internal attributions for their failures and unstable/external attributions for their successes, whereas men/boys often do the opposite (e.g., Beyer, 1998; Frieze et al., 1982; Nelson & Cooper, 1997). These gender differences are especially likely to occur in male gender-typed domains, like math and science (Beyer, 1998). In addition to gender-science stereotypes that imply negative outcomes are attributable to stable group differences, this tendency can also be accredited to gender differences in general math (Louis & Mistele, 2012) and science (Cheryan et al., 2017) self-efficacy. Since female students have lower self-efficacy in these subjects, they may be more likely than their male peers to attribute their successes to situational and unstable causes (e.g., luck) and failures to dispositional and stable causes (e.g., lack of ability and/or competency; Campbell & Hackett, 1986). These maladaptive responses to STEM-related feedback may contribute to less persistence in the field. As such, shifting the way women construe their STEM-related successes and failures may be a valuable tool for increasing their retention in STEM fields.

However, in recent years, work examining gendered causal attributions has failed to replicate the finding that women (vs. men) make more stable attributions for failure and unstable attributions for success. In one recent study examining gender differences in undergraduate students' attributions for academic achievement, the researchers only found a significant gender difference in effort-related attributions and no differences in attributions related to task difficulty, ability, or luck (García y García, 2021). Similarly, other work examining the effects of an attributional retraining intervention on a variety of outcomes for students transitioning from high school to college found no relationships between gender and attributions related to effort, test difficulty and strategy, or the professor's quality of instruction (Hamm et al., 2014). Some of these discrepancies may be due to a lack of consensus in the literature regarding the measures that best capture the content of one's attributions. The evidence for gender differences has been mixed across a multitude of studies using different methods, including ones employing multiple-choice (e.g., the Causal Dimension Scale; e.g., Cortés-Suarez & Sandiford, 2008; Russell, 1982) as well as open-ended response measures (e.g., reflective essays; Vivian et al., 2013).

Given these conflicting findings, it is currently unclear whether there are gender differences in causal attributions for STEM outcomes. Many of the studies that initially established gender differences were conducted over 30 years ago. Since then, the United States has seen major advances in gender equality (Geiger & Parker, 2018), which may positively affect women's self-efficacy in STEM and the extent to which they attribute their successes to reasons outside of their control. To promote the continued societal progress of women, it is first important to determine whether their tendency to make maladaptive attributions has persisted.

Present Work

Across two studies, we examined gender differences in causal attributions for success and failure in STEM contexts. By examining gendered attributions in STEM, we are tapping into a domain in which women have been historically excluded (Fry et al., 2021), potentially affecting how they view their performance and the reasons they use to explain success and failure (Jasko et al., 2020). Based on the mixed findings in the literature, we tested two competing predictions for how gender may affect causal attributions. The *disparity hypothesis* predicts women will make more stable/internal attributions for failure and unstable/external attributions for success, whereas men will do the opposite. Conversely, *the similarity hypothesis* predicts there will be no gender differences in the types of attributions women and men make for success and failure. Finding support for the disparity hypothesis would point to shifting women's maladaptive attributions as a critical next step for improving women's STEM outcomes, however, we first need to establish evidence of gender differences.

Study 1

As a first test of the competing hypotheses, in Study 1, undergraduate male and female STEM students imagined they had a success or failure in one of their STEM courses and provided possible reasons for the outcome. Research assistants coded their attributions for whether they were internal/external and stable/unstable. Gender differences would suggest support for the disparity hypothesis, while no gender differences would suggest support for the similarity hypothesis. Lastly, participants completed several measures assessing their experiences in STEM (e.g., sense of belonging in STEM, STEM self-efficacy, etc.) so that we could examine 1) whether attributions were related to broader STEM outcomes and 2) whether there were gender differences in the students' STEM experiences.

Study 1 Method

Participants

Participants were 162 undergraduate STEM majors at Rutgers University who participated in an online study. I conducted an *a priori* power analysis using G*Power (Faul et al., 2009) which revealed a recommended sample size of $N = 199$ to detect a small to medium effect size ($f = 0.2$) for the predicted interaction effects. We collected data until we had saturated our available participants, leaving us with $N = 162$. The study took approximately 20 minutes to complete, and participants received a $5.00 Amazon gift card for their participation. To recruit participants, we sent information about the study as well as a screening survey to a listserv of Rutgers undergraduate STEM students.

In the screening survey, students indicated their gender identity, major, and whether they were an undergraduate student at Rutgers. Based on their responses, I sent the full study to individuals who were undergraduate students at Rutgers and had a major in a STEM field predominately comprised of men (e.g., chemistry and computer science; U.S. Department of Education, 2017). Of those who completed the screening survey and met the other eligibility criteria, 88.4% had a major in a STEM field with evidence of these disparities. Six participants were removed for failing more than one attention check, leaving $N = 156$ for analyses (age: $M_{\text{age}} = 19.94$, $SD_{\text{age}} = 1.84$; gender: 55.1% women, 44.2% men, 0.6% trans women; race: 67.3% Asian, 19.2% White or Caucasian, 6.4% Middle Eastern, 5.8% Hispanic or Latinx, 3.8% Black or African American, 2.6% indicated another race/ethnicity, and 1.3% preferred not to answer; cumulative grade-point average (GPA): $M_{\text{GPA}} = 3.60$, $SD_{\text{GPA}} = 0.36$).

Procedure

After the consent process, participants indicated their gender identity and then read a short hypothetical scenario in which they imagined that in one of their STEM courses they either got a failing grade ($n = 81$) or an A ($n = 75$) on an exam. They were then prompted to list three possible causes for the outcome. Next, they completed measures of their sense of belonging in STEM, feelings toward STEM, STEM aspirations and career motivations, expectations for success in STEM, STEM identity, and STEM self-efficacy. Finally, participants completed relevant demographic items (e.g., age and race/ethnicity), speculated about the purpose of the study in an open-ended question, were thanked for their participation, and debriefed.

Measures

The following measures appear in the same order as they did for participants (see Appendix A for all questions and response scales).

Gender identity

Participants were first asked to select their gender identity from the following options: woman, man, trans woman, trans man, genderqueer/non-binary, intersex, and another. The "another" option had a corresponding textbox where participants could specify their gender identity.

Success- and failure-scenario manipulation

Participants were randomly assigned to imagine a brief hypothetical scenario in which they either got an A (success condition) or a failing grade (failure condition) on an exam in one of their STEM courses. Those in the success condition read, "for the next set of questions, please imagine that in one of your STEM courses, you got an A on an exam that you studied for." Those in the failure condition read, "for the next set of questions, please imagine that in one of your STEM courses, you got a failing grade on an exam that you studied for." Participants' condition served as the primary manipulated independent variable in analyses.

Attributions for success or failure

Participants then listed three possible causes for the success or failure outcome. A trained, independent coder unaware of the conditions or hypotheses analyzed participants' responses to all three reasons listed. The coder analyzed each response for the extent to which they were 1) stable vs. unstable and 2) internal vs. external. The coder assigned each response a "1" if the attribution appeared to be stable (i.e., a similar outcome is *likely* to happen again in the future) (e.g., "test-related anxiety and stress") or a "0" if the attribution was unstable (i.e., a similar outcome is *unlikely* to happen again in the future) (e.g., "grading error"). Similarly, the coder assigned each response a "1" if the attribution appeared to be internal (i.e., believing an outcome is because of their own abilities or traits) (e.g., "I knew the content well) or a "0" if the attribution was external (i.e., believing an outcome is caused by factors outside of the self, like the situation or environment) (e.g., "there was a curve"). The numerically coded responses were then summed such that each participant had a stability and internality score from zero to three with higher values indicating a tendency to make more stable or internal attributions, respectively, in response to the hypothetical scenarios. These two scores were the primary dependent variables used in analyses.

STEM belongingness

Participants responded to six statements assessing their perceived belongingness in STEM (Leibowitz et al., 2020). On five-point scales from 1 (*strongly disagree)* to 5 (*strongly agree*), participants respond to statements such as, "I feel a sense of belonging in my major." This measure had high internal reliability (Cronbach's α = .80) and the items were averaged together with higher values indicating more belonging in STEM.

Feelings toward STEM

Participants then indicated their feelings toward STEM fields on four, seven-point scales ranging from 1 (i.e., *bad, sad, avoid, afraid*) to 7 (i.e., *good, happy, approach, unafraid*). This measure had high internal reliability (Cronbach's α = .81) and the items were averaged together with higher values indicating more positive feelings toward STEM.

STEM career aspirations/motivations

Participants responded to four statements about their future STEM career aspirations and motivations (Moss-Racusin et al., 2018; Stake & Mares, 2001; Starr, 2018). Using five-point scales from 1 (*not at all true)* to 5 (*very true*), participants responded to statements such as, “I would enjoy a career in STEM.” This measure had high internal reliability (Cronbach’s α = .85) and the items were averaged together with higher values indicating stronger STEM career aspirations/motivations.

Expectations for STEM success

Expectations for success in STEM were measured with seven items on five-point scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) (Appianing & Van Eck, 2018). A sample item is, “I don’t think I will succeed in a STEM field.” This measure had high internal reliability (Cronbach’s α = .89) and the items were averaged together with higher values indicating expectations for greater success.

STEM identity

Participants responded to 12 statements assessing their identification with STEM (Jones et al., 2013; Leaper et al., 2012; Smith et al., 2015; Starr, 2018). On five-point scales (1 = *strongly disagree*, 5 = *strongly agree*), participants responded to statements such as, “Being good at STEM is an important part of who I am.” This measure had high internal reliability (Cronbach’s α = .86) and the items were averaged together with higher values indicating stronger identification.

STEM self-efficacy

Lastly, participants responded to eight statements measuring their self-efficacy in STEM using a five-point scale (1 = *strongly disagree*, 5 = *strongly agree*) (Leibowitz et al., 2020). A sample item is, “I will be able to achieve most of the STEM-related goals that I have set for myself.” This measure had high internal reliability (Cronbach’s α = .93) and the items were averaged together with higher values indicating stronger STEM self-efficacy.

Study 1 Results

Descriptive statistics for STEM experience measures and *t*-tests examining gender differences

I first sought to examine whether there were gender differences in any of the STEM-experience measures. Table A1 (see Appendix) includes descriptive statistics for each measure by gender as well as for the entire sample. I used independent samples *t*-tests to test for gender differences, with a positive *t* statistic indicating men (vs. women) scored higher on the respective measure.

There were significant gender differences for two of the STEM measures: STEM belongingness, $t(154) = 2.05$, $p = .043$, and feelings toward STEM, $t(154) = 1.98$, $p = .049$, such that men reported more STEM belongingness and more positive feelings toward STEM than women. There were no significant gender differences amongst the other STEM measures.

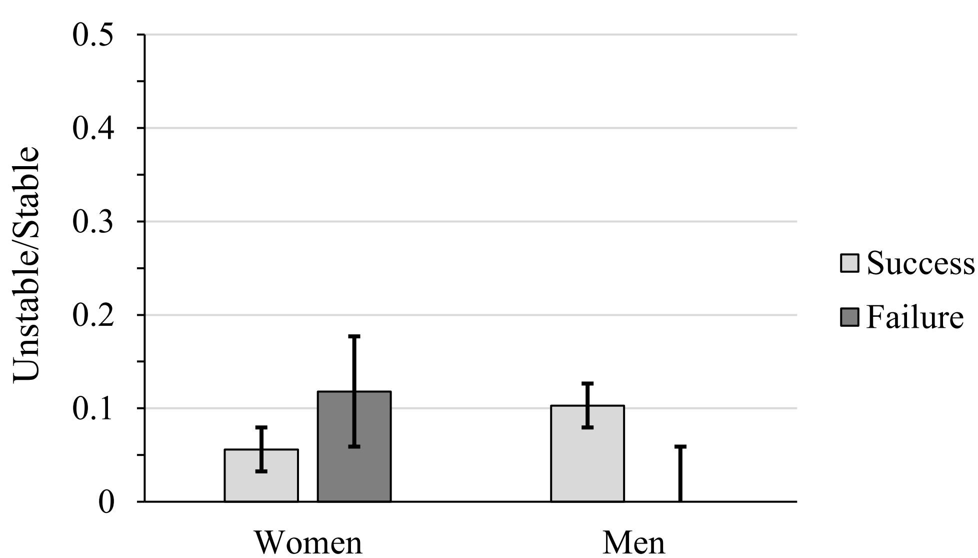
Test of competing hypotheses: Condition by gender predicting attribution type

To test my competing hypotheses that either 1) women (vs. men) will make more stable/internal attributions for failure and unstable/external attributions for success or 2) there will be no gender differences, I conducted a 2 (success vs. failure condition) × 2 (women vs. men) analysis of variance (ANOVA) examining the effects of the feedback manipulation, gender, and the interaction of the two on the attributions.

I first tested this using the stability attributions as a dependent variable. As seen in Figure 1, there were no significant main effects of gender ($F(1, 152) = 0.66$, $p = .417$) or condition ($F(1, 152) = 0.22$, $p = .642$) on the stability of the attributions. However, there was a marginally significant condition by gender interaction predicting the stability of the attributions, $F(1, 152) = 3.60$, $p = .060$. Although the effect was marginally significant, we further probed the interaction to see where differences emerged. Simple effects analyses revealed that gender differences between women and men in the failure condition were primarily driving the interaction, $F(1, 152) = 3.69$, $p = .057$. There were no attributional differences between women and men in the success condition ($F(1, 152) = 0.58$, $p = .446$), and there were no differences between women across conditions ($F(1, 152) = 1.15$, $p = .286$) or between men across conditions ($F(1, 152) = 2.52$, $p = .115$). Although marginal, the interaction partially supports the disparity hypothesis because there were gender differences that emerged, particularly in the failure condition.

Figure 1

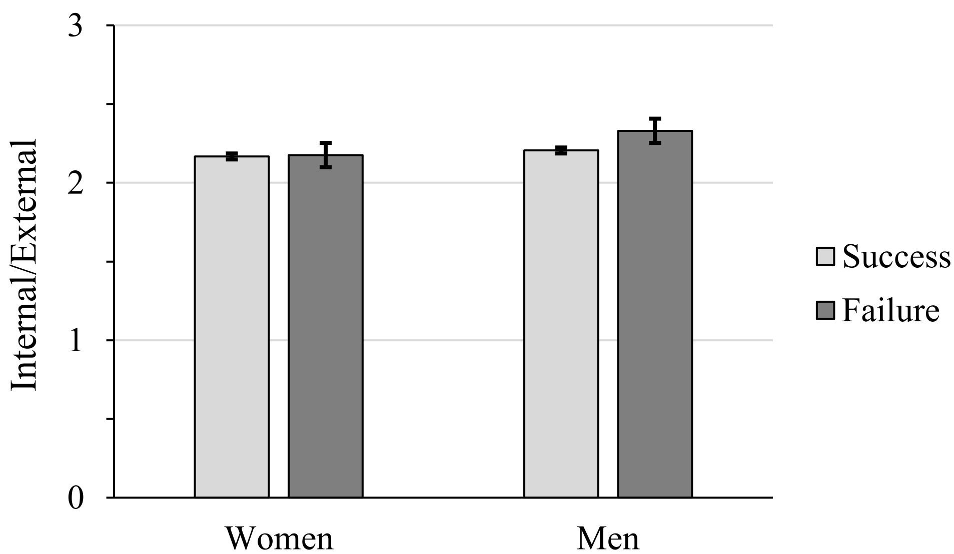
The Interaction of Condition and Gender Predicting the Stability of the Attributions



I next conducted a 2 (success vs. failure condition) × 2 (women vs. men) analysis of variance (ANOVA) examining the effects of the feedback manipulation, gender, and the interaction of the two on the internality of the attributions. As seen in Figure 2, there were no significant main effects of gender ($F(1, 152) = 0.53$, $p = .469$) or condition ($F(1, 152) = 0.26$, $p = .609$), nor a significant condition by gender interaction ($F(1, 152) = 0.19$, $p = .660$) predicting the internality of the attributions. Given that there were no gender differences, this finding supports the similarity hypothesis.

Figure 2

The Interaction of Condition and Gender Predicting the Internality of the Attributions



Relationships between STEM experiences and attributions

Finally, I conducted correlation analyses to see whether attribution type was associated with STEM experiences. I conducted separate analyses for the success and failure conditions. In Table 1 below, the correlations below the diagonal correspond to those in the success condition and those above the diagonal correspond to those in the failure condition. Only two significant correlations emerged. Participants with greater STEM self-efficacy, $r(75) = .23$, $p = .042$, and stronger STEM career aspirations/motivations, $r(75) = .28$, $p = .014$, made more internal attributions for their successes. There were no significant correlations between the STEM experiences and attributions for those in the failure condition.

Table 1

Correlations Between all Study Variables

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1. Stability of attributions		.04	.16	-.03	-.07	-.04	.04	-.14
2. Internality of attributions	-.12		.06	.18	.00	.03	.09	.05
3. STEM belongingness	.10	-.11		**.56****	**.46****	**.55****	**.69****	**.45****
4. Feelings toward STEM	.17	.17	**.30****		**.42****	**.60****	**.50****	**.60****
5. STEM career aspirations/motivations	-.04	**.28***	**.45****	**.39****		**.54****	**.46****	**.40****
6. Expectations for STEM success	-.14	.21	**.50****	**.36****	**.64****		**.53****	**.72****
7. STEM identity	.14	.12	**.44****	**.51****	**.66****	**.50****		**.54****
8. STEM self-efficacy	.01	**.23***	**.55****	**.47****	**.51****	**.80****	**.52****	

Note. All variables are continuous. * Correlation is significant at the .05 level. ** Correlation is significant at the .01 level.

Exploratory analyses: STEM experiences as moderators of condition and attribution type

Finally, it is possible that individual differences in STEM experiences moderate the relationship between condition and attributions. I tested each of the STEM measures as a moderator using PROCESS (Hayes, 2018). There were no main effects of condition, no main effects of any of the STEM measures, and no significant interactions of condition and the STEM measures on the stability of the attributions (see Table A2).

Study 2

In Study 1, we found some mixed evidence for whether women and men differ in their attributions. However, Study 1 involved a hypothetical scenario with very minimal information. Thus, in Study 2, we sought to examine participants' attributions following actual success or failure feedback. Undergraduate STEM students took a STEM test and received randomized feedback that they had either performed well or poorly. We measured the attributions they made for their performance. Again, we sought evidence for either disparity or similarity between men and women's attributions.

Pilot Work

Prior to conducting Study 2, I ran two pilot studies to develop and test 1) a novel STEM test and 2) the success- and failure-feedback manipulation.

STEM test pilot. In the STEM test pilot study, undergraduate STEM students at Rutgers ($N = 78$) completed a random subset of approximately eight (out of a possible 20) STEM-relevant problems adapted from prior research or online Intelligence Quotient (IQ) tests. After completing the problems, they provided their thoughts about each question they completed as well as the test overall. For example, participants indicated how they felt while completing the test (e.g., excited, stressed, etc.), how well they thought they performed compared to other undergraduate STEM students at Rutgers, and the extent to which each question was a good indicator of their STEM-related abilities.

On average, participants got 67% of the questions correct. In Table A3 (see Appendix), I give descriptive information for each question. For use in Study 2, I selected the 11 problems that participants rated as being most indicative of their STEM-related abilities and chose a mix of questions that varied in difficulty, based on the students' objective scores.

Feedback pilot. I then conducted a second pilot study to test the effectiveness of the success and failure feedback. Participants (*N* = 98 undergraduate STEM students recruited from Prolific) first completed the 11-problem STEM ability test and then received randomized success (scored in the 83rd percentile) or failure (scored in the 17th percentile) feedback about their performance on the test. Participants then answered a series of questions about the feedback they received, including how they felt about the feedback and how well they believe they performed on the test. Students in the success (vs. failure) condition felt significantly more positive about the feedback ($p < .001$) and indicated performing significantly better on the test ($p < .001$). Since these results were consistent with my expectations, we implemented this feedback manipulation in Study 2.

Study 2 Method

Participants

Participants were 179 undergraduate STEM students recruited via Prolific and CloudResearch to complete an online study. I conducted an *a priori* power analysis using G*Power (Faul et al., 2009) which revealed a recommended sample size of *N* = 199 to detect a small to medium effect size ($f = 0.2$) for the predicted interaction effects. We collected data until we had saturated our available participants, leaving us with *N* = 179. The study took approximately 25 minutes to complete, and participants received $4.38 for their participation. As in Study 1, we specifically focused our recruitment on students with majors in STEM fields predominately comprised of men (e.g., chemistry and computer science; U.S. Department of Education, 2017). Three participants were removed for failing more than one attention check, leaving $N = 176$ for analyses (age: $M_{age} = 23.83$, $SD_{age} = 6.16$; gender: 59.7% men, 34.7% women, 4.5% genderqueer/non-binary, 0.6% trans women, 0.6% trans men; race: 57.4% White or Caucasian, 22.2% Asian, 15.9% Black or African American, 9.1% Hispanic or Latinx, 0.6% Native American or American Indian, 0.6% Native Hawaiian or Pacific Islander, 0.6% Middle Eastern or North African, 0.6% indicated another race/ethnicity, and 1.1% preferred not to answer; cumulative grade-point average (GPA): $M_{GPA} = 3.43$, $SD_{GPA} = 0.46$).

Procedure

After the consent process, participants learned they would take a test designed to measure their STEM abilities. They read that prior research had found that those who performed well on this test were more successful in their STEM-related endeavors. They then completed the STEM test which consisted of 11 problem-solving questions related to math and science. After completing the test, participants learned that the survey program needed a few minutes to grade their tests (the tests were not actually graded) and, during that time, they completed demographic items (age, gender identity, racial/ethnic identity, etc.). Next, participants were randomly assigned to receive either success ($n = 86$) or failure ($n = 90$) feedback. They then described possible reasons (i.e., causal attributions) for the success or failure and responded to a series of Likert-style questions about these reasons. Finally, participants received a thorough debriefing and learned that the score they received was not related to their actual performance on the STEM test and that the test was developed by the researchers. Lastly, they were thanked and compensated for their participation.

Measures

The following measures appear in the same order as they did for participants (see Appendix B for all questions and response scales).

STEM ability test

Participants first completed an 11-question test that I developed, described as being

indicative of their STEM-related abilities. This test consisted of questions related to identifying

numeric (e.g., Campbell & Hackett, 1986) and non-numeric (e.g., Raven, 2000) patterns, graph

literacy (e.g., adapted from the Educational Testing Service, 2022), standard math such as algebra and geometry (e.g., the Educational Testing Service, 2021), spatial ability (e.g., Peters & Battista, 2008), and scientific reasoning (e.g., Drummond & Fischhoff, 2017). Sample questions appear in Appendix B. We scored each question for correctness and combined them for a total score out of 13 (two of the 11 questions had two parts) (Cronbach's α = .72).

Feedback manipulation

Participants were randomly assigned to receive either success or failure feedback that

they were told was related to their performance on the STEM ability test. Those in the success condition read, "Based on the amount of time you took to complete the test and the accuracy of your responses, you scored in the 83rd percentile. That means you performed better than 83% of participants who have taken the test. Your performance places you in the top 20% of STEM undergraduates and is well above average." Those in the failure condition read, "Based on the amount of time you took to complete the test and the accuracy of your responses, you scored in the 17th percentile. That means you performed better than 17% of participants who have taken the test. Your performance places you in the bottom 20% of STEM undergraduates and is well below average." Participants' condition served as the primary manipulated independent variable in analyses.

Attributions for success or failure

In a brief open-ended question, participants wrote about one possible reason for why they may have scored the way they did on the STEM ability test. They then answered two questions assessing the stability and internality of the reason they gave (adapted from Campbell & Henry, 1999; Peterson et al., 1982). For stability, the question was on a scale from 1 (*not at all likely*) to 5 (*extremely likely*) and said, "When taking STEM assessments in the future, how likely is it that this cause will again affect your performance?" For internality, the statement participants responded to was on a scale from 1 (*has nothing to do with me*) to 5 (*completely has to do with me*) and said, "Please indicate the extent to which this reason has to do with you. For example, a characteristic you have, something you did or was caused by you, etc." Participants then wrote about one more possible reason and answered the same questions about the new reason. I analyzed these Likert-type questions as four separate dependent variables.

Two independent coders unaware of condition and hypotheses also coded these reasons for the extent that they were stable vs. unstable. However, there were relatively high levels of disagreement between the two coders such that, for each reason, over 10%

of the coded responses differed. Because of this, I re-coded the open-ended responses based on guidelines from Vivian et al., 2013. Consistent with the coding in Study 1, I assigned each response a "1" if the attribution appeared to be stable (e.g., "I'm not very smart") or a "0" if the attribution was unstable (e.g., "I did not use scratch paper") and, in a separate variable, a "1" if the attribution was internal (e.g., "I put in 100% effort") or a "0" if the attribution was external (e.g., "luck"). I then used these numerically coded responses as four separate dependent variables (i.e., one for the first reason and one for the second reason for both stability and internality) in analyses.

Lastly, participants indicated on a scale from 1 (*not at all*) to 5 (*a great deal*) the extent to which 18 additional attributions explained their score on the STEM test (e.g., "I am a good test taker" and "I did not put enough effort into the test"; see Appendix B for full list of attributions) (adapted from Marsh et al., 1984; Ryckman et al., 1990).

Study 2 Results

STEM test performance by gender

First, I investigated whether there were any gender differences in scores on the STEM

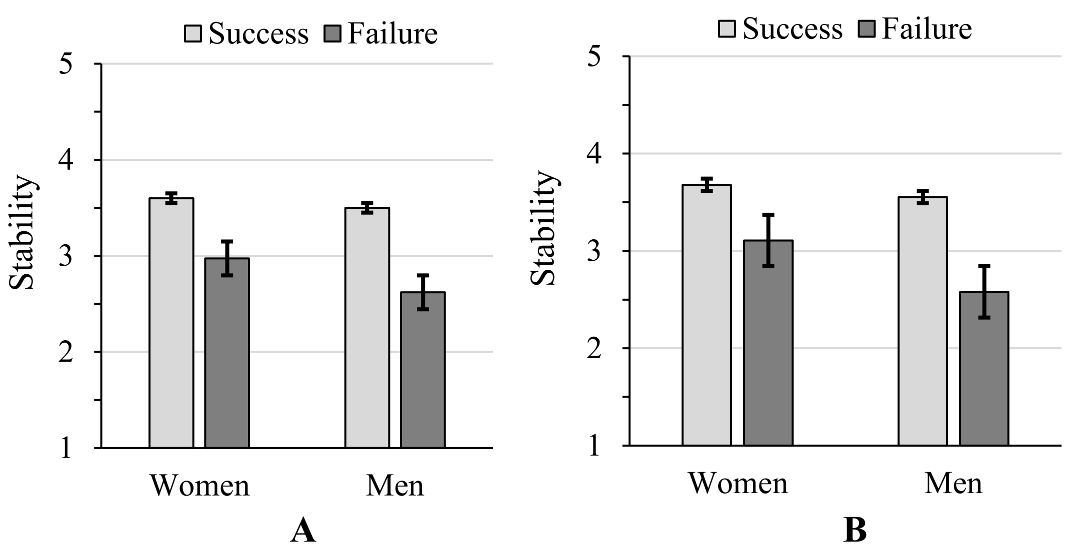
test. As expected, women (M = 6.19, SD = 2.67) and men (M = 6.19, SD = 2.89) scored similarly on the STEM test, $t(166) = -0.01$, $p = .991$.

Test of competing hypotheses: Condition and gender predicting attribution type

Self-Report Attribution Ratings. As in Study 1, to test the similarity vs. disparity hypotheses, I conducted 2 (success vs. failure condition) × 2 (women vs. men) analyses of variance (ANOVAs) examining the effects of the feedback manipulation, gender, and the interaction of the two on each attribution. In the first set of tests, the dependent variable was participants' self-report of the stability of their attributions. The results were consistent across both attributions participants listed. There were significant main effects of condition such that those in the success (vs. failure) condition made significantly more stable attributions for both the first, $F(1, 164) = 12.38$, $p < .001$ (Figure 3A), and second, $F(1, 164) = 14.20$, $p < .001$ (Figure 3B), attributions. However, there were no main effects of gender (ps > .100) or interactions of condition and gender (ps > .300) on the self-reported stability of either attribution.

Figure 3

*The Interaction of Condition and Gender Predicting the Self-Reported Stability of the First (**A**) and Second (**B**) Attributions*

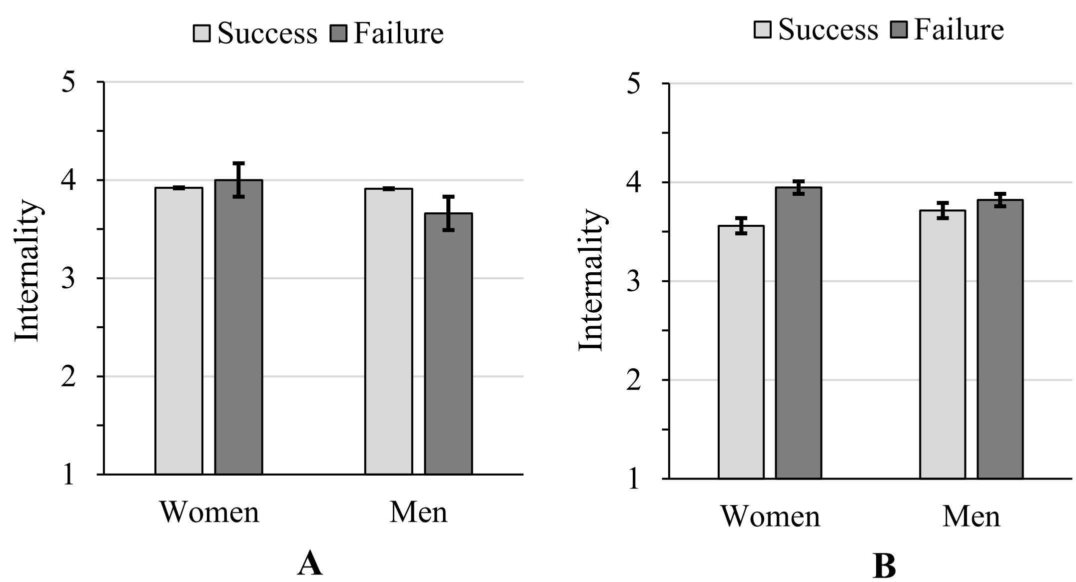


I next conducted a 2 (success vs. failure condition) × 2 (women vs. men) analysis of variance (ANOVA) examining the effects of the feedback manipulation, gender, and the interaction of the two on the self-reported internality of each attribution (Figure 4). The results were consistent across both attributions. There were no significant main effects of condition (first attribution: $F(1, 164) = 0.21$, $p = .645$; second attribution: $F(1,$

164) = 1.36, *p* = .246), gender (first attribution: *F*(1, 164) = 0.89, *p* = .346; second attribution: *F*(1, 164) = 0.01, *p* = .947), or interactions of condition and gender (first attribution: *F*(1, 164) = 0.80, *p* = .372; second attribution: *F*(1, 164) = 0.44, *p* = .508) on the self-reported internality of the attributions.

Figure 4

*The Interaction of Condition and Gender Predicting the Self-Reported Internality of the First (**A**) and Second (**B**) Attributions*



Coded Attributions. In the second set of tests, I used the coded stability of the attributions[1]. Across both attributions, there were no significant main effects of condition (first attribution: *F*(1, 157) = 0.01, *p* = .912; second attribution: *F*(1, 153) = 1.67, *p* = .198), gender (first attribution: *F*(1, 157) = 0.80, *p* = .374; second attribution: *F*(1, 153) = 0.14, *p* = .705), or condition by gender interactions (first attribution: *F*(1, 157) = 0.61, *p* =

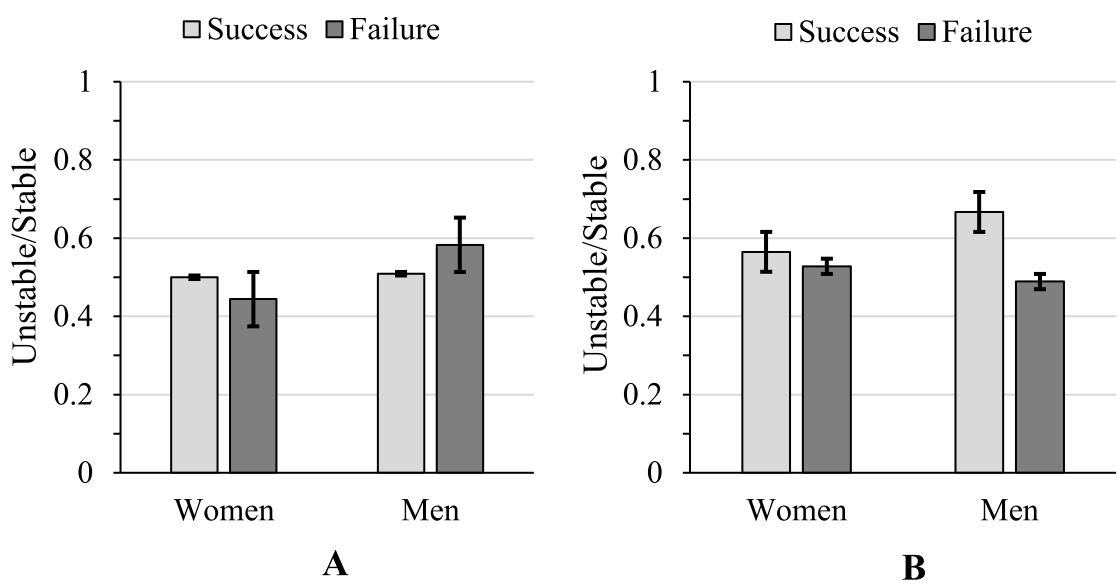
---

[1] Since we only assessed two (vs. three) attributions in Study 2 (vs. Study 1), we chose to examine each attribution as separate dependent variables. However, for both stability and internality, the results are similar when the coded attributions are summed.

.438; second attribution: $F(1, 153) = 0.71$, $p = .401$) when the coded stability of the attributions were the dependent variables (Figure 5).

Figure 5

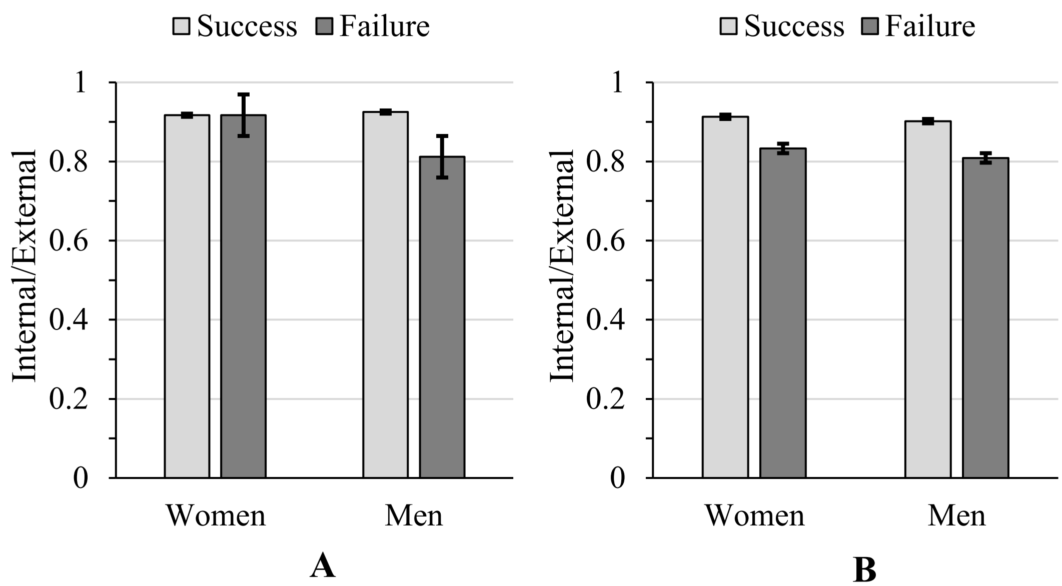
*The Interaction of Condition and Gender Predicting the Coded Stability of the First (**A**) and Second (**B**) Attributions*



I also tested the coded internality of the attributions and found similar results. Across both attributions, there were no significant main effects of condition (first attribution: $F(1, 157) = 1.16$, $p = .284$; second attribution: $F(1, 153) = 2.20$, $p = .140$), gender (first attribution: $F(1, 157) = 0.86$, $p = .356$; second attribution: $F(1, 153) = 0.10$, $p = .759$), or condition by gender interactions (first attribution: $F(1, 157) = 1.16$, $p = .284$; second attribution: $F(1, 153) = 0.01$, $p = .907$) when the coded internality of the attributions were the dependent variables (Figure 6).

Figure 6

*The Interaction of Condition and Gender Predicting the Coded Internality of the First (**A**) and Second (**B**) Attributions*



Exploratory analyses: Condition and gender predicting additional attributions

Lastly, I conducted 2 (success vs. failure condition) × 2 (women vs. men) analyses of variance (ANOVAs) examining the effects of the feedback manipulation, gender, and the interaction of the two on the extent to which participants agreed that several additional attributions (e.g., "I am a good test taker" and "I did not put enough effort into the test") were responsible for their performance on the STEM test. Across all 18 attributions, 13 of them produced a significant main effect of condition, but there were no significant main effects of gender (ps > .050) or condition by gender interactions (ps > .100; see Table A4 for condition main effects).

Discussion

Women are significantly underrepresented in STEM fields (Fry et al., 2021), and researchers have long tried to explain, and ultimately prevent, women's increased STEM attrition rates. In the present work, I investigated a possible cognitive precursor to women's disengagement from STEM: their attributional styles. Specifically, I tested whether women make less adaptive causal attributions for success and failure than do men. In Study 1, I found mixed support for the disparity hypothesis such that women made marginally more stable attributions for failure and unstable attributions for success, whereas men did the opposite. There were no gender differences in internal/external attributions. In Study 2, I found support for the similarity hypothesis as there were no gender differences in the stability or internality of the attributions for success and failure. Thus, across the studies, the results were largely in favor of the similarity hypothesis.

Interestingly, in Study 1, the marginally significant interaction was driven by gender differences in the failure condition with women making more stable attributions for their failures than men. Although I predicted gender differences in attributions for both successes and failures, a disparity hypothesis could also predict that there would only be gender differences in the failure attributions. If women are performing well, they may respond to successes similarly than men, however, they may respond to setbacks more negatively. This idea is supported by prior research that investigated gender differences in STEM persistence among undergraduate students who either passed or failed a difficult STEM pre-requisite course (Sanabria & Penner, 2017). The researchers found that women (vs. men) who failed the course were significantly more likely to disengage from STEM but did not find differences in how women and men responded to passing the course. Future research may consider focusing specifically on STEM failure responses as a contributor to the gender disparities in STEM.

Interpreting the Findings of the Present Studies

The mixed results could, in part, be explained by general study logistics, like our decision to run the studies online as opposed to in-person. Particularly for Study 2, conducting the study online, along with the general low-stakes nature of the STEM test, may have affected how much effort participants put into completing it. There is some support for this idea in prior research that has found in the context of low-stakes research situations, participants' motivation to complete an Intelligence Quotient (IQ) Test to the best of their ability deviated substantially from the maximum possible amount (Duckworth et al., 2011). Additionally, the study being conducted online may have affected the extent to which participants believed the STEM test was actually being graded or the extent to which they felt invested in or impacted by the feedback. Thus, these studies may not reflect how people respond in real high-stakes situations (e.g., real grades in their STEM courses). Future research could endeavor to collect students' attributions following successes and failures in their actual STEM courses.

Our results may have also been affected by certain characteristics of our participant samples in both Studies 1 and 2. First, participants' specific STEM major may have affected their success and failure responses. Although we focused our recruitment on STEM fields predominantly comprised of men, there are still variations in the magnitude of these gender disparities across STEM fields (Fry et al., 2021; see Table A6 for participant breakdowns by field of study). Additionally, in Study 1, our sample predominantly identified as Asian (67.3%). Prior research has found that Asian individuals are often stereotyped as having exceptional academic abilities, particularly in math and science (Ruble & Zhang, 2013; Steen, 1987). These stereotypes may subsequently affect how Asian-identifying participants interpreted and responded to the success and failure feedback (Shih et al., 2002). Future research should address this by examining women's responses to STEM-related feedback through an intersectional lens, considering both race and gender.

In Study 2, our sample was comprised primarily of students who were further along in their undergraduate educations (70% of the sample was in their third year or higher). Throughout the process of becoming more established in their STEM careers, these students needed to be successful in a number of difficult STEM pre-requisite courses and likely received positive feedback on more than one occasion. As a result of the successes they likely already experienced throughout their STEM educations, they may not have taken the failure feedback as seriously as a younger student who, in comparison, may have had fewer opportunities to receive positive feedback in STEM contexts. To examine this idea further, I conducted an ANCOVA with gender, condition, and the interaction of the two, as well as participants' year in school as a covariate, predicting the eight dependent variables I analyzed in Study 2 (i.e., self-reported stability, self-reported internality, coded stability, and coded internality for the first and second attributions participants provided). When year in school was added to the model, there was a significant condition by gender interaction predicting the coded stability of participants' first attribution (p = .044; see Table A5 for full analysis). This interaction was not in the predicted direction, but nonetheless demonstrates year in school may possibly contribute to STEM students' feedback responses.

Another possible reason for our results lie in some of the ambiguities that exist in the attribution literature. As previously mentioned, there is a lack of consensus in the literature regarding the measures that best capture the content of one's attributions and there have been varying results across studies employing multiple-choice (e.g., the Causal Dimension Scale; e.g., Russell, 1982; Cortés-Suarez & Sandiford, 2008) as well as open-ended response measures (e.g., reflective essays; Vivian et al., 2013). Relatedly, when attributions are assessed using open-ended questions and essays, there is little guidance in the literature for how to best interpret and subsequently code these responses. Some of these discrepancies were also apparent in the present work. For example, our measures for self-reported stability and internality in Study 2 were not consistently highly correlated with their corresponding coded responses. This shows that how participants interpret their own attributions may differ from the interpretations of trained coders. These major shortcomings within the attribution literature severely limit current researchers from making sense of their own results in the context of the existing literature.

It is also important to consider how advances in gender equality in the United States may affect the types of attributions women make for their STEM successes and failures. Many of the studies that initially established gender differences in causal attributions were conducted over 30 years ago. Since then, the United States has seen major advances in gender equality (Geiger & Parker, 2018), which may positively affect women's self-efficacy in STEM and the extent to which they attribute their successes to reasons outside of their control. Now, there are a number of organizations like "Women in STEM" and "She can STEM" that seek to close the STEM gender gap by providing resources and education to girls and women who are interested in pursuing STEM careers. Additionally, both within and outside these organizations, women and girls now have more access to successful exemplars, which can be critical to helping women navigate environments where they have traditionally been excluded (Downing et al., 2005; Eby et al., 2008). Given the substantial increase in resources for women interested in STEM careers along with the larger societal shifts, women may be less likely to make maladaptive attributions than they once were.

Future Research

Although we did not find gender differences in cognitive responses, gender differences may emerge in other responses to STEM feedback, such as affective or behavioral responses. Indeed, I found some support for gender differences in affective responses in the feedback pilot study. We asked participants about the feelings they were experiencing after receiving either the success or failure feedback. After receiving the failure feedback, women (vs. men) reported feeling significantly more stressed (p = .033), disappointed (p = .049), and sad (p = .006), and marginally more frustrated (p = .052). Instead of differences in cognitive attributions, there may be gender differences in earlier stage affective responses to feedback that affect STEM persistence. Moreover, past research suggests men and women have different behavioral responses to STEM setbacks. For example, men are more likely to retake college-level STEM courses after failing them than women (Penner & Willer, 2019) and are more likely to go on to graduate with a degree in STEM after failing a difficult STEM pre-requisite course (Sanabria & Penner, 2017). Similarly, in prior work I conducted, I found that women high (vs. low) in math ability were more likely to disengage from further math courses when they faced unexpected pandemic-related hardships (Svensson et al., 2022). In an upcoming study, we will measure affective and behavioral responses to feedback to gain valuable insight into other ways women and men may differ in their STEM feedback responses, thus providing us with more clarity on interventions that may be effective in bridging any gaps.

Conclusion

Overall, our findings provide more support for the similarity (vs. disparity) hypothesis such that women and men did not differ in the attributions they made for STEM successes and failures. As such, the present work may give us hope that women's attributions for success and failure may no longer be a hindrance to their progress in STEM. However, more work is still needed to test whether gender differences emerge in other types of responses to STEM hardships. Moving forward, it is critical that we continue to investigate individual-level reasons for the gender disparities that still pervade STEM fields and consider new ways in which we can intervene should differences arise.

References

Accenture. (2020). Resetting tech culture: Five strategies to keep women in tech. Retrieved September 9, 2021 from https://www.accenture.com/_acnmedia/PDF-134/Accenture-A4-GWC-Report-Final1.pdf

Appianing, J., & Van Eck, R. N. (2018). Development and validation of the Value-Expectancy STEM Assessment Scale for students in higher education. *International Journal of STEM Education, 5*(1), 1-16.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). “Doing” science versus “being” a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity. *Science Education*, *94*, 617-639.

Beyer, S. (1998). Gender differences in causal attributions by college students of performance on course examinations. *Current Psychology, 17*(4), 346-358.

Campbell, N. K., & Hackett, G. (1986). The effects of mathematics task performance on math self-efficacy and task interest. *Journal of Vocational Behavior, 28*(2), 149-162.

Campbell, C. R., & Henry, J. W. (1999). Gender differences in self-attributions: Relationship of gender to attributional consistency, style, and expectations for performance in a college course. *Sex Roles, 41*(1), 95-104.

Casad, B. J., Petzel, Z. W., & Ingalls, E. A. (2019). A model of threatening academic environments predicts women STEM majors’ self-esteem and engagement in STEM. *Sex Roles, 80*(7), 469-488.

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

Cortés-Suarez, G., & Sandiford, J. R. (2008). Causal attributions for success or failure of students in college algebra. *Community College Journal of Research and Practice, 32*(4-6), 325-346.

Crowley, K., Callanan, M. A., Tenenbaum, H. R., & Allen, E. (2001). Parents explain more often to boys than to girls during shared scientific thinking. *Psychological Science, 12*, 258-261.

Cundiff, J. L., Vescio, T. K., Loken, E., & Lo, L. (2013). Do gender–science stereotypes predict science identification and science career aspirations among undergraduate science majors?. *Social Psychology of Education*, *16*(4), 541-554.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, *82*, 766-779.

Downing, R. A., Crosby, F. J., & Blake-Beard, S. (2005). The perceived importance of developmental relationships on women undergraduates’ pursuit of science. *Psychology of Women Quarterly, 29*(4), 419–426. https://doi.org/10.1111/j.1471-6402.2005. 00242.x

Drummond, C., & Fischhoff, B. (2017). Development and validation of the scientific reasoning scale. *Journal of Behavioral Decision Making, 30*(1), 26-38.

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences*, *108*(19), 7716-7720.

Eaton, A. A., Saunders, J. F., Jacobson, R. K., & West, K. (2020). How gender and race stereotypes impact the advancement of scholars in STEM: Professors’ biased evaluations of physics and biology post-doctoral candidates. *Sex Roles, 82*(3), 127-141.

Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., & DuBois, D. L. (2008). Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of Vocational Behavior, 72*(2), 254–267. https://doi.org/ 10.1016/j.jvb.2007.04.005

Educational Testing Service. (2022). *GRE General Test: Quantitative Reasoning Sample Questions*. https://www.ets.org/gre/institutions/about/general/quantitative_reasoning_sample_questions/

Educational Testing Service. (2021). *Math Review*. https://www.ets.org/s/gre/pdf/gre_math_review.pdf

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160. doi:10.3758/BRM.41.4.1149

Frieze, I. H., Whitley, B. E., Hanusa, B. H., & McHugh, M. C. (1982). Assessing the theoretical models for sex differences in causal attributions for success and failure. *Sex Roles, 8*(4), 333-343.

Fry, R., Kennedy, B., & Funk, C. (2021). STEM jobs see uneven progress in increasing gender, racial and ethnic diversity: Higher education pipeline suggests long path ahead for increasing diversity, especially in fields like computing and engineering. *Pew Research Center*.

Geiger, A. W., & Parker, K. (2018). For women's history month, a look at gender gains – and gaps – in the U.S. *Pew Research Center*. Retrieved August 27, 2022, from http://pewrsr.ch/2HDZtxX

García y García, B. E. (2021). To what factors do university students attribute their academic success?. *Journal on Efficiency and Responsibility in Education and Science, 14*(1), 1-8.

Hamm, J. M., Perry, R. P., Clifton, R. A., Chipperfield, J. G., & Boese, G. D. (2014). Attributional retraining: A motivation treatment with differential psychosocial and performance benefits for failure prone individuals in competitive achievement settings. *Basic and Applied Social Psychology*, *36*(3), 221-237. https://doi.org/10.1080/01973533.2014.890623

Hand, S., Rice, L., & Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Social Psychology of Education*, *20*, 929-945.

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs, 85*(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100

Jasko, K., Pyrkosz-Pacyna, J., Czarnek, G., Dukała, K., & Szastok, M. (2020). The STEM graduate: Immediately after graduation, men and women already differ in job outcomes, attributions for success, and desired job characteristics. *Journal of Social Issues, 76*(3), 512-542.

Jones, B. D., Ruff, C., & Paretti, M. C. (2013). The impact of engineering identification and stereotypes on undergraduate women's achievement and persistence in engineering. *Social Psychology of Education, 16*(3), 471-493.

Leaper, C., Farkas, T., & Brown, C. S. (2012). Adolescent girls' experiences and gender-related beliefs in relation to their motivation in math/science and English. *Journal of Youth and Adolescence*, *41*(3), 268-282.

Leibowitz, J. B., Lovitt, C. F., & Seager, C. S. (2020). Development and validation of a survey to assess belonging, academic engagement, and self-efficacy in STEM RLCs. *Learning Communities: Research & Practice*, *8*(1), 3.

Louis, R. A., & Mistele, J. M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. *International Journal of Science and Mathematics Education, 10*(5), 1163-1190.

Marsh, H. W., Cairns, L., Relich, J., Barnes, J., & Debus, R. L. (1984). The relationship between dimensions of self-attribution and dimensions of self-concept. *Journal of Educational Psychology, 76*(1), 3.

Moss-Racusin, C. A., Sanzari, C., Caluori, N., & Rabasco, H. (2018). Gender bias produces gender gaps in STEM engagement. *Sex Roles*, *79*(11), 651-670.

Nelson, L. J., & Cooper, J. (1997). Gender differences in children's reactions to success and failure with computers. *Computers in Human Behavior, 13*(2), 247-267.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math= male, me= female, therefore math≠ me. *Journal of Personality and Social Psychology*, *83*(1), 44-59.

Penner, A. M., & Willer, R. (2019). Men's overpersistence and the gender gap in science and mathematics. *Socius, 5*, 2378023118821836.

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition, 66*(3), 260-264.

Peterson, C., Semmel, A., Von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. (1982). The attributional style questionnaire. *Cognitive Therapy and Research, 6*(3), 287-299.

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology, 41*(1), 1-48.

Ruble, R. A., & Zhang, Y. B. (2013). Stereotypes of Chinese international students held by Americans. *International Journal of Intercultural Relations, 37*(2), 202-211.

Russell, D. (1982). The Causal Dimension Scale: A measure of how individuals perceive causes. *Journal of Personality and Social Psychology, 42*(6), 1137.

Ryckman, D. B., Peckham, P. D., Mizokawa, D. T., & Sprague, D. G. (1990). The survey of achievement responsibility (SOAR): Reliability and validity data on an academic attribution scale. *Journal of Personality Assessment, 54*(1-2), 265-275.

Sanabria, T., & Penner, A. (2017). Weeded out? Gendered responses to failing calculus. *Social Sciences, 6*(2), 47.

Saucerman, J., & Vasquez, K. (2014). Psychological barriers to STEM participation for women over the course of development. *Adultspan Journal, 13*(1), 46-64.

Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology*, 83: 638–647.

Smith, J. L., Brown, E. R., Thoman, D. B., & Deemer, E. D. (2015). Losing its expected communal value: How stereotype threat undermines women's identity as research scientists. *Social Psychology of Education, 18*(3), 443-466.

Stake, J. E., & Mares, K. R. (2001). Science enrichment programs for gifted high school girls and boys: Predictors of program impact on science confidence and motivation. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 38*(10), 1065-1088.

Starr, C. R. (2018). "I'm not a science nerd!" STEM stereotypes, identity, and motivation among undergraduate women. *Psychology of Women Quarterly, 42*(4), 489-503.

Steen, L. A. (1987). Mathematics education: A predictor of scientific competitiveness. *Science, 237*(4812), 251-302.

Svensson, H., Shoots-Reinhard, B., Cravens-Brown, L., & Peters, E. (2022). Greater objective numeracy protects COVID-19 pandemic grades but endangers academic interest. *Scholarship of Teaching and Learning in Psychology.* Advance online publication. https://doi.org/10.1037/stl0000319

U.S. Department of Education. (2017). Gender distribution of bachelor's degrees in the humanities. *American Academy of Arts & Sciences.* Retrieved from https://www.amacad.org/humanities-indicators/higher-education/gender-distribution-bachelors-degrees-humanities

Vivian, R., Falkner, K., & Falkner, N. (2013, November). Computer science students' causal attributions for successful and unsuccessful outcomes in programming assignments. In *Proceedings of the 13th Koli Calling International Conference on Computing Education Research* (pp. 125-134).

Weiner, B. (Ed.). (1974). Achievement motivation and attribution theory. *General Learning Press*.

Appendix

Appendix A: Study 1 questions and response scales

Gender identity

What is your gender identity?

Woman, man, trans woman, trans man, genderqueer/non-binary, intersex, another (with open-ended response option)

Success- and failure-scenario manipulation

Success scenario:

For the next set of questions, please imagine that in **one of your STEM courses, you got an A on an exam that you studied for**.

Failure scenario:

For the next set of questions, please imagine that in **one of your STEM courses, you got a failing grade on an exam that you studied for**.

Attributions for success or failure

In the boxes below, please list three possible causes for this outcome.

Possible Cause #1: ____

Possible Cause #2: ____

Possible Cause #3: ____

STEM belongingness

Please respond to the following statements about your experiences in your major.

1 (*Strongly disagree*), 2 (*Somewhat disagree*), 3 (*Neither agree nor disagree*), 4 (*Somewhat agree*), 5 (*Strongly agree*)

1. I feel a sense of belonging to my major.
2. I feel that I am a member of my major's community.
3. I feel comfortable in my major's community.
4. If given the opportunity, I would choose my major again.
5. My major's community is supportive of me.
6. I struggle to feel like I belong in my major.

Feelings toward STEM

Please describe your feelings toward STEM on the following scales:

-3 (*bad, sad, avoid, afraid*) – +3 (*good, happy, approach, unafraid*)

STEM career aspirations/motivations

Indicate the extent to which the following statements represent your future plans.

1 (*Not at all true*), 2 (*Somewhat true*), 3 (*Moderately true*), 4 (*Mostly true*), 5 (*Very true*)

1. I would enjoy a career in STEM.
2. I have good feelings about a career in STEM.
3. Having a STEM career would be interesting.
4. I would like to have a career in STEM.

Expectations for STEM success

Indicate the extent to which you agree or disagree with the following statements.

1 (*Strongly disagree*), 2 (*Somewhat disagree*), 3 (*Neither agree nor disagree*), 4 (*Somewhat agree*), 5 (*Strongly agree*)

1. I don't think I will succeed in a STEM field.
2. I don't think I can make an impact if I take on a STEM-related job.

3. I would certainly feel useless in a STEM-related job.
4. I feel I have what it takes to succeed in a STEM-related job.
5. I would be able to succeed in a STEM field as well as most other people.
6. I do not think I can achieve anything meaningful as a STEM professional.
7. I feel I have a number of good qualities to be successful in a STEM field.

STEM identity

Please indicate the extent o which you agree or disagree with the following statements.

1 (*Strongly disagree*), 2 (*Somewhat disagree*), 3 (*Neither agree nor disagree*), 4 (*Somewhat agree*), 5 (*Strongly agree*)

1. Being good at STEM is an important part of who I am.
2. Doing well on STEM tasks is very important to me.
3. Success in my STEM field is very important/valuable to me.
4. It matters to me how well I perform in my STEM field.
5. I feel like I'm just like people who are good at STEM.
6. I feel that the things I like to do in my spare time are similar to what most STEM-oriented people like to do in their spare time.
7. I see myself as a STEM student.
8. I am pleased to be a STEM student.
9. I feel strong ties with other STEM students.
10. I identify with other STEM students.
11. I feel that being a STEM student is an important reflection of who I am.
12. I don't act like the typical STEM student.

STEM self-efficacy

Indicate the extent to which you agree or disagree with the following statements about yourself.

1 (*Strongly disagree*), 2 (*Somewhat disagree*), 3 (*Neither agree nor disagree*), 4 (*Somewhat agree*), 5 (*Strongly agree*)

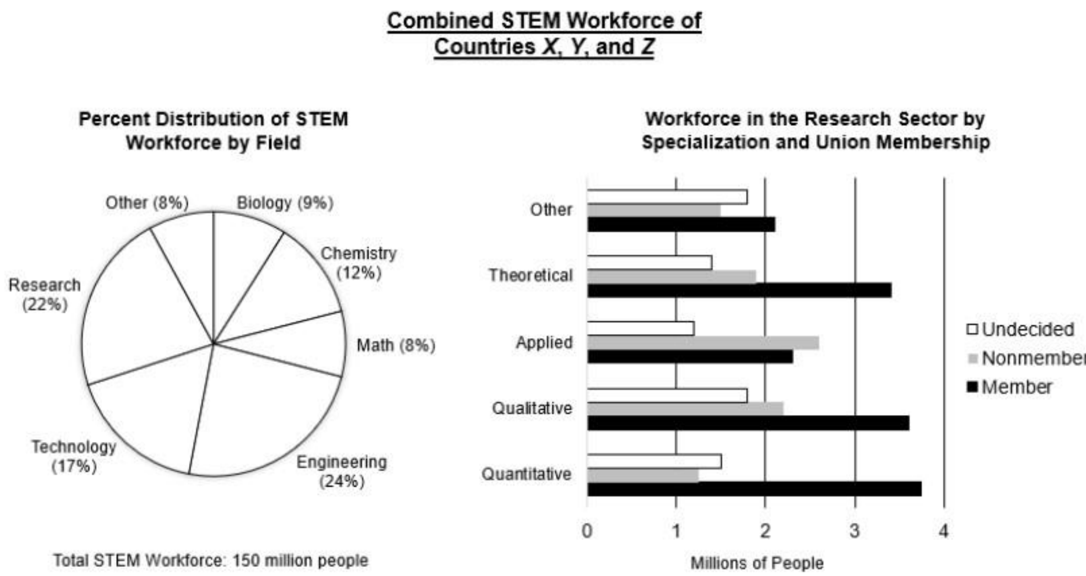
1. I will be able to achieve most of the STEM-related goals that I have set for myself.
2. When facing difficult STEM-related tasks, I am certain that I will accomplish them.
3. In general, I think that I can obtain STEM-related outcomes that are important to me.
4. I believe I can succeed at any STEM-related endeavor to which I set my mind.
5. I will be able to successfully overcome many STEM-related challenges.
6. I am confident that I can perform effectively on many different STEM-related tasks.
7. Compared to other people, I can do most STEM-related tasks very well.
8. Even when things are tough, I can perform quite well in STEM.

Appendix B: Study 2 questions and response scales

STEM ability test

Below are four examples of questions used in the STEM ability test (two difficult and two easy).

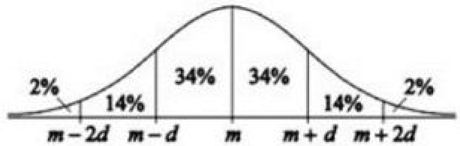
Two difficult questions:



In the combined STEM workforce of three countries, the ratio of the number of union members to the number of nonmembers is the same for the chemistry sector as it is for the quantitative specialization of the research sector. One-third of the chemistry sector is undecided in their union membership. Which of the following is closest to the number of nonmembers in the chemistry sector?

2.4 million **2.9 million** 3.9 million 10.3 million 10.8 million

The figure below shows the graph of a normal distribution with mean *m* and standard deviation *d*, including approximate percentages of the distribution corresponding to the six regions below.



Suppose the heights of a population of 3,000 adult penguins are approximately normally distributed with a mean of 65 centimeters and a standard deviation of 5 centimeters.

If an adult penguin is chosen at random from the population, approximately what is the probability that the penguin's height will be less than 60 centimeters? <u>Please round your answer to 2 decimal places.</u>

Correct Answer: **0.16**

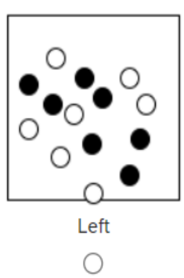
<u>Two easy questions:</u>

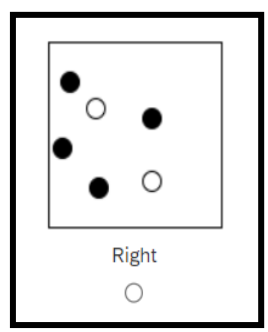
What is the next number in the sequence?

15, 60, 63, 252, 255, ______

Correct Answer: **1,020**

Please indicate which box (left or right has the higher probability of picking a black ball.





Feedback manipulation

Success feedback:

STEM Ability Test Score

Based on the amount of time you took to complete the task and the accuracy of your responses, you scored in the **83** percentile.

That means you performed better than **83%** of participants who have taken the test.

Your performance places you in the **top 20%** of STEM undergraduates and is **well above average**.

Failure feedback:

STEM Ability Test Score

Based on the amount of time you took to complete the task and the accuracy of your responses, you scored in the **17** percentile.

That means you performed better than **17%** of participants who have taken the test.

Your performance places you in the **bottom 20%** of STEM undergraduates and is **well below average**.

Attributions for success or failure

1. In box below, please reflect on the score you just received. What are the reasons you may have scored the way you did? (*Open-ended*)

Causal Attribution #1:

2. What do you believe is the biggest reason for why you scored the way you did? (*Open-ended*)
3. Please indicate the extent to which this reason **has to do with you**. For example, a characteristic you have, something you did or was caused by you, etc. 1 (*Has* ***nothing*** *to do with me*) – 5 (***Completely*** *has to do with me*)
4. When taking STEM assessments in the future, how likely is it that this cause will again affect your performance? 1 (*Not at all likely*) – 5 (*Extremely likely*)

Causal Attribution #2:

5. In the box below, please briefly describe one more main reason for why you scored the way you did. (*Open-ended*)
6. Please indicate the extent to which this reason **has to do with you**. For example, a characteristic you have, something you did or was caused by you, etc. 1 (*Has **nothing** to do with me*) – 5 (***Completely** has to do with me*)
7. When taking STEM assessments in the future, how likely is it that this cause will again affect your performance? 1 (*Not at all likely*) – 5 (*Extremely likely*)

Lastly, please indicate the extent to which you believe the following reasons explain your score on the test. 1 (*Not at all*), 2 (*A little*), 3 (*A moderate amount*), 4 (*A lot*), 5 (*A great deal*)

The test was difficult.	I was unlucky.	I am a good test taker.	Something was distracting me.	It mattered to me how well I performed.	I am smarter than other STEM students.
The test was easy.	I put a lot of effort into the test.	I am a bad test taker.	My STEM ability is strong.	It didn't matter to me how well I performed.	I am not smarter than other STEM students.
I was lucky.	I did not put enough effort into the test.	I was focused.	My STEM ability is weak.	Societal expectations about my STEM ability.	I didn't feel well/wasn't in a good mood.

Appendix C: Supplemental analyses

Table A1

Descriptive Statistics for STEM Experience Measures and t-tests Examining Gender Differences (Study 1)

	All Participants (N = 156)		**Women** (n = 87)	**Men** (n = 69)	**Gender *t* tests**	***p* value**
	M (*SD*)	Range	*M* (*SD*)	*M* (*SD*)	*t*	*p*
STEM belongingness	3.68 (0.79)	1.83 – 5.00	3.56 (0.77)	3.82 (0.80)	**2.05**	**.043**
Feelings toward STEM	5.22 (1.25)	1.50 – 7.00	5.04 (1.19)	5.44 (1.29)	**1.98**	**.049**
STEM career aspirations/ motivations	4.37 (0.71)	1.50 – 5.00	4.31 (0.71)	4.44 (0.70)	1.13	.261
Expectations for STEM success	4.23 (0.69)	1.71 – 5.00	4.17 (0.73)	4.30 (0.63)	1.19	.236
STEM identity	3.80 (0.61)	1.67 – 5.00	3.81 (0.56)	3.78 (0.68)	–0.32	.750
STEM self-efficacy	3.97 (0.70)	1.00 – 5.00	3.89 (0.71)	4.06 (0.69)	1.48	.140

Note. Significant *t* statistics and corresponding *p*-values are bolded.

Table A2

STEM Measures Moderating the Relationship Between Condition and the Stability of the Attributions (Study 1)

	Condition		**STEM Measure**		**Condition × STEM Measure**	
	b (se)	*p*	*b (se)*	*p*	*b (se)*	*p*
STEM belongingness	0.49 (3.61)	.891	0.77 (0.55)	.162	–0.15 (0.88)	.869
Feelings toward STEM	–4.02 (3.25)	.216	–0.08 (0.33)	.805	0.74 (0.57)	.192
STEM career aspirations/ motivations	–0.78 (3.35)	.817	–0.38 (0.57)	.504	0.20 (0.78)	.798
Expectations for STEM success	2.06 (3.24)	.524	–0.19 (0.54)	.730	–0.47 (0.78)	.547
STEM identity	–2.67 (4.46)	.549	0.30 (0.75)	.686	0.68 (1.11)	.539
STEM self-efficacy	–2.59 (3.17)	.413	–0.64 (0.51)	.210	0.70 (0.80)	.382

Note. Unstandardized *b*-coefficients with standard errors in parentheses and *p*-values are reported. Condition: 1 = success, 0 = failure. All other variables were continuous.

Table A3

Descriptive Statistics for Each Piloted STEM Test Question (STEM Test Pilot)

Question	**% of Sample who Answered Correctly**	**STEM Indicator** (0 (*Not at all a good measure*) – 100 (*An extremely good measure*))
Data Analysis (ETS)	**14.30%**	**64.54 (26.40)**
Objective Numeracy #2	**25.00%**	**62.35 (21.34)**
Graph Literacy #2	**34.60%**	**61.54 (28.43)**
SMAP (807)	35.70%	24.96 (23.28)
Graph Literacy #1	50.00%	55.46 (27.60)
Raven's #1	**53.80%**	**51.81 (30.06)**
Raven's #2	64.30%	34.21 (25.58)
Algebra (Pt. 1/2) (ETS)	**65.20%**	**73.29 (21.17)**
Algebra (Pt. 2/2) (ETS)	**69.60%**	**73.29 (21.17)**
Scientific Reasoning #2	**70.80%**	**58.91 (19.19)**
Geometry (Pt. 2/2) (ETS)	**72.00%**	**65.50 (27.03)**
SMAP (122)	73.10%	35.42 (26.05)
Scientific Reasoning #1	73.10%	58.88 (26.38)
Arithmetic (ETS)	**75.00%**	**61.54 (27.11)**
Spatial Ability #1	78.60%	42.50 (24.73)
Spatial Ability #2 (Dice)	**79.20%**	**56.75 (25.87)**
Number Series #1	**84.60%**	**55.85 (30.55)**
Geometry (Pt. 1/2) (ETS)	**84.60%**	**65.50 (27.03)**
Ratio Bias #1	91.70%	48.04 (25.00)
Ratio Bias #2	**92.30%**	**52.46 (31.68)**
Number Series #2	95.80%	54.22 (26.49)
Objective Numeracy #1	100.00%	43.96 (31.78)

Note. For STEM indicator, means and standard deviations in parentheses are reported.

The questions we chose for Study 2 are **bolded**.

Table A4

Condition Effects for the 18 Additional Attributions Measured in Study 2

The test was difficult.	**I was unlucky.**	**I am a good test taker.**	**Something was distracting me.**	It mattered to me how well I performed.	I am smarter than other STEM students.
p = .926	$\boldsymbol{p}$ **= .011** *Failure*	$\boldsymbol{p}$ **= .006** *Success*	$\boldsymbol{p}$ **= .014** *Failure*	p = .212	p = .468
The test was easy.	**I put a lot of effort into the test.**	**I am a bad test taker.**	**My STEM ability is strong.**	It didn't matter to me how well I performed.	**I am not smarter than other STEM students.**
$\boldsymbol{p}$ **= .010** *Success*	$\boldsymbol{p}$ **= .001** *Success*	$\boldsymbol{p}$ **< .001** *Failure*	$\boldsymbol{p}$ **= .002** *Success*	p = .089	$\boldsymbol{p}$ **= .007** *Failure*
I was lucky.	**I did not put enough effort into the test.**	**I was focused.**	**My STEM ability is weak.**	Societal expectations about my STEM ability.	**I didn't feel well/wasn't in a good mood.**
$\boldsymbol{p}$ **< .001** *Success*	$\boldsymbol{p}$ **< .001** *Failure*	$\boldsymbol{p}$ **< .001** *Success*	$\boldsymbol{p}$ **= .007** *Failure*	p = .084	$\boldsymbol{p}$ **= .032** *Failure*

Note. The condition effects are reported below each corresponding attribution. For attributions with significant condition effects (i.e., $\boldsymbol{p}$ **< .05**), the *condition* (success vs. failure) that scored higher on the corresponding attribution is included below the p-value.

Table A5

ANCOVA with Condition, Gender, their Interaction, and Covariates (Year in School, STEM Test Score, and Cumulative GPA) Predicting Study 2 Dependent Variables

Covariate Name	Dependent Variable	Covariate	Condition	Gender	Condition × Gender
Year in School	Cause #1: Self-Report Stability	.853	**<.001**	.368	.720
	Cause #2: Self-Report Stability	.107	**<.001**	.219	.285
	Cause #1: Self-Report Internality	.153	.479	.563	.251
	Cause #2: Self-Report Internality	.750	.703	.908	.834
	Cause #1: Coded Stability	.160	.262	.416	**.044**
	Cause #2: Coded Stability	.814	**.025**	.837	.691
	Cause #1: Coded Internality	.511	.329	.583	.352
	Cause #2: Coded Internality	.186	.168	.941	.902
STEM Test Score	Cause #1: Self-Report Stability	.882	**<.001**	.407	.468
	Cause #2: Self-Report Stability	.279	**<.001**	.156	.368
	Cause #1: Self-Report Internality	**.015**	.725	.380	.188
	Cause #2: Self-Report Internality	.302	.390	.769	.662
	Cause #1: Coded Stability	.343	.994	.424	.314
	Cause #2: Coded Stability	.079	.169	.739	.444
	Cause #1: Coded Internality	.500	.294	.344	.291
	Cause #2: Coded Internality	.360	.130	.740	.939
Cumulative GPA	Cause #1: Self-Report Stability	**.024**	**<.001**	.263	.418
	Cause #2: Self-Report Stability	**.032**	**<.001**	.096	.307
	Cause #1: Self-Report Internality	.756	.787	.407	.159
	Cause #2: Self-Report Internality	.281	.438	.689	.607
	Cause #1: Coded Stability	.909	.970	.423	.337
	Cause #2: Coded Stability	.905	.198	.719	.401
	Cause #1: Coded Internality	.910	.305	.350	.276
	Cause #2: Coded Internality	.230	.185	.878	.933

Note. For each dependent and predictor variable, the corresponding *p*-value is reported. Condition: 1 = success, 0 = failure; gender: 1 = woman; 0 = man; coded stability for both causes: 1 = stable, 0 = unstable; coded internality for both causes: 1 = internal, 0 = external. All other variables were continuous. Significant *p*-values (i.e., $p < .05$) are **bolded**.

Table A6

Participant Breakdown by STEM Field of Study for Studies 1 and 2

STEM Field	**% of Sample who Selected Field**	
	Study 1	Study 2
Aerospace Engineering	1.9	0.6
Applied Mathematics	0.0	1.8
Biochemistry	0.6	0.0
Biological Sciences	**16.0**	7.1
Biology	12.2	**16.0**
Biomedical Engineering	2.6	0.0
Biomedical Sciences	0.0	1.8
Biotechnology	0.6	0.0
Cell Biology & Neuroscience	**12.8**	0.6
Chemical Engineering	0.6	0.6
Chemistry	1.3	3.6
Computer Engineering	3.8	1.8
Computer Science	**24.4**	**26.6**
Computing (IT)	0.0	8.9
Earth Sciences	0.0	6.5
Economics	0.0	4.7
Electrical and Computer Engineering	3.8	0.0
Electrical Engineering	1.3	0.6
Engineering	0.0	**10.7**
Exercise Science	3.8	0.0
Genetics	1.9	0.0
Materials Science	0.0	0.6
Mathematics	2.6	3.0
Mechanical Engineering	1.9	1.8
Microbiology	1.9	0.6
Molecular Biology and Biochemistry	3.2	0.0
Neuroscience	0.0	0.6
Packaging Engineering	1.3	0.0
Physics	0.0	1.8
Statistics	0.6	0.0

Note. The three most frequently indicated STEM fields for each study sample are **bolded**.