12-8-2017

Gender and Domain Differences in College Students' Responses to Success and Failure

Elizabeth K. Lawner
University of Connecticut - Storrs, elizabeth.lawner@uconn.edu

Gender and Domain Differences in College Students' Responses to Success and Failure

Elizabeth K. Lawner, Ph.D.

University of Connecticut, 2017

Four studies examined perceptions of STEM and non-STEM college courses, gender and domain differences in responses to success and failure, and whether interventions to make failure seem normative could ameliorate negative responses to failure, particularly among women. Study 1 found that college students perceive STEM courses as more difficult than non-STEM courses, and believe that introductory STEM courses are used to weed students out of those fields. Moreover, the difference in perceptions of STEM and non-STEM courses was larger for women, particularly those who were not majoring in STEM fields in which women are most underrepresented. Study 2 piloted a novel task that was used in Studies 3 and 4. These studies did not support the idea that women are more likely than men to attribute their successes to effort and their failures to ability. However, there was some evidence that women have lower performance expectations, are less likely to believe they can succeed in STEM following failure, and are less willing to take on a challenge following success. Several of these gender differences were stronger among individuals with a fixed mindset and those who were not majoring in the STEM fields in which women are most underrepresented. In addition, Study 4 found that the normative interventions tested were not effective at promoting more resilient responses to failure among women. Together, these findings suggest that women's lower confidence in their abilities, particularly in STEM, combined with a general perception of STEM courses as more difficult and the experiences of failure embedded in STEM, may contribute to women's underrepresentation in STEM, especially in engineering, computer science, and physics.

Gender and Domain Differences in College Students' Responses to Success and Failure

Elizabeth K. Lawner

B.A., Duke University, 2010

M.A., University of Connecticut, 2014

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Gender and Domain Differences in College Students' Responses to Success and Failure

Presented by

Elizabeth K. Lawner, B.A., M.A.

Major Advisor ____

Diane M. Quinn, Ph.D.

Associate Advisor____

Blair T. Johnson, Ph.D.

Associate Advisor____

Blanca Rincón, Ph.D.

University of Connecticut

2017

Acknowledgements

My deepest appreciation goes to my advisor, Diane Quinn. She has provided tremendous guidance, encouragement, and support throughout my graduate career. I would also like to thank the members of my committee, Blair Johnson and Blanca Rincón, whose feedback on my proposal made my project stronger, as well as my readers, Nairan Ramirez Esparza and Amit Savkar, for their helpful suggestions. Thanks to the University of Connecticut's the Persistence of Women in STEM committee, Daniel Burkey and Kevin McLaughlin of the school of engineering, various professors in the mathematics, physics, and computer science department, and several STEM-related student groups who all assisted with recruiting participants for my engineering sample. I want to acknowledge the funding I received from the University of Connecticut Graduate School's Doctoral Dissertation Fellowship, which allowed me to pay participants. I would also like to thank Harriet Mosatche, whose dissertation on developmental differences in performance attributions informed and inspired my work. Finally, special thanks to my family—my parents, Harriet and Ivan, my brother, Robert, and my husband, Jonathan—for all their support and encouragement.

Table of Contents

Chapter 1

Women's Underrepresentation in Science, Technology, Engineering, and Math

Women now make up the majority of college students in the United States, earning 57 percent of the bachelor's degrees awarded in 2014 (National Science Foundation [NSF], 2017). However, when it comes to degrees in science, technology, engineering, and math (STEM), women earned 50 percent of the bachelor's degrees. That is to say, women are less likely than men to major in STEM fields. Even looking forward a cohort to the 2014 freshmen, there is a gender gap in STEM aspirations, with 49 percent of men, but only 38 percent of women intending to major in STEM fields. Furthermore, the gap increases with the level of degree: women earned 46 percent of the STEM master's degrees and 42 percent of the PhDs in STEM in 2014, despite representing 60 percent of the master's and 50 percent of the PhDs overall. Even once women have earned a degree in a STEM major, they are less likely than men to be employed in a STEM field. Instead 85 percent of employed women scientists and engineers work in STEM-related or non-STEM fields, while only 66 percent of employed men scientists and engineers do the same (NSF, 2017).

This decrease in the number of women in STEM from bachelor's degrees to graduate degrees and into the workforce demonstrates one part of what is referred to as the leaky pipeline. This name comes from the fact that there are many "leaks" in the pipeline that brings young girls from early interests in science and math all the way to leadership positions in STEM careers. Some of these leaks occur early on, as evidenced by the gender gap in freshmen's intentions to major in STEM (NSF, 2017) and AP test-taking rates in science and math (Hill, Corbett, & St. Rose, 2010). However, these leaks continue throughout the lifespan and women's career paths, underscoring the need to address issues regarding both recruitment and retention.

According to NSF (2017), STEM majors include agricultural sciences; biological sciences; computer sciences; earth, atmospheric, and ocean sciences; mathematics and statistics; physical sciences; psychology; social sciences; and engineering. Health fields are not part of the NSF definition of STEM, and instead health occupations are considered STEM-related occupations. However, not everyone uses the NSF definition. Health fields are sometimes considered part of STEM, and psychology and social sciences are sometimes left out. Women's representation in STEM varies significantly by field, and it is often the fields in which women are well represented that are not always considered to be part of STEM. For example, women earned 77 percent of the bachelor's degrees in psychology and 55 percent of the bachelor's degrees in social sciences in 2014 (NSF, 2017). On the other end of the spectrum, women earned just 18 percent of the bachelor's degrees in computer science and 20 percent of the bachelor's degrees in engineering in 2014 (NSF, 2017). Importantly, these are the STEM fields with the most jobs and best return on investment (Corbett & Hill, 2015). Furthermore, the proportion of women in computer science has actually decreased over the past 20 years; 29 percent of the bachelor's degrees in computer science in 1995 went to women (NSF, 2017). Physics is another area where women are quite underrepresented, earning just 19 percent of the bachelor's degrees in 2014.

In a 2009 review, Ceci, Williams, and Barnett examined several potential causes of women's underrepresentation in STEM. The review found little evidence for biological differences in ability. In fact, if the number of women in math-intensive careers was based solely on their math abilities, women's representation would double, thus, demonstrating that ability cannot be the primary cause of women's underrepresentation in STEM. The review concluded that four main factors contribute to women's underrepresentation in STEM careers. First, more

men than women receive the highest scores on the quantitative sections of gatekeeper tests such as the SAT and GRE. This gender difference at the top end of scores on the GRE-Q is problematic because it leads men to be admitted to graduate programs in STEM fields at greater rates than women. Second, women who do excel in math are disproportionately likely to also have superb verbal skills, compared to men who are similarly skilled in math. This gives women with high math competence greater career options outside of STEM, should they not desire a career in STEM, while the highly math-competent men are more likely to be pigeonholed in STEM by their narrower abilities. Third, these math-proficient women are more likely than math-proficient men to prefer careers in non-STEM fields, and those women who do initially choose STEM fields are more likely than men to leave those careers as they advance. Finally, in some STEM fields, women are penalized in promotion rates for having children. This review (Ceci et al., 2009) and two other reviews by Ceci and colleagues (Ceci, Ginther, Kahn, & Williams, 2014; Ceci & Williams, 2011) conclude that women's preference for non-STEM careers is the main reason for their underrepresentation in STEM. Thus, in order to begin addressing women's underrepresentation in STEM, it is important to first examine the origins of their preference for non-STEM careers, as well as why women are more likely than men to leave STEM fields. Furthermore, exploration of women's preferences needs to consider variation within STEM and focus on the areas of STEM in which women are most underrepresented.

Chapter 2

Performance Attributions

One factor that may contribute to women's preference for non-STEM careers is their attributions for success and failure. Weiner's traditional model of attributions for success and failure (Weiner et al., 1972) proposes that people interpret their performance for any achievement-related outcome in terms of four causes: ability, effort, task difficulty, and luck. Additionally, people use these same causal elements to predict future performance. This traditional model of performance attributions also provides a classification scheme for these four factors, which divides the factors along two dimensions: stability and locus of control. Ability is considered stable and internal, effort is unstable and internal, task difficulty is stable and external, and luck is unstable and external. Weiner and colleagues (1972) note that an important result of the stability dimension is that attributions to unstable causes lead people to think that their future performance may be different. They also explain that people tend to give greater weight to luck as a causal factor when performance is quite variable and even seems random, while people give greater weight to task difficulty when their performance is similar to that of others. Weiner and colleagues (1972) provide less guidance on what influences attributions to ability and effort, as well as when individuals make internal versus external attributions.

A review of the gender and motivation literature (Meece, Glienke, & Burg, 2006) reported that females were more likely than males to attribute success to effort and failure to lack of ability. Moreover, these gender differences in performance attributions seem to be specific to math and science-related tasks. One potential source of these gender differences in performance attributions that men and women make for their own performance are the gendered attributions that parents and teachers make for their children's and students' performance (Gunderson,

Ramirez, Levine, & Beilock, 2012). A review of parents′ and teachers′ influence on gender differences in math attitudes reported that the attributions that parents and teachers made for children′s math achievement were dependent on the child′s gender (Gunderson et al., 2012). Specifically, boys′ success in math is more likely to be viewed as due to ability rather than effort compared to girls′ success, and the reverse occurs for perceptions of failure, with boys′ failure blamed on lack of effort and girls′ failure attributed to lack of ability. Gunderson and colleagues (2012) suggest that children internalize the attributions that important adults in their lives make for their math achievement, resulting in children perpetuating these gendered patterns of attributions.

The influence of teachers' attributions and feedback on their students' own attributions is supported by Carol Dweck's early work on gender differences in achievement-related learned helplessness (Dweck, Davidson, Nelson, & Enna, 1978). In their first study, Dweck and her colleagues found that teachers were much more likely to attribute boys' failures to lack of motivation (effort) compared to girls' failures. Furthermore, the differences in attributions for boys' and girls' failures were accompanied by differences in the feedback teachers provided to students about their work. The vast majority of negative feedback provided to girls focused on intellectual aspects of their performance, while just over half of the negative feedback to boys focused on intellectual aspects; the rest of the negative feedback boys received was not related to the intellectual aspects of the task, and instead focused on things such as not following instructions or not being neat. In the second study, the researchers manipulated the feedback that children received to be similar to the type of feedback typically received by either girls or boys, based on the results of the first study (i.e., mostly relevant to the correctness of responses v. a mix of relevant to correctness and nonintellectual aspects of performance). They found that most

of the children in the "boy" feedback condition did not view negative feedback on a subsequent task as being due to a lack of ability; instead they generally thought it was due to lack of effort. However, children in the "girl" feedback conditions (there were two variations that had different absolute amounts of negative feedback) were more likely to say that the subsequent negative feedback was due to lack of ability (Dweck et al., 1978).

Gender differences in performance attributions may contribute to women disengaging from STEM at a greater rate than men for a couple reasons. Attributing failure to ability makes it seem that one's low performance is likely to continue in the future, no matter what one does, so there is no point in trying. In contrast, the unstable, but internal nature of effort means that attributing one's failures to effort should lead to the belief that future performance can be improved, and moreover, one is in control of that outcome. As a result, attributions of failure to effort should increase persistence and effort in the future, and can even lead to improvements in performance (Dweck, 1975; Weiner et al., 1972). Therefore, women should be more likely than men to disengage from STEM following failure, based on the gender differences in attributions for failure.

Gender differences in attributions for success can also have implications for STEM engagement. Gunderson and colleagues (2012) argue that parents' attributions of daughters' successes in math to effort and sons' successes in math to ability lead them to perceive their sons as having higher math abilities than their daughters, even when their objective performance is the same. The same process can occur with students' attributions for their own performance. When boys attribute their success to ability, they think highly of their abilities and believe they will continue to do well in the future. In contrast, when girls attribute their success to effort, that same objective success does not provide any information about their ability. Indeed, Meece and

colleagues (2006) also reported gender differences in math competency beliefs and math, science, and computer self-efficacy, and math and science expectancies, with boys more confident in their abilities and expecting to do better in the future compared to girls. This confidence gap in math has also been found in an analysis of over 17,000 eighth to tenth grade students from the National Educational Longitudinal Study of 1988, even when controlling for students' ability and performance feedback (Correll, 2001). In addition, Correll found that students' math self-assessment predicted their likelihood of enrolling in calculus in high school and choosing a quantitative major (engineering, computer programming, physical sciences, mathematics, and statistics) in college. Furthermore, she found that math self-assessment fully mediated the effect of gender on enrollment in high school calculus and partially mediated the effect of gender on choice of a quantitative major in college (Correll, 2001). Thus, gender differences in attributions can contribute to women dropping out of STEM, even among those with high competence, particularly because these women are more likely than similar men to also have high verbal competence (Ceci et al., 2009), and they do not seem to make the same pattern of attributions in verbal domains (Meece et al., 2006).

However, the past studies on gender differences in performance attributions, including those reviewed by Meece and colleagues (2006), have substantial limitations. These studies generally asked participants to make attributions for either past successes and failures (e.g., Eccles et al., 1983) or hypothetical successes and failure (e.g., Beyer, 1998; Crandall, Katkovsky, & Crandall, 1965; Mok, Kennedy, & Moore, 2010); none manipulated performance directly. Thus, these studies may not reflect the attributions that students make in real time for actual success and failure. One study that did ask participants to make attributions for task performance in real time manipulated performance through task difficulty, resulting in a

confound that likely affected results (Wiegers & Frieze, 1977). In addition, all of the studies cited by the Meece et al. (2006) review were published prior to 2000. One study mentioned above was conducted more recently (Mok et al., 2010), but it used participants in Hong Kong. With changes over time and by country in gender equality, it's unclear whether the same trends would be found in the United States today, particularly since research on other gender differences related to STEM has found that gender gaps vary by country along with gender equality (Hyde & Mertz, 2009). Thus, the first step in understanding how attributions may contribute to women's underrepresentation in STEM is to establish that gender differences in performance attributions do currently exist in the U.S., using a real task, with performance manipulated independent of task difficulty.

Chapter 3

Growth Mindset

Carol Dweck began her career combining research on attribution with research on learned helplessness to study how children cope with failure (Dweck, 2012). In exploring helpless versus mastery oriented responses to failure, Dweck discovered that it went beyond attributional styles. When mastery-oriented children were asked to talk aloud as they solved problems after difficulties, they rarely gave attributions, and instead focused on finding new ways to solve the problems (Diener & Dweck, 1978). Combining this work with achievement motivation, Dweck then found that the adoption of learning versus performance goals lead to mastery and helpless responses to failure, respectively (Elliot & Dweck, 1988). From there, Dweck and one of her students postulated that ideas about ability as either malleable or innate could underlie chronic differences in learning versus performance goals (Dweck, 2012). This began decades of research by Dweck and her colleagues on implicit theories, or beliefs on the nature of human attributes.

In terms of any given attribute, such as intelligence, individuals can be entity theorists, believing that one's level of the attribute is innate and cannot be developed, or incremental theorists, believing that one can grow the attribute with the proper motivation, opportunity, and guidance (Dweck, 2012). Entity theorists are commonly referred to as having a fixed mindset, and incremental theorists are often referred to as having a growth mindset (Dweck, 2006). These mindsets can be applied to a wide variety of traits, but we will focus on them in terms of beliefs about intelligence and academic abilities specifically.

As postulated by Dweck's original ideas about implicit theories, these mindsets lead people to set different types of goals. Those with a growth mindset tend to set learning goals so they can improve their abilities, while those with a fixed mindset tend to set performance goals,

which help prove that they are smart (Burnette, O'Boyle, VanEpps, Pollack, & Finkel, 2013; Dweck, 2012). In fact, individuals with a fixed mindset will even avoid taking steps to improve their skills if those actions would make them look stupid (e.g., remedial courses, Hong, Chiu, Dweck, Lin, & Wan, 1999).

These beliefs about intelligence and abilities also lead to differences in perceptions of effort. Adolescents with a growth mindset tend to believe that effort leads to positive outcomes—the harder you work at something, the better you will do. In contrast, adolescents with a fixed mindset tend to believe that effort will not lead to improvement and is even a sign that their ability is low—trying hard makes them feel like they are not good at what they are working on (Blackwell, Trzesniewski, & Dweck, 2007). Following from these differing views of effort, individuals with a growth versus fixed mindset make different attributions for failure. Adolescents with a fixed mindset have been found to be more likely than those with a growth mindset to make helpless attributions—attributions to low ability, unfairness, or lack of interest—for hypothetical failure (Blackwell et al., 2007). Additionally, college students with a growth mindset have been found to be more likely than those with a fixed mindset to attribute actual failure to lack of effort (Hong et al., 1999; Robins & Pals, 2002).

Since failure is much more threatening to those with a fixed mindset, who think it indicates that they are not smart enough, these individuals are more likely to experience negative emotions related to their academic performance (Burnette et al., 2013; Robins & Pals, 2002). They are also more likely to respond to failure with helpless or defensive strategies, such as choosing an easier task, reviewing material they have already mastered, cheating, decreasing effort or giving up, and comparing their exam with someone who performed worse (Blackwell et al., 2007; Burnett et al., 2013; Nussbaum & Dweck, 2008; Robins & Pals, 2002). In contrast,

those with a growth mindset engage in mastery-oriented strategies following failure, such as increasing effort, spending more time studying, reviewing material they have not yet mastered, and comparing their exam to that of someone who did better so they can learn the correct answers, (Blackwell et al., 2007; Burnett et al., 2013; Nussbaum & Dweck, 2008; Robins & Pals, 2002).

As a result of these differing strategies in response to failure, individuals with a growth mindset tend to perform better academically over time, specifically when they face difficult, new material, such as during the transition to junior high school (Blackwell et al., 2007) and college (Robins & Pals, 2002). In fact, Blackwell and colleagues (2007) found in their longitudinal study of junior high school students that there was no difference between students with a fixed versus growth mindset at the beginning of junior high school, but mindset predicted different trajectories over the course of seventh and eighth grade, with students with a growth mindset actually improving over time, resulting in a growing gap in grades between students with a fixed mindset and those with a growth mindset. Similarly, the gap in self-esteem between those with a fixed mindset and those with a growth mindset widens over the course of college, with self-esteem decreasing for those with a fixed mindset (Robins & Pals, 2002). Blackwell and colleagues' (2007) longitudinal study combined many of these differences between adolescents with a growth mindset and those with a fixed mindset into one comprehensive model. Their results demonstrated that growth mindset directly predicted learning goals and positive beliefs about effort. Positive beliefs about effort then predicted lower helpless attributions, both of which, along with learning goals, predicted positive strategies, and it was these positive strategies that predicted an improvement in grades (Blackwell et al., 2007).

While it seems that growth mindset is broadly beneficial, particularly for persistence and academic achievement, there is also some evidence that fixed mindsets might play a particular role in women's underrepresentation in certain fields, including STEM fields. One study examined the extent to which nearly 2,000 academics (professors, post-doctoral fellows, and graduate students) in 30 fields, including both STEM and non-STEM disciplines, believed that success in their field requires innate talent—essentially field-specific fixed mindset—and how much they thought others in their field held the same belief (Leslie, Cimpian, Meyer, & Freeland, 2015). They found that the belief in a field that success was based on innate talent was negatively correlated with the percentage of female Ph.D. recipients in that field, and this was true for both STEM and non-STEM fields. The fields that were high in fixed mindset also tended to endorse the belief that women are not suited to high-level work in their field and that their field is not welcoming to women. Moreover, the belief that women are not suited to high-level work and the unwelcoming environment for women significantly mediated the effect of fixed mindset on representation of women (Leslie et al., 2015).

This idea that a fixed mindset among members of a field, or even the perception of a field as having a fixed mindset, could have negative consequences for women has also specifically been examined in the context of college calculus courses (Good, Rattan, & Dweck, 2012). In this study, researchers examined male and female calculus students' perceptions of the beliefs of others in their class regarding whether math ability is innate and whether men and women have equal math abilities, as well as their sense of belonging in math, their intentions to pursue math in the future, and their math grades. They found that the combination of perceiving that one's classmates had a fixed mindset and believed gender stereotypes about math predicted lower belonging at the end of the semester only among women, even after controlling for initial

belonging. Furthermore, belonging at the end of the semester predicted intentions to pursue math and course grades (Good et al., 2012). In other words, the perception of a growth mindset among one's classmates can protect women from negative effects of stereotypical beliefs regarding their math abilities, presumably because stereotypes about a lack of ability are less important when those abilities can be developed.

Chapter 4

Failure and Persistence in STEM

Failure, poor performance, and other challenges are a common and even normative part of STEM fields. The scientific method, engineering design process, and computer coding all involve some element of trial and error, and it is impossible to have experiments, designs, or code work perfectly every time. In addition, college-level coursework in STEM is generally much more difficult than anything students experienced in high school, even in Advanced Placement courses. Thus, even students who excelled in high school can find themselves unexpectedly struggling in college. Then there is a general perception that introductory courses in science and math at the college level are often purposefully difficult in order to discourage students who do not have what it takes to succeed in the major from continuing in the field, which is exemplified by the fact that they are often referred to as "gateway" courses. Many of these courses are also graded on a curve, making it even more difficult for students to earn A's. Consequently, college students in STEM majors may experience many instances of failure or performance poor enough to be interpreted as failure.

These experiences, particularly in gateway courses, may be viewed by students as evidence that they are not suited for their particular major, or even STEM majors altogether. This interpretation seems especially likely when students are under the false impression that they are the only ones struggling. In addition, the leap from failures or low grades to switching majors should be more likely when students attribute those failures to their own lack of ability, and even more so when they believe that ability is innate. Thus, if prior research on gender differences in attributions still holds (Meece et al., 2006), then women, particularly those with a fixed mindset, should be more susceptible to interpreting their difficulties in college STEM courses as

indicating that they do not have what it takes to succeed in their major or STEM in general. Therefore, these failure experiences could cause women to disengage and drop out of STEM majors at greater rates.

Switching to non-STEM majors may be particularly likely if students believe non-STEM courses are easier or that their introductory courses do not serve as gateway courses in the same way, culling students who do not “have what it takes.” Similarly, students who have strong skills in other areas should be more likely to respond to STEM failures by switching to a non-STEM major since they can expect to excel in those fields, particularly if they are already taking non-STEM courses where they are likely not experiencing the same level of failure as in their STEM courses. Since women with strong quantitative skills are more likely than men to also have strong verbal skills (Ceci et al., 2009), these women should feel that they have other fields they can fall back on and thus be more likely than men to switch to non-STEM majors when they encounter failure.

Two existing studies provide partial support for this idea that the frequent difficulties in STEM fields, combined with women’s responses to failure, may help to explain women’s underrepresentation in STEM, including why some women drop out of STEM fields. Over three decades ago, Barbara Licht and Carol Dweck (1984) proposed that the relative lack of women in STEM might partly be due to a combination of gender differences in mastery orientation versus a more helpless attribution style, along with the fact that the way math is taught involves abrupt transitions between units (e.g., arithmetic to calculus), such that students are suddenly inundated with new concepts, which may be confusing. To test this idea, they designed a study in which helpless and mastery oriented fifth-grade students learned new material in a classroom setting. The students were randomly assigned to first learn new material that was either presented in a

confusing or straightforward way, and then all students learned a second set of new material that was presented clearly and were tested only on the second set of material. Consistent with hypotheses, they found that mastery-oriented children were not affected by the confusion manipulation, but helpless children were less likely to be able to master the material when they had first been presented with other material in a confusing manner (Licht & Dweck, 1984).

Another study demonstrates the importance of feeling that one has the skills necessary to succeed for persistence in engineering specifically. A longitudinal study of students who began college in an engineering major measured a variety of factors thought to play a role in persistence in STEM fields during participants' freshman year, and then followed up with participants in their senior year to see whether they had persisted in an engineering major (Cech, Rubineau, Silbey, & Seron, 2011). The researchers found that what they called expertise confidence—students' confidence in developing useful skills, advancing to the next level in engineering, and their ability to be successful in their career, due to their engineering courses—during students' freshmen year predicted their completion of an engineering major, as opposed to switching to a different STEM major, even after controlling for demographics, grade point average, family plans, and self-assessed skills. Importantly, women had lower expertise confidence than men, and once expertise confidence was included in their models, gender was no longer a significant predictor of retention in engineering (Cech et al., 2011). Thus, women's relative lack of confidence in their ability to succeed seems to contribute to their underrepresentation in engineering. This article does not explain why women feel less confident in their ability to succeed in engineering, but a helpless attribution style and other less resilient responses to the many difficulties students experience in gateway STEM courses in college could

play a role, and therefore be a factor in women's underrepresentation in engineering and other STEM fields.

Chapter 5

Overview of Studies and Hypotheses

A prominent factor in women's underrepresentation in STEM as a whole, and computer science, engineering, and physics, in particular, is women's choices (Ceci et al., 2009; Ceci et al., 2014; Ceci & Williams, 2011). There are numerous factors that contribute to women's initial preference for non-STEM fields, and their greater likelihood of leaving STEM fields. One factor explored in the current work is women's beliefs about their ability to succeed, and how those beliefs are shaped by gender differences in responses to success and failure. One aspect of these responses to success and failure are the attributions that men and women make for their performance. There is some evidence that females are more likely than males to attribute their successes to effort and their failures to lack of ability in STEM fields (Meece et al., 2006), leading them to believe less in their ability to succeed in the future. Furthermore, the stable nature of ability attributions makes them particularly harmful for beliefs about future performance and persistence (Dweck, 1975; Weiner et al., 1972), particularly for individuals with a fixed mindset. Thus, gender differences in attributions should have greater influence on career choices in the face of challenges and other difficulties. Because failure is a prominent part of STEM fields, particularly during introductory college courses, gender differences in attributions could help explain why women tend to switch out of STEM majors, or switch to "easier" STEM majors from fields such as engineering and computer science.

However, there are some major limitations to the prior findings on gender differences in attributions to success and failure (Meece et al., 2006). Namely, most of the research was conducted at least 15 years ago, more recent research was conducted outside of the United States (e.g., Mok et al., 2010), and studies generally did not directly manipulate performance feedback

(e.g., Eccles et al., 1983). Thus, research needs to establish that this pattern of gender differences does hold. Furthermore, gender and domain differences in other aspects of responses to success and failure, such as performance expectations, self-esteem, and persistence, should be examined to better understand the full picture of how these responses may lead women to drop out of STEM fields. In addition, since the connection between attributions and women defecting to non-STEM fields is based on the idea that students believe that a failure in an introductory STEM course means they are not cut out for the field, and that they are more likely to succeed in non-STEM majors, it needs to be empirically established that students do in fact perceive STEM introductory courses as being used to remove students who do not have what it takes to succeed and that they perceive non-STEM majors as being easier.

Four studies were designed to demonstrate that STEM majors, and introductory classes in particular, are perceived as being more difficult than non-STEM majors and even purposefully difficult to discourage students who do not have what it takes to succeed, and to examine gender differences in college students' responses to success and failure, as well as how that may vary between STEM and non-STEM domains, the consequences for persistence, and whether feedback that makes failure seem normative can lead to more resilient responses.

The purpose of Study 1 was to empirically demonstrate that college students perceive courses in general, as well as introductory courses in particular, to be more difficult in STEM majors compared to non-STEM majors, and similarly, that college students perceive introductory courses in STEM majors as being used to cull students who do not have what it takes, to a greater degree than introductory courses in non-STEM majors. In addition, analyses examined differences in perceptions of majors by gender and type of major.

Study 2 piloted a novel task to be used in the subsequent studies to ensure that participants understood the task and took it seriously. It also examined participants' perceptions of scores on the task to determine which scores were perceived as failure and which were perceived as success, so that those scores could be used to manipulate success and failure in the subsequent experimental studies. Finally, the manipulation of domain was also piloted using a manipulation check to assess whether participants paid attention to the manipulation, as well as examining preliminary differences in responses by condition as an indication that the manipulation affected participants' perceptions of the task.

Study 3 examined gender and domain differences in responses to success and failure among two samples of college students: students in psychology courses and students majoring in engineering, computer science, and physics. Specifically, persistence, as well as other variables that may impact persistence, including performance attributions, performance state self-esteem, and expectations for future performance, were examined following feedback indicating success or failure on the novel task, which was described as associated with success in either STEM or arts and humanities. This study also examined growth mindset as a potential moderator.

Finally, Study 4 focused specifically on responses to failure when the novel task was described as measuring STEM-relevant skills, since responses to failure in STEM have the potential to be most damaging to persistence in those fields. This study examined whether including information that makes failure seem more normative can increase resilient responses to failure, and whether those effects differ by gender and growth mindset. This study used two interventions: an *a priori* intervention that indicated that students generally initially do poorly on the novel task and then improve with experience, and an intervention based on participant

feedback from Study 3, which indicated that participants' performance on the task was about average compared to others.

The following hypotheses were tested:

Hypothesis 1: Students will perceive STEM courses as more difficult and being used to weed students out more compared to non-STEM courses. These differences will be larger among women than men. This hypothesis will be tested in Study 1.

Hypothesis 2: Students will react more negatively to failure than success. Specifically, they will report lower performance state self-esteem, less satisfaction with their scores, make more external attributions for their performance, expect lower scores in the future, and have lower persistence after failure. However, we also hypothesized that this effect of feedback would be qualified by a number of interactions. Specifically, we hypothesized that the more students hold growth mindsets, the smaller the negative effect of failure. Moreover, we hypothesized that the negative effects of failure will be larger for women in the STEM condition. The main hypotheses will be tested in Study 3, and the hypotheses regarding smaller effects of failure with a growth mindset and larger effects of failure among women will be tested in Studies 3 and 4.

Hypothesis 3: Women will be more likely than men to attribute success in STEM to effort and failure in STEM to ability. Attributions to success will be tested solely in Study 3, whereas attributions to failure will be tested in Studies 3 and 4.

Hypothesis 4: Both interventions that make failure seem normative will result in more positive reactions to failure relative to the control condition, in terms of performance state self-esteem, performance satisfaction, attributions to external or unstable factors, performance expectations, and persistence. However, the benefits of the interventions will be larger among women. In addition, the *a priori* intervention, which emphasizes improvement with experience,

will have a more positive impact than the comparative intervention for individuals with a fixed mindset. These hypotheses will be tested in Study 4.

Chapter 6

Study 1: Perceptions of College STEM Courses

The primary goal of the current study was to empirically test the hypothesis that college students perceive courses in STEM majors to be more difficult than those in non-STEM majors, both for courses in general and introductory courses in particular, and that they also perceive introductory STEM courses as being used to remove students who do not have what it takes to succeed in those fields, more so than introductory courses in non-STEM majors. A secondary goal was to explore whether these perceptions of majors vary by gender and type of major.

To address these questions, data were collected from two samples. The first were students from psychology courses who participate in research studies for extra credit. These students were given the opportunity to complete an online survey for research credit. The survey included measures from several researchers on a variety of topics. The relevant items to the current study were the demographics questions and three items regarding perceptions of 16 types of majors (e.g., history) or categories of majors (e.g., physical sciences), evenly split between STEM and non-STEM fields. The second sample were engineering, computer science, and physics students who volunteered to participate in a psychology research study for pay. These particular majors were chosen because they are the STEM majors in which women are most underrepresented, in general and at the University of Connecticut where the research was conducted. The vast majority came to the laboratory to participate, but a small portion of the sample participated online. Following the experimental portion of the study (see Study 3), these students answered the same items as the psychology sample.

Methods

Participants

Participants were undergraduate students at the University of Connecticut who were either part of the psychology participant pool (*N* = 869) or were engineering, computer science, and physics majors (*N* = 209), hereafter referred to as the engineering sample. Among the psychology sample, the most common fields for one's primary major or intended major were business (25.0%), health (17.1%), biological sciences (16.5%), and psychological sciences (16.5%).[1] Among the engineering sample, most were majoring in, or intended to major in, some type of engineering (74.6%), followed by computer science and engineering (21.5%), and physical sciences (1.9%). Four participants from the engineering sample indicated a primary field other than engineering, computer science and engineering, or physical sciences, meaning that while they were majoring or intending to major in one of those fields, it was not their primary major or intended major.

Most participants in the study were from the Storrs campus (90.7%; Hartford 4.3%, Waterbury 4.1%, Torrington, 0.2%, Avery Point 0.2%, did not indicate a campus 0.6%), and the majority were freshman (39.5%) or sophomores (35.9%; 18.2% juniors, 5.8% seniors, 0.6% did not indicate their year). The 1,000 participants who provided their age were between 17 and 72 years old (*M*=19.15, *SD*=2.15). The majority of the sample identified as White (71.4%)[2], non-Hispanic (88.0%), women (56.1%; see Table 1 for more detail). Twelve participants who did not identify as either women or men were eliminated from further analyses since gender was used as an independent variable, resulting in a sample of 1,066 participants.

Procedures

---

[1] When possible, "other" majors were reclassified into one of the 16 available categories.
[2] In the psychology sample, participants were allowed to check all that apply for race, including a multiracial option, while in the STEM sample, participants could only make one selection for race, including a multiracial option.

Participants from the psychology participant pool completed the survey online for research participation credit, while the engineering, computer science, and physics students received \$5 to \$10 for completing the study, depending on the semester and whether they completed the study in the laboratory (N = 198) or online (N = 11). For the psychology sample, the questions relevant to this study regarding perceptions of majors were intermixed with questions on a variety of other topics. The engineering sample was recruited using announcements that indicated that participants had to be in the engineering school, which includes computer science, or physics majors. The announcements directed students to a screener survey, which confirmed that they were in the engineering school or majoring in physics. The engineering participants completed an experimental study (see Study 3), and after completing the measures for that study, they completed the measures of perceptions of majors and demographics.

Measures

Participants responded to three questions on their perceptions of 16 different majors or types of majors. They indicated how difficult they think courses (generally) in each field are, as well as how difficult they think introductory level courses specifically in each field are, and how much they think introductory courses in each field are used to weed out students who do not have what it takes to succeed. These questions were answered using a 5-point Likert scale, with the first two questions going from very easy to very difficult, and the last question going from not at all to completely. The 16 fields were evenly split between STEM (e.g., engineering, psychological sciences) and non-STEM (e.g., history, visual and performing arts) fields, with the STEM categories based on the NSF definition of STEM. Responses for each question were averaged across the STEM and non-STEM categories, such that perceptions of the difficulty of

STEM (α = .792) and non-STEM courses in general (α = .765), and STEM (α = .838) and non-STEM introductory courses (α = .875), as well as the extent to which STEM (α = .815) and non-STEM (α = .857) introductory courses are used to weed students out were assessed.

Results

A repeated measures ANOVA was conducted for each of the three ratings comparing STEM and non-STEM fields, using gender and sample as between-subjects factors.

Perception of the Difficulty of All Courses

There was a significant main effect of field, F (1,1057) = 1,218.79, $p < .001$, $\eta_p^2 = .536$, with STEM courses ($M = 3.66$, $SD = 0.59$) perceived as more difficult than non-STEM courses ($M = 2.83$, $SD = 0.60$). There was also a significant main effect of gender, F (1,1057) = 8.77, $p = .003$, $\eta_p^2 = .008$, with women perceiving courses in general as more difficult ($M = 3.32$, $SD = 0.57$) compared to men ($M = 3.14$, $SD = 0.62$). Finally, there was a significant main effect of sample, F (1,1057) = 7.36, $p = .007$, $\eta_p^2 = .007$, with the psychology sample perceiving courses in general as more difficult ($M = 3.27$, $SD = 0.60$) compared to the engineering sample ($M = 3.14$, $SD = 0.58$).

These main effects were qualified by several interactions. There was a significant interaction between field and gender, F (1,1057) = 10.32, $p = .001$, $\eta_p^2 = .010$, such that women perceived STEM courses as significantly more difficult than men did, F (1,1057) = 18.50, $p < .001$, $\eta_p^2 = .017$, but there was no gender difference for perceived difficulty of non-STEM courses, F (1,1057) = 0.54, $p = .462$, $\eta_p^2 = .001$. There was also a significant interaction between field and sample, F (1,1057) = 9.04, $p = .003$, $\eta_p^2 = .008$, such that the psychology sample perceived non-STEM courses as more difficult than the engineering sample did, F (1,1057) = 14.80, $p < .001$, $\eta_p^2 = .014$, but there was no significant difference between samples for

perceptions of STEM courses, *F* (1,1057) = 0.44, $p = .508$, $\eta_p^2 < .001$. Finally, there was a significant interaction between sample and gender, *F* (1,1057) = 5.08, $p = .024$, $\eta_p^2 = .005$, such that in the psychology sample, women perceived courses as being more difficult than men, *F* (1,851) = 33.85, $p < .001$, $\eta_p^2 = .038$, but this gender difference was not significant in the engineering sample, *F* (1,206) = 0.16, $p = .692$, $\eta_p^2 = .001$. The three-way interaction between field, gender, and sample was not significant, *F* (1,1057) = 1.11, $p = .291$, $\eta_p^2 = .001$.

Perception of the Difficulty of Introductory Courses

There was a significant main effect of field, *F* (1,1050) = 804.00, $p < .001$, $\eta_p^2 = .434$, with STEM introductory courses (*M* = 3.10, *SD* = 0.67) perceived as more difficult than non-STEM introductory courses (*M* = 2.45, *SD* = 0.71). There was also a significant main effect of sample, *F* (1,1050) = 44.58, $p < .001$, $\eta_p^2 = .041$, with the psychology sample perceiving introductory courses as more difficult (*M* = 2.84, *SD* = 0.69) than the engineering sample (*M* = 2.52, *SD* = 0.63). There was no significant main effect of gender, *F* (1,1050) = 1.83, $p = .177$, $\eta_p^2 = .002$.

These main effects were qualified by several interactions. There was a significant interaction between field and gender, *F* (1,1050) = 22.01, $p < .001$, $\eta_p^2 = .021$, such that women perceived introductory STEM courses as significantly more difficult than men did, *F* (1,1050) = 11.88, $p = .001$, $\eta_p^2 = .011$, but there was no gender difference for perceived difficulty of introductory non-STEM courses, *F* (1,1050) = 0.78, $p = .376$, $\eta_p^2 = .001$. There was also a significant interaction between field and sample, *F* (1,1050) = 6.57, $p = .011$, $\eta_p^2 = .006$, such that the psychology sample perceived both STEM, *F* (1,1050) = 24.71, $p < .001$, $\eta_p^2 = .023$, and non-STEM introductory courses, *F* (1,1050) = 47.36, $p < .001$, $\eta_p^2 = .043$, as more difficult than the engineering sample did, but the effect of sample was larger for non-STEM introductory

courses. Finally, there was a marginally significant interaction between sample and gender, F (1,1057) = 2.77, p = .097, η_p^2 = .003, such that women perceived courses as more difficult than men in the psychology sample, F (1,845) = 10.94, p = .001, η_p^2 = .013, but not in the engineering sample, F (1,205) = 0.04, p = .852, η_p^2 < .001. The three-way interaction between field, gender, and sample was not significant, F (1,1050) = 0.10, p = .751, η_p^2 < .001.

Perception of Introductory Courses Being Used to Weed Out Students

There was a significant main effect of field, F (1,1046) = 1,212.82, p < .001, η_p^2 = .537, with STEM introductory courses (*M* = 3.21, *SD* = 0.70) perceived as being use to weed students out more so than non-STEM introductory courses (*M* = 2.42, *SD* = 0.77). There was also a significant main effect of sample, F (1,1046) = 31.00, p < .001, η_p^2 = .029, with the psychology sample (*M* = 2.87, *SD* = 0.74) perceiving introductory courses as being used to weed students out more so than the engineering sample (*M* = 2.59, *SD* = 0.64). There was no significant main effect of gender, F (1,1046) = 0.22, p = .638, η_p^2 < .001.

These main effects were qualified by several interactions. There was a significant interaction between field and gender, F (1,1046) = 22.00, p < .001, η_p^2 = .021, such that men perceived introductory non-STEM courses as being used to weed students out more than women did, F (1,1046) = 6.14, p = .013, η_p^2 = .006, but the gender difference was not significant for perceptions of STEM courses, F (1,1046) = 3.15, p = .076, η_p^2 = .003. There was also a significant interaction between field and sample, F (1,1046) = 61.58, p < .001, η_p^2 = .056, such that the psychology sample perceived non-STEM introductory courses as being used to weed students out more than the engineering sample, F (1,1046) = 68.50, p < .001, η_p^2 = .061, but there was no effect of sample for perceptions of weeding out in STEM introductory courses, F (1,1046) = 2.04, p = .153, η_p^2 = .002. The interaction between sample and gender, F (1,1046) =

0.50, p = .482, $\eta_p^2 < .001$, and the three-way interaction between sample, gender, and field, F (1,1046) = 0.00, p = .974, $\eta_p^2 < .001$, were not significant.

Discussion

The primary hypothesis was supported: participants perceived courses in STEM majors as more difficult than those in non-STEM majors, both for courses in general and introductory courses in particular, and they also perceived introductory STEM courses as being used to eliminate students who do not have what it takes to succeed in those fields, more so than introductory courses in non-STEM majors. The mean for introductory STEM classes being used to weed students out is above the mid-point, suggesting that college students do perceive these courses as being “weed out” courses. Thus, it seems plausible that students in STEM majors might respond to early struggles in these introductory courses by assuming that they do not have what it takes to succeed in the major and a future career in that particular field. Furthermore, since these students also believe that non-STEM courses are easier than STEM courses, they might then decide to switch to a non-STEM major where they believe they are more likely to succeed.

There were also main effects of gender and sample, which were qualified by several two-way interactions. Women perceived STEM courses (introductory and in general) as being more difficult than men did, and perceived introductory STEM courses as being used to cull students more so than men did. However, these gender differences were not present for perceptions of non-STEM courses, or were reversed in the case of introductory courses being used to cull students. These findings may suggest the presence of a confidence gap between men and women when it comes to STEM courses, as has been found in other research (Correll, 2001). However, women in psychology courses also perceived courses in general, regardless of field, as more

difficult than men in psychology courses did, but this gender difference was not present among engineering students. This may indicate that there may also a confidence gap that extends beyond STEM courses, at least for women who are in not in the STEM majors in which women are most underrepresented.

There were also some differences between the two samples in their perceptions of courses, but these differences were mostly restricted to perceptions of non-STEM courses—psychology students perceived non-STEM courses as being more difficult and used to remove students more than engineering and physics students did. Psychology students also perceived STEM introductory courses as being more difficult than engineering and physics students did, but this difference in perceptions was smaller than the difference in perceptions of non-STEM introductory courses. Since the engineering and physics students are less likely to have enrolled in any non-STEM courses, their perception of non-STEM courses as easier, compared to the opinions of the psychology students, may be rooted in the difference between opinions that are based on one's own experiences versus hearing about the experiences of others. In other words, the psychology students' views of non-STEM courses may be more rooted in reality, having taken such courses themselves. Alternatively, the engineering and physics students may view STEM courses as more difficult as a way to enhance self-esteem. Either way, this difference in perceptions by sample might indicate that engineering and physics students have a particularly optimistic view of the ease of non-STEM majors, which could influence choices to switch majors if they struggle in their own major.

Chapter 7

Study 2: Pilot Study of Novel Task

The purpose of this study was to pilot the novel task that was used in Studies 3 and 4. These were several aims in piloting this task: 1) to ensure that participants understood the task and took it seriously; 2) to determine the scores that were perceived as success and failure on the task, in order to use those scores as the manipulation in the subsequent studies; 3) to check whether participants paid attention and remembered the manipulation of domain; and 4) to examine whether the manipulation of domain affected perceptions of the task.

A novel task was used for this research because it allows performance to be manipulated without arousing the suspicions of participants, since they should not have specific expectations regarding their own performance without prior experience. Similarly, the novel nature of the task allows for manipulation of instructions, including information regarding the nature of the task, the skills it tests, and what other skills or fields success on the task is associated with. As such, expectations regarding the task and performance on it should be based primarily on instructions given by the researcher and not on participants' past experiences.

The novel task was based on a task that has previously used in research on causal attributions with children and college students (Mosatche, 1977) in which participants try to mark a dot in each of eight small circles on a piece of paper, but without being able to see what they are doing. This task was adapted to a computerized format so that multiple participants could complete the study at once; manipulations of condition could be done on the computer, allowing experimenters to be blind to condition; and to reduce suspicions of participants regarding their scores. In the adapted version piloted in this study, participants observe a set of eight rectangles of varying size on the computer screen. Once they are done observing, they

continue to the next screen where they must click on the previous location of each rectangle. To encourage participants to take the task seriously, they are given one practice trial to get used to the task, and then complete several regular trials. The location of the rectangles varies between trials, and the size of the rectangles decreases with each subsequent trial, making the task increasingly difficult. The visual-spatial nature of this task makes it potentially believable as testing skills related to STEM fields, as well as arts and humanities. The current study pilots this task, as well as the manipulation of domain (STEM v. non-STEM), using a sample of students from the psychology participant pool.

Methods

Participants

Forty college students were recruited from the University of Connecticut psychology participant pool. Information on the participant pool website indicated that the experiment was a study of a cognitive measure and briefly described the task. Twenty participants were female; 29 participants (72.5%) identified as White/Caucasian only, 10 as Asian/Asian American only, and 1 as both White/Caucasian and Asian/Asian American. Participants completed the study in individual cubicles. Participants were between 17 and 22 years old, with a mean age of 18.33 (*SD* = 0.98). Seventeen participants (42.5%) indicated that they were STEM majors or intended to major in a STEM field.

Procedures

Participants completed the study in the laboratory and received research participation credit. All sessions were conducted by a white female experimenter. Participants were randomly assigned by Qualtrics to the STEM or non-STEM condition, which was manipulated through the task instructions. These instructions, which were on the computer, explained that performance on

the task was associated with success in either STEM or arts and humanities and described how to complete the task:

> "Previous studies have found that performance on the task you are about to complete predicts success in science, technology, engineering, and math (STEM). The researchers are conducting this study to better understand why that is. You will complete 3 trials. On each trial, a set of 8 rectangles will appear on the screen. You should note the location and size of each rectangle. Then the rectangles will disappear and you will have to remember where they were and use the mouse to click on the previous location of each rectangle, so that if the rectangles were still visible, you would be clicking inside of each rectangle. You will have an opportunity to complete one practice trial before you begin the task."

An attention check ensured that participants read the manipulation. Participants completed one practice trial and three regular trials. For each trial, they could spend as much time as they wanted observing the rectangles and clicking on them, but only the most recent eight clicks—one for each rectangle—were recorded. Participants then responded to questions about the three trials, as well as a manipulation and suspicion check at the end. Once participants completed the study, the experimenter debriefed participants on the true nature of the study.

Measures

Participants' scores on each of the three trials were recorded, as well as the time they spent observing the rectangles and clicking on the locations for each trial. For each trial, participants were asked to give an estimate of their exact score, as well as the lowest and highest scores they think they could have possibly gotten. Participants also reported what score they thought would indicate that they had succeeded and what score would indicate that they had

failed for each trial. Finally, participants completed the suspicion check by responding to an open-ended question on what they thought the study was about, and then responded to the manipulation check question, which asked them to recall what performance on the task predicts.

Results

Repeated measures ANOVAs were conducted with trial as the within-subjects variable and condition and gender as between-subjects variables.

Manipulation and Suspicion Check

Out of forty participants, only one did not pass the manipulation check. This participant, who was in the non-STEM condition, reported at the end of the study that the instructions had indicated that success on the task predicts memory, rather than success in arts and humanities. None of the participants were suspicious of the manipulation.

Time on Task

Repeated measures ANOVAs were conducted for time spent on each part of the task. Participants spent between 3.86 and 97.46 seconds observing the rectangles for each trial (M_{trial1} = 21.77, SD_{trial1} = 14.95; M_{trial2} = 19.92, SD_{trial2} = 17.04; M_{trial3} = 20.99, SD_{trial3} = 21.62). There were no significant effects of trial, F (2,72) = 0.28, p = .760, η_p^2 = .008, gender, F (1,36) = 0.80, p = .377, η_p^2 = .022, or condition, F (1,36) = 0.02, p = .886, η_p^2 = .001. Similarly, the two-way interactions between trial and gender, F (2,72) = 0.06, p = .947, η_p^2 = .002, trial and condition, F (2,72) = 0.87, p = .423, η_p^2 = .024, and gender and condition, F (1,36) = 0.19, p = .669, η_p^2 = .005, were not significant. Finally, the three-way interaction was also not significant, F (2,72) = 1.09, p = .341, η_p^2 = .029.

Participants spent between 5.41 and 80.34 seconds clicking on the rectangles for each trial, (M_{trial1} = 15.46, SD_{trial1} = 10.09; M_{trial2} = 14.48, SD_{trial2} = 12.42; M_{trial3} = 12.19, SD_{trial3} =

6.13). There was a significant effect of trial, F (2,72) = 3.78, p = .028, η_p^2 = .095, such that participants spent less time clicking on each subsequent trial. The effect of gender, F (1,36) = 0.08, p = .786, η_p^2 = .002, and condition, F (1,36) = 0.35, p = .557, η_p^2 = .010, were not significant. The two-way interactions between trial and gender, F (2,72) = 0.88, p = .418, η_p^2 = .024, trial and condition, F (2,72) = 0.97, p = .385, η_p^2 = .026, and gender and condition, F (1,36) < 0.01, p = .997, η_p^2 < .001, were not significant. Finally, the three-way interaction was also not significant, F (2,72) = 0.43, p = .654, η_p^2 = .012.

Score

Possible scores on each trial were between 0 and 8, but the range for actual scores only went up to 7. There was a significant effect of trial on score, F (2, 72) = 7.33, p = .001, η_p^2 = .169, with scores decreasing from Trial 1 to Trial 3 (M_{trial1} = 4.08, SD_{trial1} = 1.54; M_{trial2} = 3.48, SD_{trial2} = 1.60; M_{trial3} = 2.98, SD_{trial3} = 1.39). The effects of gender, F (1,36) = 0.02, p = .892, η_p^2 = .001, and condition, F (1,36) < 0.01, p = .964, η_p^2 < .001, were not significant. The two-way interactions between trial and gender, F (2,72) = 0.06, p = .941, η_p^2 = .002, trial and condition, F (2,72) = 0.14, p = .869, η_p^2 = .004, and gender and condition, F (1,36) = 0.17, p = .684, η_p^2 = .005, were not significant. Finally, the three-way interaction was also not significant, F (2,72) = 0.79, p = .460, η_p^2 = .021.

Score Estimates

Repeated measures ANOVAs were conducted for participants’ estimates of their exact score, lowest possible score, and highest possible score. The range for participants’ estimates of their exact scores on each trial was between 1 and 8. There was a significant effect of trial, F (2,72) = 4.90, p = .010, η_p^2 = .120, with higher estimates for Trial 2 than the other two trials (M_{trial1} = 4.45, SD_{trial1} = 1.66; M_{trial2} = 5.30, SD_{trial2} = 1.50; M_{trial3} = 4.50, SD_{trial3} = 1.66). The

effect of gender, F (1,36) = 0.39, p = .537, η_p^2 = .011, and condition, F (1,36) = 1.34, p = .255, η_p^2 = .036, were not significant. The interaction between trial and condition was marginally significant, F (2,72) = 2.49, p = .090, η_p^2 = .065. This interaction is likely due to a significant difference by condition in estimated scores in Trial 1, F (1,36) = 5.57, p = .024, η_p^2 = .134, with participants in the STEM condition estimating lower scores (M = 3.85, SD = 1.42) than those in the non-STEM condition (M = 5.05, SD = 1.70). The effect of condition for score estimates for Trials 2, F (1,36) = 0.17, p = .687, η_p^2 = .005, and 3, F (1,36) = 0.03, p = .854, η_p^2 = .001, were not significant. The interaction between trial and gender, (2,72) = 0.07, p = .931, η_p^2 = .002, and gender and condition, F (1,36) = 0.07, p = .791, η_p^2 = .002, were not significant. Finally, the three-way interaction was also not significant, F (2,72) = 0.48, p = .618, η_p^2 = .013.

The range for participants' estimates of their lowest possible score on each trial was between 0 and 6 (M_{trial1} = 2.15, SD_{trial1} = 1.29; M_{trial2} = 2.49, SD_{trial2} = 1.37; M_{trial3} = 2.21, SD_{trial3} = 1.54). There were no significant effects of trial, F (2,68) = 1.54, p = .223, η_p^2 = .043, gender, F (1,34) = 0.39, p = .536, η_p^2 = .011, or condition, F (1,34) = 0.61, p = .441, η_p^2 = .018. There was a significant interaction between trial and condition, F (2,68) = 4.51, p = .015, η_p^2 = .117. Similar to the results for exact estimates, analyzing each trial separately revealed that there was a significant effect of condition for Trial 1, F (1,34) = 6.32, p = .017, η_p^2 = .157, with participants in the STEM condition estimating lower scores (M = 1.75, SD = 1.07) than those in the non-STEM condition (M = 2.72, SD = 1.67). The effect of condition was not significant for Trial 2, F (1,34) = 0.18, p = .678, η_p^2 = .005, or 3, F (1,34) = 0.29, p = .593, η_p^2 = .009. The two-way interactions between gender and trial, F (2,68) = 0.01, p = .988, η_p^2 < .001, and gender and condition, F (1,34) = 1.94, p = .173, η_p^2 = .054, were not significant. Finally, the three-way

interaction between trial, gender, and condition was not significant, F (2,68) = 0.56, p = .575, η_p^2 = .016.

The range for participants' estimates of their highest possible score on each trial was between 2 and 8. There was a significant main effect of trial, F (2,72) = 4.88, p = .010, η_p^2 = .119, with participants' estimating higher possible scores for Trial 2 (M_{trial1} = 6.08, SD_{trial1} = 1.44; M_{trial2} = 6.45, SD_{trial2} = 1.48; M_{trial3} = 5.85, SD_{trial3} = 1.55). There was also a marginally significant effect of gender, F (1,36) = 3.01, p = .091, η_p^2 = .077, with men (M = 6.47, SD = 1.18) estimating higher possible scores than women (M = 5.78, SD = 1.71). The effect of condition was not significant, F (1,36) = 1.30, p = .261, η_p^2 = .035. There was a significant interaction between trial and condition, F (2,72) = 3.62, p = .032, η_p^2 = .091. Similar to the results for exact estimates and lowest possible score, analyzing each trial separately revealed that there was a significant effect of condition for Trial 1, F (1,36) = 6.21, p = .017, η_p^2 = .147, with participants in the STEM condition estimating lower scores (M = 5.55, SD = 1.36) than those in the non-STEM condition (M = 6.60, SD = 1.35). The effect of condition was not significant for Trial 2, F (1,36) = 0.18, p = .677, η_p^2 = .005, or 3, F (1,36) = 0.05, p = .830, η_p^2 = .001. The two-way interactions between gender and trial, F (2,72) = 0.34, p = .711, η_p^2 = .009, and gender and condition, F (1,36) = 1.72, p = .198, η_p^2 = .046, were not significant. Finally, the three-way interaction between trial, gender, and condition was marginally significant, F (2,72) = 2.54, p = .086, η_p^2 = .066. Analyzing each trial separately revealed that this was due to a significant interaction between gender and condition for Trial 3, F (1,36) = 4.70, p = .037, η_p^2 = .115, with women (M = 4.90, SD = 1.37) estimating lower scores than men (M = 6.70, SD = 1.06) only in the STEM condition, F (1,18) = 10.80, p = .004, η_p^2 = .375. There was no effect of gender within the non-STEM condition for Trial 3, F (1,18) = 0.73, p = .791, η_p^2 = .004, nor was there a

significant interaction between gender and condition for Trials 1, F (1,36) = 0.13, p = .724, η_p^2 = .004, or 2, F (1,36) = 0.71, p = .406, η_p^2 = .019.

Perceptions of Scores

Repeated measures ANOVAs were conducted for participants' perception of which scores indicate failure and success. The range for the scores on each trial that participants reported would indicate that they had failed at the task were between 0 and 7, (M_{trial1}=2.50, SD_{trial1}=1.68; M_{trial2}=2.69, SD_{trial2}=1.47; M_{trial3}=2.64, SD_{trial3}=1.61). There were no significant effects of trial, F (2,70) = 1.25, p = .294, η_p^2 = .034, gender, F (1,35) = 2.07, p = .159, η_p^2 = .056, or condition, F (1,35) = 0.28, p = .603, η_p^2 = .008. The interaction between gender and condition was significant, F (1,35) = 6.27, p = .017, η_p^2 = .152, with men indicating higher scores as indicative of failure (M = 3.59, SD = 2.02) than women (M = 1.87, SD = 1.23) in the STEM condition only, F (1,17) = 6.10, p = .024, η_p^2 = .264. There was no effect of gender among participants in the non-STEM condition, F (1,18) = 0.76, p = .396, η_p^2 = .040. There was no significant interaction between trial and gender, F (2,70) = 0.15, p = .860, η_p^2 = .004, or trial and condition, F (2,70) = 1.25, p = .294, η_p^2 = .034. The three-way interaction was also not significant, F (2,70) = 0.69, p = .504, η_p^2 = .019.

The range for the scores on each trial that participants reported would indicate that they had succeeded at the task were between 2 and 8, (M_{trial1}=5.65, SD_{trial1}=1.31; M_{trial2}=5.79, SD_{trial2}=1.21; M_{trial3}=5.39, SD_{trial3}=1.26). There were no significant effects of trial, F (2,68) = 1.70, p = .190, η_p^2 = .048, gender, F (1,34) = 0.23, p = .632, η_p^2 = .007, or condition, F (1,34) = 0.05, p = .825, η_p^2 = .001. Similarly, the two-way interactions between trial and gender, F (2,68) = 0.90, p = .411, η_p^2 = .026, trial and condition, F (2,68) = 0.30, p = .744, η_p^2 = .009, and gender

and condition, F (1,34) = 1.86, p = .181, η_p^2 = .052, were not significant. Finally, the three-way interaction was also not significant, F (2,72) = 0.11, p = .898, η_p^2 = .003.

Discussion

On average, participants spent about 34 seconds total on each trial and scored 3.5 out of 8 for each trial. Based on the average scores and amount time spent on each trial, it seems that participants understood the task and took it seriously. The amount of time participants spent on the clicking portion of the task decreased from Trial 1 to Trial 3, which may be because participants felt more comfortable with the task. The average scores also decreased from Trial 1 to Trial 3, from 4 out of 8 on Trial 1 to 3 out of 8 on Trial 3, which likely indicates that the trials became progressively more difficult, as intended. However, participants' estimates of their scores did not decrease from Trial 1 to Trial 3, and in fact, their exact estimate and estimate of their highest possible score were highest on Trial 2. Combined with the decrease in time spent clicking, this may indicate that participants felt they were improving as they continued, but ultimately did recognize that Trial 3 was more difficult.

Across all three trials, participants indicated that on average a score of 2.61 would indicate that they had failed at the task. Participants' average estimate across all three trials of their lowest possible score was slightly lower: 2.28. Based on these responses, it seems that scores of 2 or 3 out of 8 on each trial would be perceived as failure and would be believable to participants, and thus are appropriate scores for the failure condition of future studies. Across all three trials, participants indicated that on average a score of 5.61 would indicate that they had succeeded at the task. Participants' average estimate across all three trials of their highest possible score was slightly higher: 6.13. Based on these responses, it seems that scores of 5 to 6

out of 8 on each trial would be perceived as success and would be believable to participants, and thus are appropriate scores for the success condition of subsequent studies.

Only one of the forty participants failed the manipulation check and none were suspicious of the manipulation, indicating that participants paid attention to the manipulation and the vast majority still remembered it at the end of the study. Further supporting the successful manipulation of domain are some of the preliminary results indicating significant interactions between condition and trial and condition and gender. For Trial 1 only, participants in the STEM condition reported significantly lower estimates of their lowest and highest possible scores compared to participants in the non-STEM condition. This could indicate that participants who thought the task was related to success in STEM were less confident in their ability to do well on the task or found it more difficult. Across all three trials, there was a significant interaction between condition and gender for the scores participants reported would indicate failure, with males reporting higher scores in the STEM condition compared to the non-STEM condition, and females showing the reverse pattern. Similar to the interpretation of the effect of condition on score estimates, this may indicate that females in the STEM condition felt the task was harder or that lower scores were more acceptable compared to females in the non-STEM condition, while men may have perceived the task as harder when it was described as relating to success in arts and humanities than when it was described as relating to success in STEM. These preliminary results need to be replicated in further studies, particularly because of the small sample size in this study. However, these results indicate that not only did participants pay attention to the manipulation of domain, but their responses were affected by it.

Chapter 8

Study 3: Gender and Domain Differences in Responses to Success and Failure

The goal of Study 3 was to examine gender and domain differences in responses to success and failure among college students using the novel task that was piloted in Study 2. Specifically, we aimed to 1) replicate past findings (Meece et al., 2006) that female students are more likely than male students to attribute success to effort and failure to lack of ability on STEM-related tasks; 2) examine gender and domain differences in attributions to other factors, including the rest of Weiner's traditional classification scheme (difficulty and luck; Weiner et al., 1972) and factors that we previously found college students made spontaneously (strategy and experience; Lawner & Quinn, 2016); and 3) investigate gender and domain differences in other types of responses to success and failure, specifically performance state self-esteem, performance satisfaction, expectations for future performance, and persistence.

Based on the preliminary findings of domain and gender differences in perceptions of scores in Study 2, we also decided to include questions regarding the task before participants received their scores in order to better understand differences in perceived difficulty and initial performance expectations. We posited that the participants in general, and particularly women, might perceive a task that is associated with STEM as more difficult and expect to receive lower scores, compared to if they believe the task is associated with arts and humanities. This may be related to the findings of Study 1 in which STEM courses were perceived as more difficult. Finally, we also sought to explore the role of growth mindset on responses to success and failure, since it has been found to be related to persistence following failure (Dweck & Sorich, 1999), as well as whether the benefits of a growth mindset vary by gender.

In order to address issues of generalizability, the current study involved two samples: students from psychology courses who participate in research for extra credit and the same sample of engineering and physics students used in Study 1, who were paid for their participation. We compared results for the two samples to better understand how gender and domain differences in responses to success and failure may affect retention of women in the STEM majors where women are most underrepresented. It is possible that women who go into these fields react more similarly to men. However, it may also be the case that failure on a STEM-related task prompts even stronger reactions among women in these STEM fields in which women are so underrepresented, due to the stereotypes that question their ability and belonging in STEM. For example, researchers have suggested that the chilly climate toward women in engineering could lead them to interpret poor performance as evidence that they do not belong in the field, and found that an intervention to address belonging improved the confidence and performance on women in the most male-dominated engineering majors (Walton, Logel, Spencer, Peach, & Zanna, 2015).

It was hypothesized that students would generally react more negatively to failure than success: having lower performance state self-esteem, being less satisfied with their scores, attributing their performance more to external factors, expecting lower scores in the future, and having lower persistence. However, we also hypothesized that this effect of feedback would be qualified by a number of interactions. Specifically, we hypothesized that the more of a growth mindset one held, the smaller the negative effect of failure, and that the negative effects of failure would be larger for women in the STEM condition.

Methods

Participants

Participants were undergraduate students at the University of Connecticut who were either part of the psychology participant pool (*N* = 249) or were engineering, computer science, and physics majors (*N* = 209), hereafter referred to as the engineering sample. Among the psychology sample, the most common fields for one's primary major or intended major were business (26.5%), health (23.7%), biological sciences (14.5%), psychological sciences (11.6%), and social sciences (7.6%).[3] Among the engineering sample, most were majoring in, or intended to major in, some type of engineering (74.6%), followed by computer science and engineering (21.5%), and physical sciences (1.9%). Four participants from the engineering sample indicated a primary field other than engineering, computer science and engineering, or physical sciences, meaning that while they majoring or intending to major in one of those fields, it was not their primary major or intended major.

The majority of participants were freshman (52.0%; 28.6% sophomores; 12.7% juniors, 6.6% seniors, 0.2% did not indicate their year), and the 426 participants who provided their age were between 17 and 29 years old (*M*=18.75, *SD*=1.33). The majority of the sample identified their ethnicity as not Hispanic or Latino (88.6%) and their race as White (68.8; see Table 2 for more detail). The sample was close to evenly split between men (52.6%) and women (47.2%) One participant did not identify as either a woman or man and was eliminated from further analyses.

Measures

Value of domain. Participants responded to four items that were adapted from the devaluation subscale of Schmader, Major, and Gramzow's (2001) disengagement scale to refer to the importance of doing well in either science, technology, engineering, and math, or arts and

[3] When possible, "other" majors were reclassified into the 16 available categories.

humanities, depending on participants' condition (e.g., "Being good at [science, technology, engineering, and math] is an important part of who I am.") Participants indicated their agreement with each statement on a 7-point Likert scale, from strongly disagree to strongly agree. Two items that indicated that the domain in question was not important (e.g., "Success in [science, technology, engineering, and math] is not very valuable to me.") were reverse coded, and the four items were averaged together to create a scale in which higher values indicated that the participant placed greater importance on doing well in the particular domain ($\alpha = .888$).

Growth mindset. Participants responded to a shortened three-item version of a scale of beliefs about the malleability of intelligence (Dweck, 2000; e.g., "To be honest, you can't really change how intelligent you are.") Participants indicated their agreement with each statement on a 7-point Likert scale, from strongly disagree to strongly agree. Since all three items indicated a belief that intelligence was fixed, the items were averaged together and then reverse coded so that higher values on the scale would indicate having more of a growth mindset ($\alpha = .947$). The psychology sample completed this measure during prescreening at the beginning of the semester, while the engineering sample completed this measure in the lab, after completing the dependent variable measures and before responding to the manipulation and suspicion check.

Performance expectations. After completing Trial 1, but before receiving their scores, participants indicated how well they expected to do by responding to the question "What do you think your score was on this trial?" and choosing a number between 0 and 8. After participants completed all four trials and had received their scores, they indicated how well they expected to do on a future trial by responding to the question "If you were to complete another trial, of similar difficulty to the 4 trials you just completed, what score do you think you would get?" and choosing a number between 0 and 8.

Perceived task difficulty. After completing Trial 1, but before receiving their scores, participants indicated how difficult they thought the trial was on a 7-point Likert scale from very easy to very difficult.

Performance satisfaction. After participants completed all four trials and had received their scores, they indicated how satisfied they were with their total score on a 7-point Likert scale from very dissatisfied to very satisfied.

Performance attributions. After completing all four trials and receiving their scores, participants made attributions for their performance on the task by responding to the question: "How much do you think each of the following factors contributed to your performance on this task? For each factor, indicate the percent that you think it contributed to your performance. The total of all the factors should add up to 100." Participants indicated the percent that each of the following six factors contributed to their performance: task ease/difficulty, ability, effort, strategy, experience, and luck/random chance.

Performance state self-esteem. Participants completed the six-item performance subscale of Heatherton and Polivy's (1991) state self-esteem scale (e.g., "I feel confident about my abilities.") Participants indicated their agreement with each statement on a 7-point Likert scale, from strongly disagree to strongly agree. Three items that indicated low performance state self-esteem (e.g., I feel that I have less scholastic ability right now than others.) were reverse coded and averaged with the other items so that higher values on the scale indicate having higher performance state self-esteem ($\alpha = .861$).

Persistence. After completing all four trials, receiving their scores, and responding to the above measures (except for growth mindset, which was either answered during prescreening or at the end of the study), participants were given the opportunity to complete up to four additional

trials and to choose the level of difficulty they would like for the additional trial(s) on a 7-point Likert scale from much easier [than the trials just completed] to much more difficult [than the trials just completed]. Persistence was conceptualized as the number of additional trials selected, the level of difficulty requested, and the total amount of time spent on the additional trial(s).

Belonging. The engineering sample completed a 10-item measure of belonging in engineering or physics, depending on the participant's major, adapted from Walton and colleagues' (2015) measure of belonging in engineering (e.g., "I fit in well in [engineering/physics] at UConn.") Participants indicated their agreement with each statement on a 7-point Likert scale, from strongly disagree to strongly agree. Three items that indicated a lack of belonging (e.g., "I feel alienated from [engineering/physics] at UConn." were reverse coded and averaged together with the other items so that higher values on the scale would indicate greater belonging (α = .860).

Intentions to switch majors. The engineering sample indicated the likelihood that they would switch majors by the end of the year on a 5-point Likert scale from not at all likely to extremely likely.

Feedback or information that would increase interest in the task. After completing any additional trials and before the manipulation and suspicion check, the psychology sample responded to the open-ended question: "What kind of feedback about your performance or information about the task would make you more interested in trying the task again?"

Manipulation and suspicion check. At the end of the study, two questions were used to gauge participants' memory of the manipulation of condition and feedback. Specifically, participants were asked what the instructions said performance on the task predicted, with the options of GPA, success in STEM, success in arts and humanities, memory, or hand-eye

coordination, and they were asked what their total score on Trials 1-4 was out of 32 possible points. To gauge participants' suspicions regarding the manipulations, they then responded to an open-ended question regarding what they thought the purpose of the study was.

Procedures

Students in the psychology participant pool were eligible to participate if they had completed the growth mindset measure during prescreening at the beginning of the semester and indicated their gender as woman or man on the same prescreening survey. Information on the participant pool website indicated that the study involved completing a task on the computer and responding to questions about the task and oneself. Filters based on gender were used to alternate between allowing only women or men into the study in order to achieve a close to an even split between female and male participants. Participants from the psychology participant pool completed the study in the laboratory for research participation credit.

The engineering sample was recruited using announcements that indicated that participants had to be physics majors or in the engineering school, which includes computer science. The announcements directed students to a screener survey, which confirmed that they were in the engineering school or majoring in physics, and asked for their gender and information needed to schedule the laboratory session or send the online survey. Once sufficient numbers of male participants had signed up for the study, only female participants were allowed to sign up for the study. The engineering and physics students received $5 to $10 for completing the study, depending on the semester and whether they completed the study in the laboratory (N = 198) or online (N = 11).

All laboratory sessions were conducted by a white female experimenter. Participants were randomly assigned to the STEM or non-STEM condition and to receive success or failure

feedback. Domain was manipulated in the same manner as in Study 2, with instructions indicating that performance on the task predicts success in either science, technology, engineering, and math, or arts and humanities. The same attention check from Study 2 was used. Participants completed one practice trial and 4 regular trials of the same task used in Study 2, and the computer gave them their purported score, based on their feedback condition, after each of the regular trials. Those in the *failure* condition were told that their scores were 2, 3, 2, and 3, respectively, while those in the *success* condition were told that their scores were 5, 6, 5, and 6, respectively. After completing the first regular trial, but before receiving their score, participants responded to two questions about the trial (expected score and perceived difficulty). After completing all four trials, participants were reminded of their total score across the 4 trials and then completed the measures of performance satisfaction, expected future score, performance attributions, value of domain, and performance state self-esteem. They were then asked how many additional trials they wanted to complete and their preferred level of difficulty for those trials, and completed the number of additional trials they selected. All trials were of similar difficulty, no matter the level of difficulty the participant selected. Participants then completed the open-ended question. At this point, the engineering sample completed the measures of perceptions of majors for Study 1, as well as the questions that the psychology sample completed during prescreening and the measures of belonging and intentions to switch majors. Finally, all participants completed the manipulation and suspicion checks, and then read the debriefing on the computer. For laboratory sessions, the experimenter also verbally explained the most important parts of the debriefing and noted whether participants were interested in finding out their actual scores on the task, which were later emailed to interested participants. For the online

sessions, instructions were given as to how participants could obtain their actual scores if they were interested.

Results

Manipulation and Suspicion Check

Participants who incorrectly remembered the domain from the instructions (n = 19), were off by more than six points in their recollection of their total score (n = 5), or did not answer one of the manipulation check questions (n = 1) were dropped from further analyses. An additional 26 participants were dropped from further analyses due to suspicions regarding the true nature of the study. Along with dropping one participant from the engineering sample who did not indicate their gender as woman or man, this left a final analytic sample of 406 participants.

Value of Domain

Although participants completed the measure of value of domain after receiving their scores, there was no effect of feedback, F (1,390) < 0.01, p = .959, η_p^2 < .001. There was, however, a significant effect of domain, F (1,390) = 164.15, p < .001, η_p^2 = .296, and a significant effect of sample, F (1,390) = 21.98, p < .001, η_p^2 = .053, which were qualified by an interaction between domain and sample, F (1,390) = 10.20, p = .002, η_p^2 = .025. Analyzing the STEM and non-STEM domains separately indicated that the engineering sample valued STEM (M = 6.15, SD = 0.78) more than the psychology sample (M = 5.25, SD = 1.25), F (1,197) = 37.07, p < .001, η_p^2 = .158, but the was no effect of sample within the non-STEM domain, F (1,193) = 0.96, p = .329, η_p^2 = .005. Since participants' responses on this measure were not affected by their purported scores, this measure can be taken as a measure of trait-level valuing of the domain the participant was assigned to, and as such, will be used as a control variable in further analyses.

Growth Mindset

There were no significant differences between the two samples, F (1,401) = 0.19, p = .660, η_p^2 < .001, or between men and women in growth mindset, F (1,401) = 2.99, p = .085, η_p^2 = .007. The interaction between sample and gender was not significant, F (1,401) = 0.27, p = .603, η_p^2 = .001. Since the engineering sample completed the growth mindset measure after receiving their scores, further analyses were conducted to determine whether their responses were affected by condition. Growth mindset was not affected by domain, F (1,185) = 0.09, p = .759, η_p^2 = .001, feedback, F (1,185) = 0.11, p = .737, η_p^2 = .001, nor the interaction between domain and feedback, F (1,185) < 0.01, p = .959, η_p^2 < .001. Thus, growth mindset can be considered an individual difference and used to predict responses to success and failure.

Data Analytic Plan

Analyses on the dependent variables were conducted using hierarchical linear regression. Sample, domain, feedback, gender, growth mindset (mean-centered), and value of domain (control variable) were added in Step 1. The two-way interaction terms (sample x domain, sample x gender, sample x feedback, domain x gender, feedback x gender, feedback x growth mindset, and gender x growth mindset) were added in Step 2. The three-way interaction terms (sample x domain x gender, sample x domain x feedback, sample x feedback x gender, domain x feedback x gender, and feedback x gender x growth mindset) were added in Step 3. Finally, the four-way interaction term sample x domain x gender x feedback was added in the Step 4. For variables that were measured before participants received their scores (e.g., expected score), feedback and all interactions involving feedback were not included. Similarly, for variables that were only measured among the engineering sample, sample and all interactions involving sample were not included. Because the way attributions were measured means that attributions to each

individual factor is dependent on responses to the other factors (i.e., to increase an attribution to one factor, participants must decrease their attribution to another factor), multivariate regression was used for performance attributions. When the multivariate results were significant, univariate analyses were used to better understand the pattern of results. In addition, complete univariate analyses were conducted for attributions to effort and ability, even when the multivariate analyses were not significant, given my *a priori* hypotheses about attributions to effort and ability. Means and standard deviations by gender and condition are presented in Tables 3 through 9.

Expected Score

Main effects: There was a significant effect of gender, with women expecting lower scores than men (see Table 3), β = -.19, t (380) = -3.73, p < .001. The other main effects were not significant: domain, β = .06, t (380) = 1.03, p = .304, sample, β = .06, t (380) = 1.13, p = .259, or growth mindset, β = .06, t (380) = 1.11, p = .269.

Two-way interactions: None of the two-way interactions were significant: gender and mindset, β = .36, t (376) = 1.89, p = .060, sample and domain, β = -.13, t (376) = -1.55, p = .121, sample and gender, β = -.10, t (376) = -1.23, p = .220, and domain and gender, β = .02, t (376) = 0.18, p = .856.

Three-way interaction: Not significant (sample, domain, and gender, β = .04, t (375) = 0.31, p = .758).

Perceived Task Difficulty

Main effects: There were no main effects: Domain, β = -.05, t (398) = -0.79, p = .431, sample, β = -.04, t (398) = -0.85, p = .393, gender, β = .04, t (398) = 0.74, p = .459, or growth mindset, β = -.02, t (398) = -0.48, p = .634.

Two-way interactions: There was a significant interaction between gender and mindset, *β* = -.14, *t* (394) = -2.08, *p* = .038. Examining men and women separately indicated that growth mindset was non-significantly related to *greater* perceived difficulty among men, *β* = .07, *t* (209) = 0.94, *p* = .351, and non-significantly related to *decreased* perceived difficulty among women, *β* = -.14, *t* (185) = -1.88, *p* = .062. There were no other significant two-way interactions: sample and domain, *β* = .10, *t* (394) = 1.11, *p* = .267, sample and gender, *β* = .02, *t* (394) = 0.22, *p* = .827, and domain and gender, *β* = -.05, *t* (394) = -0.64, *p* = .525.

Three-way interaction: Not significant (sample, domain, and gender, *β* = -.13, *t* (393) = -1.01, *p* = .311).

Performance Satisfaction

Main effects: There was a significant effect of feedback, *β* = -.61, *t* (398) = -15.33, *p* < .001, with those in the failure condition less satisfied with their scores than those in the success condition (see Table 4). There were no other significant main effects: domain, *β* = -.08, *t* (398) = -1.73, *p* = .084, sample, *β* = -.01, *t* (398) = -0.27, *p* = .786, gender, *β* = -.02, *t* (398) = -0.46, *p* = .645, or growth mindset, *β* = 0.01, *t* (398) = 0.31, *p* = .759.

Two-way interactions: None of the two-way interactions were significant: gender and mindset, *β* = .10, *t* (390) = 1.86, *p* = .064, sample and domain, *β* = -.03, *t* (390) = -0.45, *p* = .652, sample and gender, *β* = -.10, *t* (390) = -1.60, *p* = .111, sample and feedback, *β* = -.04, *t* (390) = -0.67, *p* = .502, domain and gender, *β* = .11, *t* (390) = 1.58, *p* = .116, domain and feedback, *β* < .01, *t* (390) = 0.07, *p* = .948, feedback and gender, *β* = .04, *t* (390) = 0.68, *p* = .498, and feedback and mindset, *β* = -.05, *t* (390) = -0.83, *p* = .410.

Three-way interactions: None of the three-way interactions were significant: sample, domain, and gender, *β* = .01, *t* (385) = 0.12, *p* = .903, sample, domain, and feedback, *β* = -.04, *t*

(385) = -0.39, p = .700, sample, feedback, and gender, β = -.02, t (385) = -0.24, p = .813, domain, feedback, and gender, β = .03, t (385) = 0.30, p = .764, and feedback, gender, and growth mindset, β = -.06, t (385) = -0.83, p = .405.

Four-way interaction: Not significant (sample, domain, gender, and feedback, β = .13, t (384) = 0.95, p = .343).

Expected Future Score

There was a significant effect of feedback, β = -.79, t (397) = -26.25, p < .001, with those in the failure condition expecting to receive lower scores in the future compared to those in the success condition (see Table 4). There was also a significant effect of gender, β = -.13, t (397) = -4.29, p < .001, with women (M = 4.40, SD = 1.38) expecting to receive lower scores in the future compared to men (M = 4.70, SD = 1.44). There were no other significant main effects: sample, β = -.05, t (397) = -1.46, p = .145, domain, β = .01, t (397) = 0.36, p = .717, or growth mindset, β = .02, t (397) = .50, p = .615.

Two-way interactions: There was a significant interaction between gender and sample, β = .10, t (389) = 2.08, p = .038. Examining each sample indicated there was a significant effect of gender in the psychology sample, β = -.48, t (209) = -4.11, p < .001, with women (M = 4.38, SD = 1.39) expecting lower future scores compared to men (M = 4.83, SD = 1.48), but within the engineering sample, men (M = 4.55, SD = 1.38) did not expect significantly higher scores compared to women (M = 4.43, SD = 1.36), β = -.07, t (183) = -1.73, p = .086. The other two-way interactions were not significant: sample and domain, β > -.01, t (389) = -0.06, p = .954, sample and feedback, β = -.01, t (389) = -0.16, p = .873, domain and gender, β = -.05, t (389) = -0.96, p = .340, domain and feedback, β > -.01, t (389) = -0.03, p = .976, feedback and gender, β

$< .01$, *t* (389) = 0.06, *p* = .803, feedback and mindset, *β* = .01, *t* (389) = 0.25, *p* = .803, and gender and mindset, *β* = .01, *t* (389) = 0.31, *p* = .759.

Three-way interactions: None of the three-way interactions were significant: sample, domain, and feedback, *β* = -.13, *t* (384) = -1.75, *p* = .081, sample, domain, and gender, *β* = -.05, *t* (384) = -0.69, *p* = .492, sample, feedback, and gender, *β* = .05, *t* (384) = 0.65, *p* = .514, domain, feedback, and gender, *β* = -.10, *t* (384) = -1.34, *p* = .182, and feedback, gender, and growth mindset, *β* = .06, *t* (384) = 0.99, *p* = .325.

Four-way interaction: Not significant (sample, domain, gender, and feedback, *β* = -.10, *t* (383) = -1.04, *p* = .299).

Performance Attributions

Main effects: There was a significant effect of feedback, *F* (5,394) = 6.70, $p < .001$, η_p^2 = .078. Univariate analyses indicated that feedback was a significant predictor of attributions to strategy, *β* = -.21, *t* (398) = -4.31, $p < .001$, task difficulty, *β* = .11, *t* (398) = 2.31, *p* = .021, and luck, *β* = .18, *t* (398) = 3.58, $p < .001$, with participants more likely to attribute their performance to strategy when they succeeded and more likely to attribute their performance to task difficulty and luck when they failed (see Tables 5-6). Feedback was not a significant predictor for the other attributions: ability, *β* = .01, *t* (398) = 0.17, *p* = .866, effort, *β* = -.06, *t* (398) = -1.26, *p* = .210, or experience, *β* = .08, *t* (398) = 1.63, *p* = .104. There was also a significant effect of gender, *F* (5,394) = 3.89, *p* = .002, η_p^2 = .047. Univariate analyses indicated that gender was a significant predictor of attributions to ability, *β* = -.20, *t* (398) = -4.01, $p < .001$, and strategy, *β* = .10, *t* (398) = 1.99, *p* = .047, with women less likely to attribute their performance to ability and more likely to attribute their performance to strategy compared to men (see Tables 5-6). Gender was not a significant predictor for the other attributions: effort, *β* = .08, *t* (398) = 1.51, *p* = .131, task

difficulty, β = -.03, t (398) = -0.68, p = .500, experience, β = .02, t (398) = 0.36, p = .717, or luck, β = -.03, t (398) = -0.52, p = .600. None of the other main effects were significant: sample, F (5,394) = 1.24, p = .291, η_p^2 = .015, domain, F (5,394) = 0.77, p = .575, η_p^2 = .010, or growth mindset, F (5,394) = 0.79, p = .556, η_p^2 = .010.

Two-way interactions: None of the two-way interactions were significant: sample and domain, F (5,386) = 1.71, p = .131, η_p^2 = .022, sample and gender, F (5,386) = 0.53, p = .753, η_p^2 = .007, sample and feedback, F (5,386) = 1.81, p = .110, η_p^2 = .023, domain and gender, F (5,386) = 1.15, p = .333, η_p^2 = .015, domain and feedback, F (5,386) = 1.12, p = .349, η_p^2 = .014, feedback and gender, F (5,386) = 0.35, p = .882, η_p^2 = .005, feedback and mindset, F (5,386) = 1.55, p = .173, η_p^2 = .020, and gender and mindset, F (5,386) = 0.87, p = .500, η_p^2 = .011.

Three-way interactions: None of the three-way interactions were significant: sample, domain and gender, F (5,381) = 2.18, p = .056, η_p^2 = .028, sample, domain and feedback, F (5,381) = 1.36, p = .239, η_p^2 = .018, sample, feedback, and gender, F (5,381) = 0.68, p = .637, η_p^2 = .009, domain, feedback, and gender, F (5,381) = 1.78, p = .117, η_p^2 = .023, and feedback, gender, and mindset, F (5,381) = 0.42, p = .836, η_p^2 = .005.

Four-way interaction: Not significant (sample, domain, gender, and feedback, F (5,380) = 0.99, p = .422, η_p^2 = .013).

Ability. Main effects: There was a significant effect of gender, β = -.20, t (398) = -4.01, p < .001, with women less likely to attribute their performance to ability than men (see Table 5). The other main effects were not significant: sample, β = -.05, t (398) = -1.01, p = .314, domain, β < .01, t (398) < 0.01, p = .998, feedback, β = .01, t (398) = 0.17, p = .866, or growth mindset, β = -.02, t (398) = -0.47, p = .642.

Two-way interactions: None of the two way interactions were significant: sample and domain, β = .15, *t* (390) = 1.77, *p* = .078, sample and gender, β = -.03, *t* (390) = -0.40, *p* = .687, sample and feedback, β = -.08, *t* (390) = -1.08, *p* = .282, domain and gender, β = .03, *t* (390) = 0.37, *p* = .712, domain and feedback, β = -.11, *t* (390) = -1.34, *p* = .183, feedback and gender, β = -.01, *t* (390) = -0.14, *p* = .890, feedback and mindset, β = .04, *t* (390) = 0.60, *p* = .548, and gender and mindset, β = -.08, *t* (390) = -1.11, *p* = .269.

Three-way interactions: None of the three-way interactions were significant: sample, feedback, and gender, β = .20, *t* (385) = 1.66, *p* = .099, sample, domain, and gender, β = -.18, *t* (385) = -1.47, *p* = .143, sample, domain, and feedback, β = -.01, *t* (385) = -0.10, *p* = .918, domain, feedback, and gender, β = .13, *t* (385) = 1.08, *p* = .283, and feedback, gender, and growth mindset, β = -.05, *t* (385) = -0.51, *p* = .609.

Four-way interaction: Not significant (sample, domain, gender, and feedback, β = -.05, *t* (384) = -0.27, *p* = .786).

Effort. Main effects: None of the main effects were significant: sample, β = .10, *t* (398) = 1.91, *p* = .057, domain, β = .10, *t* (398) = 1.69, *p* = .091, gender, β = .08, *t* (398) = 1.51, *p* = .131, feedback, β = -.06, *t* (398) = -1.26, *p* = .210, or growth mindset, β = .04, *t* (398) = 0.83, *p* = .407.

Two-way interactions: None of the two-way interactions were significant: domain and feedback, β = .15, *t* (390) = 1.83, *p* = .068, sample and feedback, β = -.14, *t* (390) = -1.69, *p* = .091, sample and domain, β = -.04, *t* (390) = -0.41, *p* = .681, sample and gender, β = .07, *t* (390) = 0.86, *p* = .388, domain and gender, β = .07, *t* (390) = 0.78, *p* = .437, feedback and gender, β = .06, *t* (390) = 0.68, *p* = .498, feedback and mindset, β = -.08, *t* (390) = -1.09, *p* = .276, and gender and mindset, β = -.01, *t* (390) = -0.18, *p* = .857.

Three-way interactions: There was a significant three-way interaction between sample, domain, and feedback, $\beta = .31$, $t\ (385) = 2.49$, $p = .013$. Examining each sample separately revealed that the interaction between feedback and domain was not significant in psychology sample, $\beta = -.04$, $t\ (205) = -0.34$, $p = .735$, but it was significant in the engineering sample, $\beta = .35$, $t\ (178) = 2.96$, $p = .004$. As shown in Figure 1, engineering students made relatively similar attributions to effort regardless of the feedback they received when they were in the STEM condition, but in the non-STEM condition they attributed their performance more to effort when they succeeded. None of the other three-way interactions were significant: sample, domain, and gender, $\beta = .08$, $t\ (385) = 0.64$, $p = .525$, sample, feedback, and gender, $\beta = .04$, $t\ (385) = 0.37$, $p = .713$, domain, feedback, and gender, $\beta > -.01$, $t\ (385) = -0.02$, $p = .986$, and feedback, gender, and growth mindset, $\beta = -.11$, $t\ (385) = -1.13$, $p = .261$.

Four-way interaction: Not significant (sample, domain, gender, and feedback, $\beta = -.02$, $t\ (384) = -0.13$, $p = .897$).

Performance State Self-Esteem

Main effects: There was a significant effect of feedback, $\beta = -.12$, $t\ (396) = -2.40$, $p = .017$, with participants having lower self-esteem when they failed than when they succeeded (see Table 8). There was also a significant effect of gender, $\beta = -.11$, $t\ (396) = -2.26$, $p = .025$, with women having lower self-esteem than men (see Table 8). None of the other main effects were significant: domain, $\beta = -.11$, $t\ (396) = -1.83$, $p = .068$, sample, $\beta = -.08$, $t\ (396) = -1.51$, $p = .131$, or growth mindset, $\beta = .03$, $t\ (396) = 0.63$, $p = .530$.

Two-way interactions: There was a significant two-way interaction between gender and mindset, $\beta = .20$, $t\ (388) = 3.01$, $p = .003$. Examining men and women separately indicated that growth mindset did not have a significant effect for men, $\beta = -.11$, $t\ (206) = -1.57$, $p = .117$, but it

had a significant effect for women, β = .18, *t* (185) = 2.45, *p* = .015, with growth mindset increasing women's self-esteem, such that there was no gender difference in self-esteem for those with a growth mindset, as shown in Figure 2. None of the other two-way interactions were significant: sample and domain, β = -.09, *t* (388) = -1.07, *p* = .284, sample and gender, β = .10, *t* (388) = 1.23, *p* = .220, sample and feedback, β = -.05, *t* (388) = -0.55, *p* = .584, domain and gender, β = .05, *t* (388) = 0.59, *p* = .553, domain and feedback, β = -.04, *t* (388) = -0.44, *p* = .661, feedback and gender, β = -.06, *t* (388) = -0.71, *p* = .476, and feedback and mindset, β = .03, *t* (388) = 0.47, *p* = .641.

Three-way interactions: None of the three-way interactions were significant: sample, domain, and gender, β = -.11, *t* (383) = -0.85, *p* = .396, sample, domain, and feedback, β = -.12, *t* (383) = -0.95, *p* = .343, sample, feedback, and gender, β = -.16, *t* (383) = -1.35, *p* = .177, domain, feedback, and gender, β = -.09, *t* (383) = -0.74, *p* = .458, and feedback, gender, and growth mindset, β = -.06, *t* (383) = -0.63, *p* = .528.

Four-way interaction: The four-way interaction between sample, domain, gender, and feedback was significant, β = -.37, *t* (382) = -2.26, *p* = .024. Examining men and women separately indicated that the three-way interaction between sample, domain, and feedback was not significant among men, β = .16, *t* (201) = 0.93, *p* = .356, but was significant among women, β = -.32, *t* (176) = -1.88, *p* = .063. As shown in Figure 3, women in the engineering sample had lower performance state self-esteem following failure in the STEM condition, but their self-esteem was unaffected by failure in the non-STEM condition, and women in the psychology sample had similar self-esteem regardless of condition.

Persistence

Number of additional trials. Main effects: There was a significant effect of sample, β = .13, t (387) = 2.55, p = .011, with engineering students (M = 3.06, SD = 1.14) completing more additional trials than psychology students (M = 2.72, SD = 1.20). None of the other main effects were significant: domain, β = .09, t (387) = 1.44, p = .151, gender, β = -.07, t (387) = -1.36, p = .173, feedback, β = .04, t (387) = 0.70, p = .483, or growth mindset, β = .01, t (387) = 0.26, p = .792.

Two-way interactions: There was a significant two-way interaction between gender and mindset, β = -.15, t (379) = -2.13, p = .034. Examining men and women separately indicated that the effect of mindset was negatively, but not significantly related to the number of trials for women, β = -.10, t (179) = -1.34, p = .183, and was positively, but not significantly related to the number of trials for men, β = .12, t (203) = 1.73, p = .086. None of the other two-way interactions were significant: sample and domain, β = -.07, t (379) = -0.82, p = .413, sample and gender, β = .01, t (379) = 0.07, p = .947, sample and feedback, β = .14, t (379) = 1.65, p = .101, domain and gender, β = .02, t (379) = 0.17, p = .863, domain and feedback, β = -.04, t (379) = -0.49, p = .624, feedback and gender, β = -.08, t (379) = -1.02, p = .309, and feedback and mindset, β = -.02, t (379) = -0.27, p = .790.

Three-way interactions: None of the three-way interactions were significant: sample, domain, and gender, β = .14, t (374) = 1.12, p = .265, sample, domain, and feedback, β = -.14, t (374) = -1.09, p = .277, sample, feedback, and gender, β = .19, t (374) = 1.52, p = .131, domain, feedback, and gender, β = -.01, t (374) = -0.10, p = .919, and feedback, gender, and growth mindset, β = -.11, t (374) = -1.11, p = .270.

Four-way interaction: Not significant (sample, domain, gender, and feedback, β < .01, t (373) = 0.03, p = .980).

Difficulty of additional trials. Main effects: There was a significant effect of gender, β = -.24, t (384) = -5.26, p < .001, with women wanting easier trials compared to men (see Table 9). There was a significant effect of feedback, β = -.37, t (384) = -8.29, p < .001, with participants wanting easier trials after failure than after success (see Table 9). There was a significant effect of sample, β = .11, t (384) = 2.37, p = .018, with engineering students (*M* = 4.32, *SD* = 1.48) wanting more difficult trials compared to the psychology students (*M* = 3.94, *SD* = 1.45). The other main effects were not significant: domain, β = .04, t (384) = 0.76, p = .449, or growth mindset, β = -.01, t (384) = -0.30, p = .765.

Two-way interactions: There was a significant two-way interaction between feedback and gender, β = .16, t (376) = 2.11, p = .036. Examining men and women separately indicated that the effect of feedback was significant for both women, β = -.33, t (177) = -4.70, p < .001, and men, β = -.43, t (202) = -6.80, p < .001, but, as illustrated in Figure 4, the effect was larger for men, resulting in a larger gender difference in the success condition. None of the other two-way interactions were significant: sample and feedback, β = -.14, t (376) = -1.82, p = .070, sample and domain, β = .08, t (376) = 0.98, p = .326, sample and gender, β < .01, t (376) = 0.05, p = .964, domain and gender, β = .04, t (376) = 0.46, p = .646, domain and feedback, β = -.03, t (376) = -0.42, p = .673, feedback and mindset, β = .01, t (376) = 0.08, p = .939, and gender and mindset, β = .06, t (376) = 1.03, p = .302.

Three-way interactions: None of the three-way interactions were significant: sample, domain, and feedback, β = -.21, t (371) = -1.90, p = .058, sample, domain, and gender, β = .05, t (371) = 0.43, p = .667, sample, feedback, and gender, β = .11, t (371) = 0.94, p = .348, domain, feedback, and gender, β = -.06, t (371) = -0.51, p = .607, and feedback, gender, and growth mindset, β = .02, t (371) = 0.16, p = .870.

Four-way interaction: Not significant: (sample, domain, gender, and feedback, *β* = -.01, *t* (370) = -0.09, *p* = .928).

Time spent on additional trials. Main effects: There was a significant effect of sample, *β* = .27, *t* (384) = 5.48, *p* < .001, with engineering students (*M* = 105.11, *SD* = 76.98) spending more time on the additional trials than the psychology students (*M* = 69.25, *SD* = 49.90). There was a significant effect of feedback, *β* = .15, *t* (384) = 3.05, *p* = .002, with participants spending more time on the additional trials after failure than after success (see Table 9). There was a significant effect of domain, *β* = .16, *t* (384) = 2.73, *p* = .007, with participants spending more time on the additional trials in the STEM condition than in the non-STEM condition (see Table 9). The other main effects were not significant: gender, *β* = .04, *t* (384) = 0.82, *p* = .415, or growth mindset, *β* = .05, *t* (384) = 1.12, *p* = .264.

Two-way interactions: There was a significant two-way interaction between domain and feedback, *β* = .19, *t* (376) = 2.35, *p* = .019. Examining the STEM and non-STEM conditions separately indicated that the effect of feedback was not significant in the non-STEM condition, *β* = .04, *t* (188) = 0.57, *p* = .571, but was significant in the STEM condition, *β* = .22, *t* (191) = 3.29, *p* = .001, with participants in the STEM condition spending more time on the additional trials when they failed than when they succeeded (see Table 9). None of the other two-way interactions were significant: sample and feedback, *β* = .14, *t* (376) = 1.81, *p* = .072, sample and gender, *β* = .14, *t* (376) = 1.75, *p* = .081, sample and domain, *β* = 108, *t* (376) = 1.25, *p* = .212, domain and gender, *β* = -.13, *t* (376) = -1.54, *p* = .124, feedback and gender, *β* = -.11, *t* (376) = -1.35, *p* = .178, feedback and mindset, *β* = -.05, *t* (376) = -0.70, *p* = .486, and gender and mindset, *β* = -.07, *t* (376) = -1.04, *p* = .297.

Three-way interactions: None of the three-way interactions were significant: feedback, gender, and growth mindset, β = -.18, t (371) = -1.91, p = .057, sample, domain, and gender, β = .01, t (371) = 0.11, p = .915, sample, domain, and feedback, β = -.03, t (371) = -0.22, p = .829, sample, feedback, and gender, β = .15, t (371) = 1.29, p = .196, and domain, feedback, and gender, β = -.06, t (371) = -0.54, p = .591.

Four-way interaction: Not significant (sample, domain, gender, and feedback, β = -.08, t (370) = -0.52, p = .607).

Belonging

Main effects: None of the main effects were significant: growth mindset, β = -.13, t (183) = -1.82, p = .070, domain, β = -.10, t (183) = -0.99, p = .324, gender, β = -.01, t (183) = -0.16, p = .876, or feedback, β = .03, t (183) = 0.47, p = .639.

Two-way interactions: None of the two-way interactions were significant: domain and gender, β = .01, t (178) = 0.11, p = .916, domain and feedback, β = .06, t (178) = 0.45, p = .657, feedback and gender, β = -.01, t (178) = -0.01, p = .970, feedback and mindset, β = -.04, t (178) = -0.36, p = .718, and gender and mindset, β = -.04, t (178) = -0.45, p = .655.

Three-way interactions: Neither of the three-way interactions were significant: domain, feedback, and gender, β = -.22, t (176) = -1.25, p = .213, and feedback, gender, and growth mindset, β = -.10, t (176) = -0.66, p = .513.

Intentions to Switch Majors

Main effects: There was a significant effect of feedback, β = .18, t (182) = 2.40, p = .018, with engineering students indicating a greater likelihood of switching majors after failing on the task than after succeeding (see Table 10). None of the other main effects were significant:

domain, $\beta = .05$, $t\ (182) = 0.46$, $p = .646$, gender, $\beta = .02$, $t\ (182) = 0.32$, $p = .748$, or growth mindset, $\beta = .05$, $t\ (183) = 0.69$, $p = .493$.

Two-way interactions: None of the two-way interactions were significant: domain and gender, $\beta > -.01$, $t\ (177) = -0.03$, $p = .974$, domain and feedback, $\beta = -.13$, $t\ (177) = -1.08$, $p = .282$, feedback and gender, $\beta = -.02$, $t\ (177) = -0.17$, $p = .865$, feedback and mindset, $\beta = -.14$, $t\ (177) = -1.23$, $p = .219$, and gender and mindset, $\beta = -.02$, $t\ (177) = -0.24$, $p = .814$.

Three-way interactions: The three-way interaction between feedback, gender, and growth mindset was significant, $\beta = -.30$, $t\ (175) = -2.05$, $p = .042$. Examining men and women separately indicated that the interaction between feedback and growth mindset was not significant among men, $\beta = .05$, $t\ (95) = 0.35$, $p = .726$, but was significant among women, $\beta = -.39$, $t\ (79) = -2.34$, $p = .022$. As illustrated in Figure 5, growth mindset seemed to increase intentions to switch majors for women in the success condition. The three-way interaction between domain, feedback, and gender was not significant, $\beta = -.06$, $t\ (175) = -0.34$, $p = .733$.

Feedback or Information that would Increase Interest in the Task

Responses to the open-ended question from the psychology sample were examined by two trained research assistants. The lead research assistant first read through all of the responses and drafted coding categories and subcategories. The second research assistant then read through all of the responses along with the draft coding scheme, to ensure that all responses could be categorized, adding or editing categories as necessary. Both research assistants then read through the responses again, and used the finalized coding scheme to tally the number of responses to each category or subcategory. Since some participants mentioned multiple factors that would have made them more interested in completing the task again, the categories are not mutually exclusive. In addition, since the purpose of the coding was to get a general sense of the most

frequent responses in order to create a new intervention condition for Study 4, disagreements between the two research assistants in their frequency counts for each category were not reconciled. However, their disagreements only change the order of frequency of the categories for the less common categories. The two most common responses were 1) more detailed objective information on one's score or performance, such as being able to view where they clicked in relation to the targets, information on how far off they were from the targets, or seeing which targets they got correct; and 2) viewing their results compared to others, such as compared to the average, a successful student, or students from various majors. See Table 11 for the complete ordered list of categories.

Discussion

The hypothesis that participants would react more negatively to failure than success was supported. Participants in the failure condition were less satisfied with their scores, expected lower scores in the future, had lower performance state self-esteem, and requested easier additional trials. Participants in the engineering sample even had higher intentions to switch majors in the failure condition. There was also some evidence of a self-serving bias, with participants in the failure condition attributing their performance less to the strategy they used and more to the difficulty of the task and luck, compared to those in the success condition.

However, several of these effects of feedback were qualified by interactions with other factors, and there were interactions between feedback and other factors for dependent variables that did not have overall main effects of feedback. First of all, there was a four-way interaction between sample, domain, gender, and feedback for performance state self-esteem. Specifically, the interaction between domain, feedback, and sample was only significant among women, with women in the engineering sample having lower performance state self-esteem when they failed

than when they succeeded only in the STEM condition. This is one example of the negative effects of failure being specific to women in STEM. There was also a significant interaction between feedback and gender for requested difficulty of additional trials. Overall, participants asked for more difficult trials after success than after failure; the positive feedback likely made them feel ready to take on more of a challenge. However, this boost in requested difficulty was smaller for women. This may imply that success does not increase women's confidence in their ability to do well in the future or take on a challenge as much as it does for men.

There was one instance in which failure seemed to have a positive effect, at least for certain participants or under certain conditions. There was a significant interaction between feedback and domain for time spent on additional trials. Participants in the STEM condition tended to spend more time on the additional trials after failure, which could indicate an interest in improvement and greater persistence.

Contrary to hypotheses, participants did not view the task as more difficult or expect lower scores in the STEM condition, compared to the non-STEM condition. However, there were some gender differences that suggest a confidence gap. Specifically, women expected lower scores initially and lower future scores, had lower performance state self-esteem, and requested easier additional trials. A number of these main effects of gender were qualified by interactions with feedback, sample, and/or growth mindset. As mentioned previously, there was an interaction between feedback and gender for requested difficulty of additional trials, with the difference in requested difficulty being larger for men than for women, resulting in a larger gender difference in the success condition. The interaction between gender and sample for expected future score suggests that the confidence gap is larger among psychology students than engineering students. It is possible that the young women who suffer most from a confidence gap

choose not to major in engineering and physics (or dropped out of those majors early on) because they view those fields as more difficult and are not confident that they can do well in those fields.

One result seems to suggest that growth mindset might have particular benefits for women, effectively reducing gender gaps. Specifically, growth mindset predicted increased performance state self-esteem among women, such that the gender gap in self-esteem was not present among those with more of a growth mindset. However, there was also one area in which growth mindset may have had a negative effect. There was a significant three-way interaction between feedback, gender, and growth mindset for intentions to switch majors, which indicated that growth mindset tended to increase intentions to switch majors among women in the success condition. However, since the participants in the engineering condition came from a number of different majors, it is possible that these women were considering switching to a more difficult major or a major in which women are less represented.

The hypothesis that women would attribute success to effort and failure to ability more so than men in the STEM condition, in accordance with past research (Meece et al., 2006), was not supported. Women were less likely to attribute their performance to ability, but this effect occurred in both the success and failure conditions, and regardless of whether the task was purportedly associated with success in STEM or arts and humanities. In addition, there were no significant effects of gender or interactions between gender and other factors on attributions to effort. There was, however, an effect of gender on attributions to strategy, with women more likely to attribute their performance to strategy. Taken together with the effect of gender on ability, which is in the opposite direction, this suggests the possibility that women tend to endorse strategy instead of ability. Strategy may be seen as a more modest way of taking credit

for one's performance, compared to ability, since strategy is more about making the correct decision about how to approach a task or being intelligent about figuring out the best way to do something, rather than having innate ability.

Finally, the results for the open-ended question on what would increase interest in repeating the task can be used to develop an additional intervention condition for Study 4. The most common category of more information on one's performance might be specific to the exact task being used here—many of these responses referred to distance from the target or visual feedback on the task. For the same reason, an intervention based on that category of feedback would be specific to the exact task used in this study and could not be directly used on other tasks or contexts. Therefore, it makes more sense to use the second most common category, information comparing one's performance to others', as the basis for an additional intervention in Study 4. Specifically, participants can be given information on their percentile, along with their score, to demonstrate that even though they did poorly in an objective sense, compared to others, their performance was closer to average, and thus normative.

Chapter 9

Study 4: Intervention Including Norm Information with Failure Feedback

This study aims to build on the findings of Study 3 by focusing specifically on responses to failure, since those responses can be most damaging to persistence in a field, and examining responses solely on a STEM-related task, since that most directly relates to the overall purpose of the current work. In addition, this study examines whether including information that makes failure seem more normative increases resilient responses to failure. We specifically use two related but distinct interventions that, if successful, can be easily adapted to be feasibly used by instructors in the classroom. The *a priori* intervention involves telling participants that, when considering their score, they should be aware that most people initially do poorly on the task, but then improve with experience. This message not only conveys the point that low scores are normative, but also suggests the potential for improvement. The second intervention, which is based on participant feedback in Study 3, also makes participants' low scores seem normative, but does not include a message on potential for improvement. Specifically, participants are told that their scores put them at the 48th percentile—essentially indicating that their performance is average.

We hypothesize that both of these interventions will result in more positive reactions to failure, relative to the control condition of receiving one's score with no additional information. However, we also hypothesize that the impact of these interventions will vary by gender. Study 3 found that women react more negatively to failure than men in some respects. In addition, women face negative stereotypes regarding their competence in STEM. Thus, interventions that make their low performance seem normal could have a greater positive impact on women,

similar to how belonging interventions that make social struggles seem typical improve belonging for women in engineering (Walton et al., 2015).

Both interventions should make participants feel better about their low scores by making that level of performance seem normal. However, increased persistence will likely only follow if participants believe they can improve, which can either come from the intervention itself, in the case of the improvement intervention, or from a participant's own beliefs about the malleability of intelligence. Thus, we hypothesize that the impact of the interventions will vary by growth mindset, with the improvement intervention having a greater impact, relative to the percentile intervention and control condition, for those with more of a fixed mindset. Those with a growth mindset already have the message of potential for improvement internally, so the value added from having that message explicit in the intervention should be smaller for those participants. Additionally, Study 3 found that growth mindset reduced and even eliminated some of the existing gender gaps. Thus, we can expect interactions between feedback, gender, and growth mindset, such that the larger impact of the interventions for women compared to men only occurs among those with more of a fixed mindset.

Methods

Participants

Participants were 276 students from the psychology participant pool at the University of Connecticut who valued doing well in STEM, based on a median split of prescreening responses. Most of the participants were women (54.0%) and identified as White (64.5%) and not Hispanic or Latino (88.0%; see Table 12 for more detail). Most participants were freshmen (46.0%) or sophomores (31.5%; 18.1% juniors, 4.3% seniors), and the 275 participants who reported their age were between 17 and 28 (M =19.03, SD = 1.20). The most common majors (declared or

anticipated)[4] were in biological sciences (30.1%), health (23.3%), business (15.2%), and psychological sciences (10.1%).

Measures

Value of domain. During prescreening participants responded to the same value of domain measure that was used in Study 3, with all items referring to the importance of doing well in science, technology, and math ($\alpha = .496$). This measure was only used to screen students into the study. During the laboratory portion of the study, participants again completed this measure of value of domain, but this time the items referred to the importance of doing well on visual-spatial tasks, since the task was described as measuring visual-spatial skills ($\alpha = .841$).

Growth mindset. Participants responded during prescreening to the same measure of growth mindset as in Study 3, with the items averaged together and reverse coded so that higher values on the scale would indicate having more of a growth mindset ($\alpha = .944$).

Performance expectations. After completing the four trials, but before receiving their scores, participants indicated how well they expected to do by responding to the question "What do you think your score was on this task (out of a possible total of 32)?" and choosing a number between 0 and 32. Then after participants received feedback on their performance, they indicated how well they expected to do on a future trial using the same item as was used in Study 3.

Perceived task difficulty. After completing the four trials, but before receiving their scores, participants indicated how difficult they thought the task was on a 7-point Likert scale from very easy to very difficult.

Performance satisfaction. After participants completed all four trials and had received feedback on their performance, they indicated how satisfied they were with their total score on a

[4] When possible, "other" majors were reclassified into the 16 available categories.

7-point Likert scale from very dissatisfied to very satisfied. They also indicated their agreement on a 7-point Likert scale from strongly disagree to strongly agree with two statements on how they felt about their scores: "I am upset about my score on this task" and "I am happy with my score on this task." The item regarding being upset about one's scores was reversed, and the three items were averaged together ($\alpha = .762$)

Performance attributions. After completing all four trials and receiving feedback on their performance, participants made attributions for their performance on the task in the same manner as in Study 3.

Performance state self-esteem. Participants completed the performance state self-esteem scale used in Study 3 ($\alpha = .833$).

Reactions to task and performance. To better understand how participants felt about the task and the feedback they received, participants indicated their agreement with four statements on a 7-point Likert scale from strongly disagree to strongly agree. The statements were: 1) "I have very strong visual-spatial skills"; 2) "I can improve my visual-spatial skills"; 3) "I have what it takes to succeed in STEM"; and 4) "My score on this task tells me nothing about my ability to succeed in STEM."

Persistence. Persistence was measured in the same manner as in Study 3—based on the number and difficulty of the trials selected, as well as the amount of time spent on the additional trials.

Procedures

Students in the psychology participant pool were eligible to participate if they were above the median (5.00) on the value of domain measure included in prescreening, had completed the growth mindset measure, reported their SAT or ACT scores, and indicated their gender as

woman or man on the prescreening survey. Information on the participant pool website indicated that the study involved completing a task on the computer and responding to questions about the task and oneself. Filters based on gender were used to try to achieve a close to an even split between female and male participants. Participants from the psychology participant pool completed the study in the laboratory for research participation credit.

All laboratory sessions were conducted by a white female experimenter. For all participants, the instructions indicated that the task measures visual-spatial skills, which are important for success in STEM:

> "You are going to complete a task that measures visual spatial skills, which are important for success in science, technology, engineering, and math (STEM). You will complete 4 trials. On each trial, a set of 8 rectangles will appear on the screen. You should note the location and size of each rectangle. Then the rectangles will disappear and you will have to remember where they were and use the mouse to click on the previous location of each rectangle, so that if the rectangles were still visible, you would be clicking inside of each rectangle. You will have an opportunity to complete one practice trial before you begin the task."

A two-question attention check was used to make sure participants paid attention to both what the task measured and what those skills are important for. Participants completed one practice trial and 4 regular trials of the same task used in Studies 2 and 3 before the computer gave them their purported scores and additional feedback, based on their feedback condition. Participants were randomly assigned to one of three feedback conditions. All participants received scores indicative of failure (a total of 10 out of 32, based on the failure condition of Study 3). However, the feedback that accompanied the scores differed by condition. In the

control condition participants only received their score and no additional information. In the percentile condition, participants were told that their score put them at the 48^{th} percentile, meaning that they scored better than 48 percent of people who had completed the task. In the improvement condition, participants were told to keep in mind that most people do poorly at first on the task, but then tend to improve as they gain experience.

After completing the trials, but before receiving their scores or feedback, participants responded to two questions about the trial (expected score and perceived difficulty). After receiving the feedback according to their condition, participants completed the remaining measures. They were then asked how many additional trials they wanted to complete and their preferred level of difficulty for those trials, and completed the number of additional trials they selected. As in Study 3, all additional trials were of similar difficulty, no matter the level of difficulty the participant selected. Finally, all participants completed the manipulation and suspicion checks, and then read the debriefing on the computer. The experimenter also explained the most important parts of the debriefing and noted whether participants were interested in finding out their actual scores on the task, which were later emailed to interested participants.

Results

Manipulation and Suspicion Check

Participants who did not remember that the instructions said the task measured visual-spatial skills (n = 5) and that those skills are important for STEM (n = 4) or recalled that their scores were higher than they were told by more than six points (n = 9) were dropped from further analyses. In addition, those in the improvement condition who reported that there was no information provided about how experience affects performance on the task (n = 7), those in the percentile condition who reported that they performed over the 50^{th} percentile (n = 2), and any participants who mentioned during the suspicion check that they thought the purpose of the study

was related to poor performance (*n* = 12), were dropped from further analyses. There was some overlap between the reasons that participants were dropped, resulting in the elimination of 35 participants and a final analytic sample of 241 participants.

Value of Domain

Although participants completed the measure of value of domain after receiving their scores, there was no effect of feedback, F (2,235) = 2.15, p = .118, η_p^2 = .018, or gender, F (1,235) = 0.02, p = .890, η_p^2 < .001, and the interaction between feedback and gender was also not significant, F (2,235) = 0.61, p = .546, η_p^2 = .005. Since participants' responses on this measure were not affected by their feedback, this measure can be taken as a measure of trait-level valuing of the skills participants thought the task was testing, and as such, will be used as a control variable in further analyses.

Growth Mindset

Similar to Study 3, there was a marginally significant difference between men and women in growth mindset, F (1,239) = 3.25, p = .073, η_p^2 = .013, with men (M = 5.21, SD = 1.53) having a somewhat higher growth mindset than women (M = 4.85, SD = 1.53).

Data Analytic Plan

Analyses on the dependent variables were conducted using hierarchical linear regression. Feedback (dummy coded to compare each intervention to the control condition), gender, growth mindset (mean-centered), and value of domain (control variable) were added in Step 1. The two-way interaction terms (percentile feedback x gender, improvement feedback x gender, percentile feedback x growth mindset, improvement feedback x growth mindset, and gender x growth mindset) were added in Step 2. The three-way interaction terms (percentile feedback x gender x growth mindset and improvement feedback x gender x growth mindset) were added in Step 3.

For variables that were measured before participants received their scores (e.g., expected score), feedback and all interactions involving feedback were not included. When both interventions were significantly different from the control condition, the analysis was rerun with dummy coding that used the improvement intervention as the reference group in order to determine whether the two interventions were significantly different from each other. Because the way attributions were measured means that attributions to each individual factor is dependent on responses to the other factors (i.e., to increase an attribution to one factor, participants must decrease their attribution to another factor), multivariate regression was used for performance attributions. When the multivariate results were significant, univariate analyses were used to better understand the pattern of results. In addition, complete univariate analyses were conducted for attributions to effort and ability, even when the multivariate analyses were not significant, since attributions to effort and ability were central to the hypotheses. Means and standard deviations by gender and condition are presented in Tables 13 through 17.

Expected Score

Main effects: Neither of the main effects were significant: gender, β = -.05, t (237) = -0.76, p = .450, or growth mindset, β = .01, t (237) = 0.20, p = .844.

Two-way interaction: The interaction between gender and growth mindset was not significant, β = .05, t (236) = 0.53, p = .594.

Perceived Task Difficulty

Main effects: Neither of the main effects were significant: gender, β = -.07, t (237) = -1.13, p = .262, or growth mindset, β = .03, t (237) = 0.39, p = .699.

Two-way interaction: The interaction between gender and growth mindset was not significant, β = -.19, t (236) = -1.94, p = .053.

Performance Satisfaction

Main effects: None of the main effects were significant: gender, β = -.04, t (235) = -0.66, p = .513, growth mindset, β = .02, t (235) = 0.35, p = .724, percentile feedback, β = .08, t (235) = 0.79, p = .429, or improvement feedback, β = .02, t (235) = 0.22, p = .830.

Two-way interactions: The interactions between gender and improvement feedback, β = .24, t (230) = 2.00, p = .047, and between gender and percentile feedback, β = .24, t (230) = 2.08, p = .038, were both significant. As shown in Figure 6, examining the conditions separately indicated that there was a significant main effect of gender in the control condition, β = -.25, t (80) = -2.35, p = .021, with women less satisfied with their scores than men. However, the main effect of gender was not significant in the improvement condition, β = .10, t (75) = 0.88, p = .382, or the percentile condition, β = .06, t (74) = 0.55, p = .584. Switching the reference group indicated that the effect of gender was not significantly different between the improvement and percentile conditions, β = .01, t (230) = 0.08, p = .938. None of the other two-way interactions were significant: growth mindset and gender, β = -.14, t (230) = -1.44, p = .150, growth mindset and percentile feedback, β = .04, t (230) = 0.46, p = .643, and growth mindset and improvement feedback, β = -.09, t (230) = -0.89, p = .375.

Three-way interactions: Neither of the three-way interactions were significant: gender, growth mindset, and percentile feedback, β = -.18, t (228) = -1.35, p = .180, and gender, growth mindset, and improvement feedback, β = -.26, t (228) = -1.92, p = .056.

Expected Future Score

Main effects: There was a significant effect of gender, β = -.14, t (235) = -2.17, p = .031, with women estimating lower future scores than men (see Table 14). The other main effects were

not significant: growth mindset, *β* = -.01, *t* (235) = -0.21, *p* = .835, percentile feedback, *β* = .03, *t* (235) = 0.44, *p* = .662, or improvement feedback, *β* = .03, *t* (235) = 0.38, *p* = .703.

Two-way interactions: There was a significant interaction between gender and improvement feedback, *β* = -.24, *t* (230) = -2.00, *p* = .046. Examining the conditions separately indicated that there was a significant effect of gender in the improvement condition, *β* = -.37, *t* (75) =-3.36, *p* = .001, with women estimating lower scores than men, but there was no effect of gender in the percentile, *β* = -.14, *t* (74) = -1.19, *p* = .237, or control conditions, *β* = .02, *t* (80) = 0.15, *p* = .878, as illustrated in Figure 7. The other two-way interactions were not significant: gender and percentile feedback, *β* = -.14, *t* (230) = -1.16, *p* = .248, growth mindset and percentile feedback, *β* = -.03, *t* (230) = -0.35, *p* = .731, growth mindset and improvement feedback, *β* = -.05, *t* (230) = -0.50, *p* = .618, and gender and growth mindset, *β* = .10, *t* (230) = 1.04, *p* = .300.

Three-way interactions: Neither of the three-way interactions were significant: gender, growth mindset, and percentile feedback, *β* = -.15, *t* (228) = -1.09, *p* = .275, and between gender, growth mindset, and improvement feedback were not significant, *β* = -.08, *t* (226) = -0.56, *p* = .578.

Performance Attributions

Main effects: There was a significant effect of percentile feedback, *F* (5,231) = 2.63, *p* = .025, η_p^2 = .054. Univariate analyses indicated that percentile feedback was a significant predictor of attributions to ability, *β* = .23, *t* (235) = 3.12, *p* = .002, with those in the percentile condition attributing their performance more to ability than those in the control condition (see Table 15). Percentile feedback was not a significant predictor for the other attributions: effort, *β* = -.14, *t* (235) = -1.93, *p* = .055, strategy, *β* = .08, *t* (235) = 1.03, *p* = .306, task difficulty, *β* = -.08, *t* (235) = -1.11, *p* = .270, experience, *β* = -.05, *t* (235) = -0.63, *p* = .529, and luck, *β* = -.03, *t*

(235) = -0.39, p = .696. There was a significant main effect of improvement feedback, F (5,231) = 2.47, p = .034, η_p^2 = .051. Univariate analyses indicated that improvement feedback was a significant predictor of attributions to effort, β = -.16, t (235) = -2.19, p = .030, with participants in the improvement condition reporting effort as less important than those in the control condition, and to experience, β = .20, t (235) = 2.79, p = .006, with participants in the improvement condition attributing their performance more to experience than participants in the control condition (see Table 15). Improvement feedback was not a significant predictor for the other attributions: ability, β = .05, t (235) = 0.66, p = .513, strategy, β = .06, t (235) = 0.76, p = .447, task difficulty, β = -.02, t (235) = -0.30, p = .766, or luck, β = -.11, t (235) = -1.47, p = .144. Neither of the other main effects were significant in the multivariate analysis: gender, F (5,231) = 1.92, p = .093, η_p^2 = .040, or growth mindset, F (5,231) = 0.50, p = .775, η_p^2 = .011.

Two-way interactions: None of the two-way interactions were significant: gender and percentile feedback, F (5,226) = 1.55, p = .176, η_p^2 = .033, gender and improvement feedback, F (5,226) = 1.75, p = .124, η_p^2 = .037, growth mindset and percentile feedback, F (5,226) = 0.77, p = .576, η_p^2 = .017, growth mindset and improvement feedback, F (5,226) = 1.87, p = .100, η_p^2 = .040, and gender and growth mindset, F (5,226) = 0.64, p = .666, η_p^2 = .010.

Three-way interactions: Neither of the three-way interactions were significant: gender, growth mindset and improvement feedback, F (5,224) = 0.91, p = .477, η_p^2 = .020, or gender, growth mindset and percentile feedback, F (5,224) = 0.79, p = .557, η_p^2 = .017.

Ability. Main effects: As discussed above, there was a significant effect of percentile feedback, β = .23, t (235) = 3.12, p = .002, with those in the percentile condition attributing their performance more to ability than those in the control condition (see Table 15). The other main

effects were not significant: gender, β = -.04, t (235) = -0.58, p = .564, growth mindset, β = .01, t (235) = 0.11, p = .913, or improvement feedback, β = .05, t (235) = 0.66, p = .513.

Two-way interactions: None of the two-way interactions were significant: growth mindset and percentile feedback, β = .15, t (230) = 1.76, p = .080, gender and percentile feedback, β = -.03, t (230) = -0.24, p = .808, gender and improvement feedback, β = .09, t (230) = 0.75, p = .457, growth mindset and improvement feedback, β = -.03, t (230) = -0.30, p = .764, and growth mindset and gender, β = -.02, t (230) = -0.19, p = .849.

Three-way interactions: Neither of the three-way interactions were significant: gender, growth mindset, and percentile feedback, β = -.10, t (228) = -0.77, p = .444, and gender, growth mindset, and improvement feedback, β = -.11, t (228) = -0.79, p = .431.

Effort. Main effects: There was a significant effect of gender, β = .13, t (235) = 2.02, p = .044, with women reporting that effort was a more important contributor to their performance than men did. There was also a significant effect of improvement feedback, β = -.16, t (235) = -2.19, p = .030, with participants in the improvement condition reporting effort as less important than those in the control condition (see Table 15). Neither of the other main effects were significant: percentile feedback, β = -.14, t (235) = -1.93, p = .055, or growth mindset, β = -.01, t (235) = -0.22, p = .824.

Two-way interactions: None of the two-way interactions were significant: gender and percentile feedback, β = -.10, t (230) = -0.82, p = .411, gender and improvement feedback, β = -.01, t (230) = -0.10, p = .920, growth mindset and percentile feedback, β = -.04, t (230) = -0.49, p = .622, growth mindset and improvement feedback, β = .02, t (230) = 0.18, p = .855, and growth mindset and gender, β = .12, t (230) = 1.26, p = .209.

Three-way interactions: Neither of the three-way interactions were significant: gender, growth mindset, and percentile feedback, $\beta = .20$, t (228) = 1.48, $p = .141$, and gender, growth mindset, and improvement feedback, $\beta = .13$, t (228) = 0.90, $p = .367$.

Performance State Self-Esteem

Main effects: There was a significant effect of gender, $\beta = -.23$, t (235) = -3.65, $p < .001$, with women having lower self-esteem than men (see Table 14). There was also a significant effect of growth mindset, $\beta = .17$, t (235) = 2.67, $p = .008$, with participants with more of a growth mindset having higher self-esteem. Neither of the other main effects were significant: percentile feedback, $\beta = -.13$, t (235) = -1.81, $p = .072$, or improvement feedback, $\beta = -.09$, t (235) = -1.31, $p = .193$.

Two-way interactions: None of the two-way interactions were significant: gender and percentile feedback, $\beta = .05$, t (230) = 0.39, $p = .695$, gender and improvement feedback, $\beta = .06$, t (230) = 0.54, $p = .592$, growth mindset and percentile feedback, $\beta = .11$, t (230) = 1.33, $p = .186$, growth mindset and improvement feedback, $\beta = .06$, t (230) = 0.61, $p = .541$, and growth mindset and gender, $\beta = .09$, t (230) = 0.98, $p = .329$.

Three-way interactions: None of the three-way interactions were significant: gender, growth mindset, and percentile feedback, $\beta = .06$, t (228) = 0.43, $p = .667$, and gender, growth mindset, and improvement feedback, $\beta = .03$, t (228) = 0.22, $p = .828$.

Task Reactions

Strong visual spatial skills. Main effects: There was a significant effect of gender, $\beta = -.25$, t (232) = -3.95, $p < .001$, with women less likely to believe that they had strong visual spatial skills compared to men (see Table 16). None of the other main effects were significant: growth

mindset, β = -.02, t (232) = -0.31, p = .759, percentile feedback, β = .04, t (232) = 0.62, p = .538, or improvement feedback, β = .04, t (232) = 0.55, p = .581.

Two-way interactions: None of the two-way interactions were significant: gender and percentile feedback, β = .05, t (227) = 0.39, p = .697, gender and improvement feedback, β = -.13, t (227) = -1.08, p = .281, growth mindset and percentile feedback, β = -.01, t (227) = -0.12, p = .908, growth mindset and improvement feedback, β = -.03, t (227) = -0.28, p = .779, and growth mindset and gender, β = -.05, t (227) = -0.54, p = .590.

Three-way interactions: Neither of the three-way interactions were significant gender, growth mindset, and percentile feedback, β = .06, t (225) = 0.44, p = .663, and gender, growth mindset, and improvement feedback, β = -.21, t (225) = -1.57, p = .119.

Ability to improve visual spatial skills. Main effects: None of the main effects were significant: gender, β = .02, t (232) = 0.29, p = .773, growth mindset, β = .06, t (232) = 0.91, p = .364, percentile feedback, β = -.01, t (232) = -0.15, p = .878, or improvement feedback, β = .10, t (232) = 1.34, p = .182.

Two-way interactions: None of the two-way interactions were significant: gender and percentile feedback, β = .01, t (227) = 0.09, p = .927, gender and improvement feedback, β = -.19, t (227) = -1.56, p = .119, growth mindset and percentile feedback, β = -.07, t (227) = -0.80, p = .424, growth mindset and improvement feedback, β = -.07, t (227) = -0.66, p = .512, and growth mindset and gender, β = .06, t (227) = 0.62, p = .534.

Three-way interactions: Neither of the three-way interactions were significant: gender, growth mindset, and percentile feedback, β = -.10, t (225) = -0.69, p = .491, and gender, growth mindset, and improvement feedback, β = -.08, t (225) = -0.59, p = .559.

Ability to succeed in STEM. Main effects: There was a significant effect of gender, β = -.17, t (233) = -2.73, p = .007, with women less likely to believe that they had what it takes to succeed in STEM (see Table 16). There was also a significant effect of growth mindset, β = .16, t (233) = 2.52, p = .012, with participants with more of a growth mindset agreeing more that they have the ability to succeed in STEM. Neither of the other main effects were significant: percentile feedback, β = -.10, t (233) = -1.43, p = .154, or improvement feedback, β = -.05, t (233) = -0.62, p = .539.

Two-way interactions: There was a significant interaction between growth mindset and improvement feedback, β = .33, t (228) = 3.48, p = .001. Examining the improvement and control conditions separately indicated that growth mindset was significant in the improvement condition, β = .44, t (75) = 4.34, p < .001, with participants with more of a growth mindset agreeing more that they have what it takes to succeed in STEM, but it was not significant in the control condition, β = -.12, t (79) = -1.08, p = .285 (see Figure 8). None of the other two-way interactions were significant: growth mindset and percentile feedback, β = .14, t (227) = 1.66, p = .098, gender and percentile feedback, β = -.04, t (228) = -0.34, p = .733, gender and improvement feedback, β = -.08, t (228) = -0.67, p = .506, and growth mindset and gender, β = .01, t (228) = 0.07, p = .942.

Three-way interactions: Neither of the three-way interactions were significant: gender, growth mindset, and percentile feedback, β = .06, t (226) = 0.44, p = .663, and gender, growth mindset, and improvement feedback, β = .17, t (226) = 1.29, p = .198.

Relation between task and success in STEM. Main effects: There was a significant effect of gender, β = -.19, t (233) = -2.93, p = .004, with men more likely to agree that their score does not reflect their ability to succeed in STEM (see Table 16). None of the other main effects

were significant: growth mindset, *β* = -.05, *t* (233) = -0.79, *p* = .433, percentile feedback, *β* = -.04, *t* (233) = -0.56, *p* = .576, or improvement feedback, *β* = -.11, *t* (233) = -1.48, *p* = .140.

Two-way interactions: None of the two-way interactions were significant: growth mindset and gender, *β* = -.16, *t* (228) = -1.67, *p* = .097, gender and percentile feedback, *β* = .16, *t* (228) = 1.36, *p* = .177, gender and improvement feedback, *β* = .04, *t* (228) = 0.36, *p* = .719, growth mindset and percentile feedback, *β* = -.02, *t* (227) = -0.24, *p* = .808, and growth mindset and improvement feedback, *β* = -.03, *t* (228) = -0.35, *p* = .725.

Three-way interactions: Neither of the three-way interactions were significant: gender, growth mindset, and percentile feedback, *β* = -.04, *t* (226) = -0.31, *p* = .757, and gender, growth mindset, and improvement feedback, *β* = .06, *t* (226) = 0.47, *p* = .637.

Persistence

Number of additional trials. Main effects: None of the main effects were significant: gender, *β* = .03, *t* (228) = 0.50, *p* = .619, growth mindset, *β* = -.02, *t* (228) = -0.22, *p* = .825, percentile feedback, *β* = -.12, *t* (228) = -1.56, *p* = .120, or improvement feedback, *β* = -.06, *t* (228) = -0.83, *p* = .407.

Two-way interactions: None of the two-way interactions were significant: gender and percentile feedback, *β* = .19, *t* (223) = 1.54, *p* = .126, gender and improvement feedback, *β* = .04, *t* (223) = 0.31, *p* = .759, growth mindset and percentile feedback, *β* = .06, *t* (223) = 0.70, *p* = .482, growth mindset and improvement feedback, *β* = -.10, *t* (223) = -0.99, *p* = .325, and growth mindset and gender, *β* = .08, *t* (223) = 0.81, *p* = .420.

Three-way interactions: There was a significant three-way interaction between gender, growth mindset, and percentile feedback, *β* = -.37, *t* (221) = -2.65, *p* = .009. Examining men and women separately indicated that the interaction between growth mindset and percentile feedback

was significant among men, β = .33, t (100) = 2.64, p = .010, but not among women, β = -.12, t (120) = -0.97, p = .335. As illustrated in Figure 9, men in the percentile condition with more of a fixed mindset request fewer trials than other participants. The three-way interaction between gender, growth mindset, and improvement feedback was not significant, β = .15, t (223) = 1.06, p = .293.

Difficulty of additional trials. Main effects: There was a significant effect of gender, β = -.20, t (218) = -3.03, p = .003, with women selecting less difficult trials (see Table 17). The other main effects were not significant: growth mindset, β = .10, t (218) = 1.55, p = .124, percentile feedback, β = -.05, t (218) = -0.70, p = .482, or improvement feedback, β = -.06, t (218) = -0.75, p = .456.

Two-way interactions: None of the two-way interactions were significant: gender and percentile feedback, β = .18, t (213) = 1.44, p = .152, gender and improvement feedback, β = -.04, t (213) = -0.34, p = .735, growth mindset and percentile feedback, β = -.03, t (213) = -0.35, p = .727, growth mindset and improvement feedback, β = .16, t (213) = 1.64, p = .102, and growth mindset and gender, β = .02, t (213) = 0.23, p = .815.

Three-way interactions: Neither of the three-way interactions were significant: gender, growth mindset, and percentile feedback, β = -.08, t (211) = -0.54, p = .589, and gender, growth mindset, and improvement feedback, β = -.07, t (211) = -0.49, p = .626.

Time on additional trials. Main effects: None of the main effects were significant: gender, β = .05, t (220) = 0.79, p = .430, growth mindset, β = .01, t (220) = 0.20, p = .841, percentile feedback, β = .09, t (220) = 1.14, p = .255, or improvement feedback, β < .01, t (220) = 0.02, p = .983.

Two-way interactions: The interaction between growth mindset and improvement feedback was significant, $\beta = -.22$, $t\ (215) = -2.09$, $p = .038$. Examining the conditions separately indicated that growth mindset was significant in the control condition, $\beta = .23$, $t\ (75) = 2.03$, $p = .046$, with greater growth mindset predicting more time spent on the additional trials, but was not significant in the improvement, $\beta = -.15$, $t\ (72) = -1.25$, $p = .217$, or percentile conditions, $\beta = -.02$, $t\ (67) = -0.16$, $p = .872$. None of the other two-way interactions were significant: gender and percentile feedback, $\beta = .04$, $t\ (215) = 0.32$, $p = .751$, gender and improvement feedback, $\beta = -.01$, $t\ (220) = 0.06$, $p = .950$, growth mindset and percentile feedback, $\beta = -.13$, $t\ (220) = -1.40$, $p = .162$, and growth mindset and gender, $\beta = .06$, $t\ (220) = 0.58$, $p = .565$.

Three-way interactions: None of the three-way interactions were significant: gender, growth mindset, and percentile feedback, $\beta = -.08$, $t\ (218) = -0.52$, $p = .606$, and gender, growth mindset, and improvement feedback, $\beta = .05$, $t\ (218) = 0.32$, $p = .753$.

Discussion

Study 4 replicated the finding from Study 3 that women had lower performance state self-esteem compared to men. In Study 4, however, we also found that growth mindset predicted increased performance state self-esteem. This result was not found in Study 3, even in the failure and STEM conditions, which parallel the procedures of Study 4. Study 4 also replicated the gender difference in selected difficulty of additional trials, with women requesting easier trials compared to men. Expected patterns of gender differences were found in two of participants' reactions to the task and their scores. Specifically, women agreed less than men with the statements that they had strong visual-spatial skills and had the ability to succeed in STEM. Again, the findings from prior research of women attributing failure to ability more than men and

to effort less than men were not replicated. In fact, the opposite pattern was found for effort, with women attributing their performance to effort more than men.

We hypothesized that both interventions would improve reactions to low scores relative to the control condition, that these effects would be greater for women, and that the improvement intervention would be more beneficial than the percentile intervention for those with a fixed mindset. These hypotheses were generally not supported, with one exception. In the control condition, men were more satisfied with their performance than women were, but this gender difference disappeared in both intervention conditions. However, the attribution results demonstrated some negative impacts of the interventions. First of all, the improvement intervention decreased attributions to effort, which are posited to be to be the most beneficial attributions for failure because the internal, unstable nature of effort prompts greater effort in the future, resulting in improvement (Dweck, 1975; Weiner et al., 1972). Similarly, participants in the percentile condition made greater attributions to (lack of) ability compared to the control condition, which can lead to less effort and persistence in the future because ability is generally thought of as a stable attribution (Weiner et al., 1972). There was also one negative effect of the percentile intervention for men specifically, but only in combination with a fixed mindset: men in the percentile condition with a fixed mindset chose to complete fewer additional trials. In addition, in contrast to our prediction that a growth mindset would be most beneficial in the percentile condition, the benefits of a growth mindset seemed to either occur regardless of condition (i.e., higher performance state self-esteem) or occurred specifically in the improvement condition (i.e., greater agreement that one has what it takes to succeed in STEM).

Chapter 10

General Discussion

A prominent factor in women's underrepresentation in STEM, especially in the key fields of engineering, computer science, and physics, is women's preference for non-STEM careers and their tendency to drop out of STEM fields as they advance (Ceci et al., 2009). This dissertation aimed to explore how gender and domain differences in responses to success and failure, including performance attributions and expectations, might contribute to these tendencies by manipulating domain and performance feedback on a novel task. In addition, since perceptions of STEM and non-STEM majors play a role in the connection between failure and dropping out of STEM, this dissertation also empirically established that college STEM courses are indeed perceived as more difficult than non-STEM courses, and STEM introductory courses are perceiving as being used to cull students who will not be able to succeed in those fields more so than non-STEM introductory courses. Furthermore, this dissertation examined whether making failure seem more normative could promote more resilient responses to failure on a STEM-relevant novel task, particularly among women.

Summary of Findings

Perceptions of STEM and non-STEM majors. As expected, Study 1 demonstrated that college students from a variety of majors perceive courses in STEM majors as more difficult than those in non-STEM majors, both for courses in general and introductory courses in particular, and they also perceive introductory STEM courses as being used to eliminate students who do not have what it takes to succeed in those fields. In addition, women perceived STEM courses as more difficult and being used to eliminate students more than men, but this gender difference did not occur for perceptions of non-STEM courses, which is in line with past findings

on the confidence gap between men and women (Correll, 2001). Interestingly, among students from psychology courses—who almost all came from non-STEM majors or STEM majors in which women are well represented—this gender gap was present for both STEM and non-STEM courses, which was not the case for students majoring in engineering, computer science, and physics. This might suggest that the women who are least confident in their overall academic abilities are less likely to choose STEM majors in which women are most underrepresented.

Regardless of gender, the engineering, computer science, and physics students tended to perceive non-STEM courses as easier than the psychology students did. This difference in perception by participants' major may reflect differing levels of experience with non-STEM courses or it may be a self-esteem enhancement strategy in which students believe their own fields are more difficult.

Responses to success and failure. As expected, Study 3 found that participants generally reacted more positively to success compared to failure. Specifically, when participants were given very low scores, they were less satisfied with their scores, expected lower scores in the future, had lower performance state self-esteem, and requested easier additional trials. Additionally, some of their attributions demonstrated a self-serving bias, with participants who failed giving greater weight to the external factors of task difficulty and luck and less weight to the internal factor of strategy.

More interesting are the interactions between performance feedback and other factors, including gender. For example, a four-way interaction in Study 3 indicated that among women engineering and physics students, performance state self-esteem was lower following failure compared to success, but only in STEM condition. Failure is clearly most relevant to STEM students when they believe the task predicts STEM success, but the fact that men's self-esteem

was unaffected by feedback demonstrates the greater impact of failure in STEM on women, possibly because of the negative stereotypes they face in that domain. In addition, while participants generally requested more difficult trials following success feedback, this effect of feedback was smaller among women, which might suggest that success does not increase women's confidence in their ability to do well in the future or take on a challenge as much as it does for men, which could be contributing to the confidence gap. The overall gender effect was replicated in Study 4, with women requesting easier trials. Importantly, past findings on gender differences in attributions for success and failure in STEM (Meece et al., 2006) were not replicated in either Study 3 or Study 4. In fact, Study 4, which only involved failure, found that women actually made greater attributions to effort than men did.

There were also ways in which men and women reacted differently to the task, regardless of the type of feedback they received or before they had received any feedback. Specifically, in Study 3, women predicted that they had done worse on the task than men did, before they received any feedback, and after receiving feedback, women predicted that they would do worse on a future trial than men did. Furthermore, this gender gap in expected future scores was larger among psychology students, compared to students majoring in engineering, computer science, and physics. Similar to the findings on gender differences in perceptions of course difficulty, these results might suggest that women who suffer most from the confidence gap are choosing not to major in engineering, computer science, and physics, or are dropping out of those majors. In addition, while the gender gap in performance state self-esteem in Study 3 was qualified by an interaction indicating that only women majoring in engineering, computer science, and physics were negatively affected by failure, Study 4, which only involved psychology students, found that women reported lower self-esteem regardless of the type of failure feedback they received.

Study 4 also found that women were less likely than men to believe that they had strong visual-spatial skills and could succeed in STEM.

Growth mindset was not found to be reliably beneficial for women specifically. In Study 3, growth mindset predicted higher performance state self-esteem among women, thus reducing the gender gap among those with more of a growth mindset. However, in Study 4, this benefit of growth mindset for self-esteem was found for both men and women. In addition, growth mindset did not have benefits for any of the other outcomes.

Interventions to make failure seem normative. Although we hypothesized that both interventions would improve reactions to failure, particularly among women, this was not the case, with one exception. In the control condition, men were more satisfied with their performance than women were, but this gender difference disappeared in both intervention conditions. One other benefit of the improvement intervention only occurred among men, with men in the improvement condition expecting higher scores in the future. In addition, the interventions had some negative effects on attributions. The improvement decreased attributions to effort, which are generally thought to be beneficial attributions for failure because they are internal and unstable (Weiner et al., 1972). The percentile intervention also increased attributions to ability, which are seen as detrimental for persistence because it gives people the idea that they cannot improve (Weiner et al., 1972).

Our hypothesis that growth mindset would be most beneficial for participants in the percentile condition was also not supported. Instead, growth mindset had benefits for participants regardless of condition, including higher performance state self-esteem, or the benefits were specific to participants in the improvement condition, as was the case for believing that one has the ability to succeed in STEM.

Limitations

While the use of college student participants is appropriate given the overarching research question, and the recruitment of participants from the areas of STEM in which women are most underrepresented for Studies 1 and 3 is a strength of this dissertation, there are also limitations to the samples used. Studies 2 and 4 only used students from the psychology participant pool. While psychology is considered to be a STEM major by NSF (2017), it is an area in which women are overrepresented, and students may not view psychology as part of STEM. In addition, many of the students from the psychology participant pool are not psychology majors, with some coming from non-STEM majors. Even the students whose majors are broadly considered to be viewed as STEM, tend to be in STEM majors in which women are not underrepresented, such as biology. Thus, results may not generalize to students in STEM majors, particularly those where women are underrepresented, as indicated by some of the effects of sample seen in Studies 1 and 3. The use of solely psychology students in the pilot of the novel task in Study 2 has some potential implications for the results of Study 3, which manipulated success and failure based on the results of Study 2. It is possible that engineering, computer science, and physics majors would have a different definition of success and failure on the novel task compared to the sample of psychology students used in Study 2. Thus, although we consider the two levels of performance to represent success and failure, they may not be interpreted in that way by the engineering sample. However, the two feedback conditions still can be considered better and worse compared to each other, and thus comparison of the conditions is still useful in that regard.

Even within the engineering sample, there is a limitation in that not all of the participants are freshmen, and some of the freshmen completed the study toward the end of their freshmen

year. Therefore, the sample at least partially represented students who have stayed in engineering, computer science, and physics, and these students might be somewhat different from the larger group of students who begin their college career in those majors. Since the purpose of this dissertation relates to retention of women in STEM fields, that limitation is particularly relevant. For example, it is possible that the students who are retained in these male-dominated STEM fields are more resilient to failure than the students who dropped out, and that could partly explain some of the differences by sample. In addition, differences in the incentives between the two samples—extra credit for the psychology sample and money for the engineering sample—could have led to different levels of engagement. Furthermore, since the psychology students are enrolled in at least one psychology course and complete multiple studies over the course of the semester, they are likely more suspicious of the cover story than the engineering students, who rarely participate in this type of research. However, the use of attention, manipulation, and suspicion checks helps to overcome these particular limitations.

Another limitation comes from the novel task. While the novel nature removes the influence of expectations from prior experience, making the manipulation of performance more believable, the task may not have been particularly believable as predictive of STEM or arts and humanities success, as purported by the instructions in Studies 2 and 3. Twenty-six participants were dropped from analyses in Study 3 because their responses to the suspicion check indicated that they thought performance or domain had been manipulated. Quite a few other participants who were kept in analyses had indicated in the suspicion check that they thought the task was actually testing memory, which is not in line with the description of the task from the instructions. This interpretation of the task may have affected participants' responses since it could have changed how much they cared about their performance.

In addition, since the task does not actually test any skills, and certainly not any related to STEM, we are not able to examine how performance is affected by feedback, domain, gender, and growth mindset. Similarly, because the novel task is very different from the type of tasks that students generally complete in their courses, there are some issues with external validity, particularly since persistence was measured in terms of additional trials on the novel task. In a related vein, measures were all collected at the same time, immediately after success or failure, so it is not clear how long effects would last or how they accumulate over multiple performances.

Implications and Future Directions

Effects of sample and gender on perceptions of majors and future score expectations, with larger gender gaps within the psychology sample, imply that the women who choose non-STEM majors and STEM majors in which women are well represented perceive courses as more difficult and have lower expectations for their own performance, perhaps indicating low self-efficacy. Future research should further explore this relationship, including using longitudinal and experimental studies to determine if this lack of confidence in one's ability causes some women to shy away from engineering, computer science, and physics. Future research should also examine whether a similar pattern is found for non-STEM fields in which women are also underrepresented, such as philosophy, which also tend to be fields where a fixed mindset is commonplace (Leslie et al., 2015).

Several results indicate women have less confidence in their abilities, and a few results additionally suggest that for women, these views might be more negatively impacted by failure and less bolstered by success than they are for men. Future research should replicate these

findings and further explore the boundary conditions, as well as when this confidence gap develops, to determine how it can be reduced.

None of the results replicated past findings (Meece et al., 2006) on gender differences in attributions of effort and ability for success and failure. However, it is unclear whether this is due to changes over time in gender socialization, or to limitations of the prior methodology, with past studies not involving attributions made in the moment, or conflating performance with task difficulty. Since past research also found that teachers and parents make different attributions for boys' and girls' performance, which can then affect children's attitudes (Gunderson et al., 2012), future research should explore observers' attributions for students' performance by gender, particularly among teachers and parents.

The interventions to make failure seem normative that were tested in the current work were generally ineffective. It is possible that the percentile intervention did not do enough to make participants feel better about their performance and increase their persistence because students generally receive the message that being average is not enough. Therefore, finding out that they performed better than 48 percent of students does not change their view of their performance as failure, particularly since a grade of 59 or below out of 100 is generally considered a failing grade. Future research should examine whether a higher percentile, such as the 70^{th} percentile, is high enough to change perceptions of and reactions to a low absolute score on a task. In addition, implementing this type of percentile intervention in the classroom would mean that some students learn that their scores put them at the bottom of the class. Thus, future research should test whether that that approach is harmful for the lowest-performing students, and whether any increase in persistence among average or above average students, who may

otherwise have considering themselves to be low-performing, is enough to increase retention overall, particularly among students from underrepresented groups.

The improvement intervention was also not effective, potentially because participants expected that they should have already improved over the course of the four trials. It is possible then that such an intervention would be more effective if the scores indicated a gradual increase in performance, or if the feedback specified the amount of experience needed to improve or suggested strategies for improvement. Future research should examine whether such changes to the improvement intervention would make it more effective. Alternatively, the issue may lie not in the interventions themselves, but in the context in which they were applied. Students may not have been particularly engaged in the novel task because it is not directly relevant to their coursework, and they did not expect to complete it again after the end of the study. Thus, they may not have been interested in trying to improve. Future research should examine these interventions, as well as the variations mentioned above, in the field. Students could be randomly assigned to receive one or two sentences of feedback along with grades on their first assignment or exam in gateway STEM courses, and then their subsequent engagement in class and during office hours and their grades could be measured.

Another avenue for developing a better intervention to improve resilience among women in the face of failure is to directly encourage a growth mindset, since results indicated some positive effects of a growth mindset on performance state self-esteem, which was specific to women in Study 3 such that it narrowed the gender gap in performance state self-esteem. However, the fact that the benefits of a growth mindset were limited to performance state self-esteem and were found for both men and women in Study 4, suggests that growth mindset interventions might not be the best solution for gender gaps in STEM representation. In addition,

interventions to inculcate a growth mindset that have been tested in past research have been fairly involved (e.g., Good, Aronson, & Inzlicht, 2003), so it may make more sense to focus an in-depth intervention on a strategy with greater impacts.

Conclusion

Women are underrepresented in several important areas of STEM, including engineering, computer science, and physics. This dissertation proposed that common experiences of failure in STEM, combined with gender differences in responses to success and failure, particularly in terms of performance attributions, could partially explain women's underrepresentation, and that interventions to make failure seem more normative could be used to increase women's persistence in the face of failure. The current work confirmed that STEM courses are perceived as more difficult than non-STEM courses, and that STEM gateway courses in particular are perceived by college students as being used to cull students who would not be able to succeed in those fields. Furthermore, these perceptions were more pronounced among women, particularly those who had been recruited from psychology courses. The experimental studies in this dissertation found that, in contrast with past research (Meece et al., 2006), women did not attribute success more to effort and failure more to ability compared to men. However, there were other gender differences, suggestive of a confidence gap, particularly among psychology students and those with more of a fixed mindset. Neither normative intervention achieved their intended effect. Overall, the findings suggest that some women are less confident than men in their abilities, and this confidence gap, along with the perception of STEM courses as more difficult, may be a factor in these women's avoidance of engineering, computer science, and physics.

References

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

Blackwell, L.S., Trzesniewski, K.H., & Dweck, C.S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78,* 246-263.

Burnette, J.L., O'Boyle, E.H., VanEpps, E.M., Pollack, J.M., & Finkel, E.J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139,* 655-701.

Cech, E., Rubineau, B., Silbey, S., & Seron, C. (2011). Professional role confidence and gendered persistence in engineering. *American Sociological Review, 76,* 641-666.

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest, 15,* 75-141.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academies of Science, 108,* 3157-3162.

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135,* 218–261.

Corbett, C., & Hill, C. (2015). *Solving the equation: The variables for women's success in engineering and computing.* Washington, DC: American Association of University Women.

Correll, S.J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology, 106,* 1691-1730.

Crandall, V.C., Katkovsky, W., & Crandall, V.J. (1965). Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. *Child Development*, *36*, 91-109.

Diener, C.I., & Dweck, C.S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology*, *36*, 451-462.

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

Dweck, C.S. (2000). *Self-theories: Their role in motivation, personality and development.* Taylor & Francis: Philadelphia, PA.

Dweck, C.S. (2006). *Mindset: The new psychology of success.* New York: Random House.

Dweck, C.S. (2012). Implicit theories. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social psychology* (Vol. 2, pp. 43–61). Thousand Oaks, CA: Sage.

Dweck, C.S., Davidson, W., Nelson, S., & Enna, B. (1978). Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. *Developmental Psychology, 14,* 268-276.

Dweck, C. S., & Sorich, L. A. (1999). Mastery oriented thinking. In C. R. Snyder (Ed.), *Coping: The psychology of what works* (pp. 232–251). New York, NY: Oxford University Press.

Eccles, J.S., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J.L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J.T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75-146). San Francisco, CA: W.H. Freeman & Company.

Elliott, E.S., & Dweck, C.S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54,* 5-12.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24,* 645-662.

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.

Gunderson, E.A., Ramirez, G., Levine, S.C., & Beilock, S.L. (2012). The role of parents and teachers in the development of math related attitudes. *Sex Roles, 66,* 153-166.

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60,* 895-910.

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics.* Washington, DC: American Association of University Women.

Hong, Y.Y., Chiu, C., Dweck, C. S., Lin, D., & Wan, W. (1999) Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*(3), 588–599.

Hyde, J.S., & Mertz, J.E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, 106,* 8801-8807.

Lawner, E.K., & Quinn, D.M. (2016, January 28). *Responses to success and failure: The role of gender and domain.* Poster presented at the convention of the Society for Personality and Social Psychology, San Diego, CA.

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

Licht, B.G., & Dweck, C.S. (1984). Determinants of academic achievement: The interaction of children's achievement orientations with skill area. *Developmental Psychology, 20,* 628-636.

Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. *Journal of School Psychology, 44,* 351-373

Mok, M. M. C., Kennedy, K. J., & Moore, P. J. (2010). Academic attribution of secondary students: Gender, year level, and achievement level. *International Journal of Experimental Educational Psychology, 31,* 87-104.

Mosatche. H. S. (1977). *A social-cognitive analysis of the causal attributions made by children and adults.* (Unpublished doctoral dissertation). City University of New York: New York, NY.

National Science Foundation. (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017.* Arlington, VA: Author. Retrieved from https://www.nsf.gov/statistics/2017/nsf17310/

Nussbaum, D.A., & Dweck, C.S. (2008). Defensiveness vs. remediation: Self-theories and modes of self-esteem maintenance. *Personality and Social Psychology Bulletin, 34,* 599-612.

Robins, R.W., & Pals, J.L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, and self-esteem change. *Self and Identity, 1,* 313-336.

Schmader, T., Major, B., & Gramzow, R. H. (2001). Coping with ethnic stereotypes in the academic domain: Perceived injustice and psychological disengagement. *Journal of Social Issues, 57*, 93-111.

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

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1972). Perceiving the causes of success and failure. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* (pp.95-120). Morristown, NJ: General Learning Press.

Wiegers, R.M., & Frieze, I.H. (1977). Gender, female traditionality, achievement level, and cognitions of success and failure. *Psychology of Women Quarterly, 2,* 125-137.

Table 1

Demographic Characteristics of Participants in Study 1

Demographic Variable	N (%)
Race	
White	770 (71.4%)
Asian	182 (16.9%)
Black or African American	85 (7.9%)
Native American or Alaska Native	6 (0.5%)
Native Hawaiian or Other Pacific Islander	1 (0.1%)
Multiracial	61 (5.7%)
Did not answer	26 (2.4%)
Ethnicity	
Hispanic/Latino	118 (10.9%)
Not Hispanic/Latino	949 (88.0%)
Did not answer	11 (1.0%)
Gender	
Woman	605 (56.1%)
Man	461 (42.8%)
Gender-queer or gender-fluid	3 (0.3%)
Transgender Woman	1 (0.1%)
Transgender Man	1 (0.1%)
Other transgender	2 (0.2%)
Other gender	2 (0.2%)
Did not indicate gender	8 (0.7%)

Note. In the psychology sample, participants were allowed to check all that apply for race, including a multiracial option (called more than one race), while in the engineering, physics, and computer science sample, participants could only make one selection for race, including a multiracial option. Participants could check all that apply for gender, but wording varied between the samples and certain options were not available for both samples: gender-queer/gender-fluid and other gender were only available for the STEM sample; other transgender was only available for the psychology sample.

Table 2

Demographic Characteristics of Participants in Study 3

Demographic Variable	N (%)
Race	
White	315 (68.8%)
Asian	86 (18.8%)
Black or African American	23 (5.0%)
Native American or Alaska Native	0
Native Hawaiian or Other Pacific Islander	0
More than one race	27 (5.9%)
Did not answer	7 (1.5%)
Ethnicity	
Hispanic/Latino	48 (10.5%)
Not Hispanic/Latino	406 (88.6%)
Did not answer	4 (0.9%)

Note. Participants could only select one race.

Table 3

Means and Standard Deviations for Initial Scores Estimate and Perceived Difficulty as a Function of Domain and Participant Gender in Study 3

	Gender Domain	All	Women	Men
Estimated Score	All	5.02 (1.32) *N* = 386	4.76 (1.35) *N* = 185	5.26 (1.26) *N* = 201
	STEM	5.15 (1.38) *N* = 197	4.90 (1.44) *N* = 94	5.38 (1.28) *N* = 103
	Non-STEM	4.88 (1.25) *N* = 189	4.62 (1.24) *N* =91	5.13 (1.22) *N* = 98
Perceived Difficulty	All	4.08 (1.17) *N* = 404	4.12 (1.15) *N* = 190	4.04 (1.18) *N* = 214
	STEM	4.00 (1.25) *N* = 204	4.00 (1.22) *N* = 97	3.99 (1.29) *N* = 107
	Non-STEM	4.17 (1.07) *N* = 200	4.25 (1.06) *N* =93	4.09 (1.08) *N* = 107

Table 4

Means and Standard Deviations for Performance Satisfaction and Future Score Estimate as a Function of Domain, Participant Gender, and Feedback in Study 3

Variable	Gender	Feedback / Domain	All	Success	Failure
Performance Satisfaction	All	All	3.34 (1.38) *N* = 405	4.13 (1.23) *N* = 213	2.47 (0.97) *N* = 192
		STEM	3.20 (1.41) *N* = 205	3.95 (1.31) *N* = 113	2.29 (0.91) *N* = 92
		Non-STEM	3.49 (1.35) *N* = 200	4.33 (1.10) *N* = 100	2.64 (0.99) *N* = 100
	Men	All	3.36 (1.39) *N* = 214	4.21 (1.21) *N* = 110	2.47 (0.92) *N* = 104
		STEM	3.09 (1.41) *N* = 107	3.96 (1.32) *N* = 54	2.21 (0.84) *N* = 53
		Non-STEM	3.64 (1.31) *N* = 107	4.45 (1.06) *N* = 56	2.75 (0.94) *N* = 51
	Women	All	3.32 (1.38) *N* = 191	4.04 (1.24) *N* = 103	2.48 (1.02) *N* = 88
		STEM	3.33 (1.41) *N* = 98	3.93 (1.31) *N* = 59	2.41 (0.99) *N* = 39
		Non-STEM	3.31 (1.37) *N* = 93	4.18 (1.15) *N* = 44	2.53 (1.04) *N* = 49
Future Score Estimate	All	All	4.55 (1.42) *N* = 404	5.61 (0.76) *N* = 213	3.38 (0.98) *N* = 191
		STEM	4.68 (1.43) *N* = 204	5.65 (0.75) *N* = 113	3.46 (1.09) *N* = 91
		Non-STEM	4.43 (1.40) *N* = 200	5.56 (0.77) *N* = 100	3.30 (0.87) *N* = 100
	Men	All	4.69 (1.44) *N* = 213	5.78 (0.75) *N* = 110	3.53 (1.04) *N* = 103
		STEM	4.78 (1.45) *N* = 106	5.83 (0.75) *N* = 54	3.69 (1.16) *N* = 52
		Non-STEM	4.61 (1.43) *N* = 107	5.73 (0.75) *N* = 56	3.37 (0.87) *N* = 51
	Women	All	4.40 (1.38) *N* = 191	5.43 (0.74) *N* = 103	3.19 (0.88) *N* = 88
		STEM	4.56 (1.40) *N* = 98	5.49 (0.73) *N* = 59	3.15 (0.90) *N* = 39
		Non-STEM	4.23 (1.34) *N* = 93	5.34 (0.75) *N* = 44	3.22 (0.87) *N* = 49

Table 5

Means and Standard Deviations for Attributions to Strategy and Task Difficulty as a Function of Domain, Participant Gender, and Feedback in Study 3

Variable	Gender	Feedback Domain	All	Success	Failure
Strategy	All	All	26.48 (15.96) *N* = 405	29.70 (17.31) *N* = 213	22.91 (13.49) *N* = 192
		STEM	26.78 (15.99) *N* = 205	29.73 (17.23) *N* = 113	23.16 (13.56) *N* = 92
		Non-STEM	26.18 (15.96) *N* = 200	29.68 (17.48) *N* = 100	22.68 (13.48) *N* = 100
	Men	All	24.82 (15.93) *N* = 214	28.16 (17.94) *N* = 110	21.29 (12.63) *N* = 104
		STEM	24.10 (14.90) *N* = 107	25.72 (16.12) *N* = 54	22.45 (13.51) *N* = 53
		Non-STEM	25.54 (16.93) *N* = 107	30.52 (19.39) *N* = 56	20.08 (11.67) *N* = 51
	Women	All	28.35 (15.83) *N* = 191	31.35 (16.53) *N* = 103	24.83 (14.26) *N* = 88
		STEM	29.70 (16.70) *N* = 98	33.39 (17.53) *N* = 59	24.13 (13.76) *N* = 39
		Non-STEM	26.91 (14.81) *N* = 93	28.61 (14.83) *N* = 44	25.39 (14.77) *N* = 49
Task Difficulty	All	All	16.90 (12.91) *N* = 405	15.52 (12.08) *N* = 213	18.43 (13.65) *N* = 192
		STEM	17.00 (12.91) *N* = 205	15.71 (12.33) *N* = 113	18.59 (13.49) *N* = 92
		Non-STEM	16.80 (12.94) *N* = 200	15.31 (11.84) *N* = 100	18.28 (13.86) *N* = 100
	Men	All	17.11 (12.99) *N* = 214	15.58 (11.35) *N* = 110	18.73 (14.41) *N* = 104
		STEM	18.64 (13.88) *N* = 107	16.81 (12.37) *N* = 54	20.49 (15.15) *N* = 53
		Non-STEM	15.59 (11.91) *N* = 107	14.39 (10.24) *N* = 56	16.90 (13.50) *N* = 51
	Women	All	16.66 (12.85) *N* = 191	15.46 (12.87) *N* = 103	18.07 (12.76) *N* = 88
		STEM	15.21 (11.58) *N* = 98	14.69 (12.32) *N* = 59	16.00 (10.47) *N* = 39
		Non-STEM	18.18 (13.97) *N* = 93	16.48 (13.65) *N* = 44	19.71 (14.22) *N* = 49

Table 6

Means and Standard Deviations for Attributions to Luck and Ability as a Function of Domain, Participant Gender, and Feedback in Study 3

Variable	Gender	Feedback Domain	All	Success	Failure
Luck/Random Chance	All	All	11.10 (10.84) *N* = 405	9.24 (9.90) *N* = 213	13.16 (11.47) *N* = 192
		STEM	9.95 (9.98) *N* = 205	8.75 (10.32) *N* = 113	11.41 (9.39) *N* = 92
		Non-STEM	12.29 (11.56) *N* = 200	9.80 (9.42) *N* = 100	14.77 (12.94) *N* = 100
	Men	All	11.50 (11.76) *N* = 214	9.22 (10.88) *N* = 110	13.92 (12.22) *N* = 104
		STEM	10.64 (11.34) *N* = 107	9.80 (12.99) *N* = 54	11.49 (9.42) *N* = 53
		Non-STEM	12.37 (12.16) *N* = 107	8.66 (8.44) *N* = 56	16.45 (14.23) *N* = 51
	Women	All	10.65 (9.71) *N* = 191	9.27 (8.79) *N* = 103	12.26 (10.51) *N* = 88
		STEM	9.19 (8.23) *N* = 98	7.80 (7.04) *N* = 59	11.31 (9.46) *N* = 39
		Non-STEM	12.18 (10.90) *N* = 93	11.25 (10.46) *N* = 44	13.02 (11.31) *N* = 49
Ability	All	All	14.21 (10.40) *N* = 405	14.09 (10.29) *N* = 213	14.34 (10.55) *N* = 192
		STEM	14.25 (9.90) *N* = 205	14.64 (10.92) *N* = 113	13.77 (8.52) *N* = 92
		Non-STEM	14.17 (10.92) *N* = 200	13.47 (9.53) *N* = 100	14.86 (12.15) *N* = 100
	Men	All	16.12 (11.15) *N* = 214	15.96 (11.29) *N* = 110	16.29 (11.06) *N* = 104
		STEM	15.98 (10.35) *N* = 107	16.98 (12.43) *N* = 54	14.96 (7.67) *N* = 53
		Non-STEM	16.26 (11.95) *N* = 107	14.98 (10.08) *N* = 56	17.67 (13.68) *N* = 51
	Women	All	12.06 (9.05) *N* = 191	12.09 (8.72) *N* = 103	12.03 (9.48) *N* = 88
		STEM	12.36 (9.07) *N* = 98	12.49 (8.91) *N* = 59	12.15 (9.41) *N* = 39
		Non-STEM	11.75 (9.07) *N* = 93	11.55 (8.52) *N* = 44	11.94 (9.62) *N* = 49

Table 7

Means and Standard Deviations for Attributions to Effort and Experience as a Function of Domain, Participant Gender, and Feedback in Study 3

Variable	Gender	Feedback Domain	All	Success	Failure
Effort	All	All	21.64 (13.94) *N* = 405	22.49 (14.48) *N* = 213	20.70 (13.30) *N* = 192
		STEM	22.30 (14.28) *N* = 205	22.05 (13.32) *N* = 113	22.61 (15.45) *N* = 92
		Non-STEM	20.97 (13.59) *N* = 200	22.99 (15.73) *N* = 100	18.94 (10.75) *N* = 100
	Men	All	20.84 (14.09) *N* = 214	22.10 (15.85) *N* = 110	19.51 (11.87) *N* = 104
		STEM	21.13 (13.28) *N* = 107	21.26 (13.49) *N* = 54	21.00 (13.20) *N* = 53
		Non-STEM	20.55 (14.91) *N* = 107	22.91 (17.93) *N* = 56	17.96 (10.22) *N* = 51
	Women	All	22.54 (13.76) *N* = 191	22.91 (12.91) *N* = 103	22.10 (14.76) *N* = 88
		STEM	23.58 (15.27) *N* = 98	22.78 (13.24) *N* = 59	24.79 (18.02) *N* = 39
		Non-STEM	21.44 (11.96) *N* = 93	23.09 (12.59) *N* = 44	19.96 (11.28) *N* = 49
Experience	All	All	9.67 (9.38) *N* = 405	8.95 (8.58) *N* = 213	10.46 (10.15) *N* = 192
		STEM	9.72 (9.71) *N* = 205	9.12 (9.27) *N* = 113	10.46 (10.23) *N* = 92
		Non-STEM	9.61 (9.04) *N* = 200	8.75 (7.78) *N* = 100	10.47 (10.12) *N* = 100
	Men	All	9.60 (9.47) *N* = 214	8.97 (7.90) *N* = 110	10.26 (10.88) *N* = 104
		STEM	9.51 (9.47) *N* = 107	9.43 (8.52) *N* = 54	9.60 (10.44) *N* = 53
		Non-STEM	9.68 (9.50) *N* = 107	8.54 (7.30) *N* = 56	10.94 (11.39) *N* = 51
	Women	All	9.74 (9.30) *N* = 191	8.92 (9.30) *N* = 103	10.70 (9.26) *N* = 88
		STEM	9.95 (10.01) *N* = 98	8.85 (9.97) *N* = 59	11.62 (9.96) *N* = 39
		Non-STEM	9.53 (8.54) *N* = 93	9.02 (8.42) *N* = 44	9.98 (8.70) *N* = 49

Table 8

Means and Standard Deviations for Performance Self-Esteem and Number of Additional Trials Selected as a Function of Domain, Participant Gender, and Feedback in Study 3

Variable	Gender	Feedback Domain	All	Success	Failure
Performance State Self-Esteem	All	All	4.86 (1.06) N = 403	4.97 (1.03) N = 211	4.74 (1.08) N = 192
		STEM	4.75 (1.11) N = 204	4.86 (1.06) N = 112	4.61 (1.16) N = 92
		Non-STEM	4.97 (1.00) N = 199	5.10 (0.99) N = 99	4.85 (1.00) N = 100
	Men	All	4.97 (0.99) N = 212	5.06 (1.03) N = 108	4.88 (0948) N = 104
		STEM	4.82 (1.03) N = 106	4.88 (1.13) N = 53	4.76 (0.92) N = 53
		Non-STEM	5.13 (0.92) N = 106	5.23 (0.90) N = 55	5.01 (0.94) N = 51
	Women	All	4.73 (1.13) N = 191	4.88 (1.03) N = 103	4.57 (1.21) N = 88
		STEM	4.67 (1.20) N = 98	4.84 (1.01) N = 59	4.41 (1.41) N = 39
		Non-STEM	4.80 (1.06) N = 93	4.93 (1.08) N = 44	4.69 (1.03) N = 49
Number of Additional Trials	All	All	2.88 (1.18) N = 394	2.85 (1.22) N = 205	2.92 (1.15) N = 189
		STEM	3.02 (1.17) N = 199	3.01 (1.21) N = 109	3.02 (1.14) N = 90
		Non-STEM	2.74 (1.18) N = 195	2.67 (1.21) N = 96	2.82 (1.15) N = 99
	Men	All	2.95 (1.15) N = 209	2.86 (1.16) N = 108	3.05 (1.13) N = 101
		STEM	3.08 (1.17) N = 105	3.02 (1.21) N = 54	3.14 (1.15) N = 51
		Non-STEM	2.83 (1.11) N = 104	2.70 (1.11) N = 54	2.96 (1.11) N = 50
	Women	All	2.80 (1.22) N = 185	2.84 (1.28) N = 97	2.76 (1.15) N = 88
		STEM	2.95 (1.18) N = 94	3.00 (1.22) N = 55	2.87 (1.13) N = 39
		Non-STEM	2.65 (1.25) N = 91	2.62 (1.34) N = 42	2.67 (1.18) N = 49

Table 9

Means and Standard Deviations for Difficulty of and Time on Additional Trials as a Function of Domain, Participant Gender, and Feedback in Study 3

Variable	Gender	Feedback Domain	All	Success	Failure
Difficulty of Trials	All	All	4.12 (1.48) *N* = 391	4.66 (1.53) *N* = 203	3.54 (1.18) *N* = 188
		STEM	4.25 (1.43) *N* = 197	4.76 (1.40) *N* = 108	3.64 (1.21) *N* = 89
		Non-STEM	3.98 (1.52) *N* = 194	4.55 (1.66) *N* = 95	3.44 (1.15) *N* = 99
	Men	All	4.44 (1.60) *N* = 208	5.10 (1.59) *N* = 108	3.73 (1.28) *N* = 100
		STEM	4.53 (1.56) *N* = 104	5.19 (1.54) *N* = 54	3.82 (1.26) *N* = 50
		Non-STEM	4.36 (1.64) *N* = 104	5.02 (1.65) *N* = 54	3.64 (1.31) *N* = 50
	Women	All	3.75 (1.23) *N* = 183	4.16 (1.27) *N* = 95	3.32 (1.02) *N* = 88
		STEM	3.95 (1.19) *N* = 93	4.33 (1.10) *N* = 54	3.41 (1.12) *N* = 39
		Non-STEM	3.56 (1.25) *N* = 90	3.93 (1.46) *N* = 41	3.24 (0.95) *N* = 49
Time on Trials	All	All	86.33 (66.71) *N* = 391	77.78 (51.10) *N* = 203	95.55 (79.35) *N* = 188
		STEM	95.98 (76.04) *N* = 197	81.26 (53.43) *N* = 108	113.83 (93.91) *N* = 90
		Non-STEM	76.53 (54.13) *N* = 194	73.83 (48.28) *N* = 95	79.12 (59.33) *N* = 99
	Men	All	84.10 (60.55) *N* = 208	71.56 (42.13) *N* = 108	97.65 (73.42) *N* = 100
		STEM	99.12 (73.03) *N* = 104	78.80 (48.35) *N* = 54	121.08 (87.95) *N* = 50
		Non-STEM	69.08 (39.74) *N* = 104	64.32 (33.74) *N* = 54	74.21 (45.13) *N* = 50
	Women	All	88.85 (73.18) *N* = 183	84.86 (59.12) *N* = 95	93.17 (85.96) *N* = 88
		STEM	92.45 (79.52) *N* = 93	83.72 (58.42) *N* = 54	104.54 (101.44) *N* =39
		Non-STEM	85.13 (66.23) *N* = 90	86.35 (60.72) *N* = 41	84.12 (71.12) *N* =49

Table 10

Means and Standard Deviations for Belonging and Intentions to Switch Majors as a Function of Domain, Participant Gender, and Feedback in Study 3

Variable	Gender	Feedback Domain	All	Success	Failure
Belonging	All	All	5.08 (0.82) *N* = 189	5.06 (0.80) *N* = 98	5.11 (0.85) *N* = 91
		STEM	5.08 (0.83) *N* = 96	5.03 (0.78) *N* = 53	5.13 (0.88) *N* = 43
		Non-STEM	5.09 (0.82) *N* = 93	5.09 (0.82) *N* = 45	5.09 (0.83) *N* = 48
	Men	All	5.07 (0.74) *N* = 102	5.04 (0.74) *N* = 52	5.10 (0.75) *N* = 50
		STEM	5.06 (0.72) *N* = 52	4.94 (0.78) *N* = 27	5.19 (0.63) *N* = 25
		Non-STEM	5.09 (0.77) *N* = 50	5.16 (0.70) *N* = 25	5.02 (0.85) *N* = 25
	Women	All	5.10 (0.91) *N* = 87	5.07 (0.86) *N* = 46	5.12 (0.97) *N* = 41
		STEM	5.10 (0.95) *N* = 44	5.13 (0.79) *N* = 26	5.06 (1.16) *N* = 18
		Non-STEM	5.09 (0.89) *N* = 43	5.00 (0.97) *N* = 20	5.17 (0.82) *N* = 23
Intentions to Switch Majors	All	All	1.37 (0.75) *N* = 188	1.24 (0.56) *N* = 97	1.51 (0.90) *N* = 91
		STEM	1.34 (0.68) *N* = 95	1.27 (0.63) *N* = 52	1.42 (0.73) *N* = 43
		Non-STEM	1.40 (0.82) *N* = 93	1.20 (0.46) *N* = 45	1.58 (1.03) *N* = 48
	Men	All	1.36 (0.72) *N* = 102	1.23 (0.51) *N* = 52	1.50 (0.86) *N* = 50
		STEM	1.35 (0.65) *N* = 52	1.26 (0.53) *N* = 27	1.44 (0.77) *N* = 25
		Non-STEM	1.38 (0.78) *N* = 50	1.20 (0.50) *N* = 25	1.56 (0.96) *N* = 25
	Women	All	1.37 (0.80) *N* = 86	1.24 (0.61) *N* = 45	1.51 (0.95) *N* = 41
		STEM	1.33 (0.72) *N* = 43	1.28 (0.74) *N* = 24	1.39 (0.70) *N* =18
		Non-STEM	1.42 (0.88) *N* = 43	1.20 (0.41) *N* = 20	1.61 (1.12) *N* =23

Table 11

Frequency of Responses for Feedback or Information that Would Make Participants More Interested in Completing the Task Again

	Lead research assistant (N)	Second research assistant (N)
1.	More detailed objective information on own performance/scores (48)	More detailed objective information on own performance/scores (43)
2.	View results compared to others (29)	View results compared to others (29)
3.	Changes to feedback (27)	Changes to feedback (24)
4.	More information on exactly what abilities it's testing or correlated with (26)	More information on exactly what abilities it's testing or correlated with (23)
5.	Make it more interesting/fun (20)	Make it more interesting/fun (20)
6.	Provide a strategy for the next trials (16)	Provide a strategy for the next trials (16)
7.	Understanding what their data helps to achieve overall (12)	Timing (9)
8.	More information on what scores mean (9)	Other changes to task (8)
9.	Timing (6)	Understanding what their data helps to achieve overall (7)
10.	Receiving an incentive (6)	More information on what scores mean (6)
11.	Difficulty (5)	Difficulty (6)
12.	Other changes to task (5)	Receiving an incentive (5)

Table 12

Demographic Characteristics of Participants in Study 4

Demographic Variable	N (%)
Race	
White	178 (64.5%)
Asian	53 (19.2%)
Black or African American	15 (5.4%)
Native American or Alaska Native	1 (0.4%)
Native Hawaiian or Other Pacific Islander	2 (0.7%)
More than one race	22 (8.0%)
Did not answer	5 (1.8%)
Ethnicity	
Hispanic/Latino	32 (11.6%)
Not Hispanic/Latino	243 (88.0%)
Did not answer	1 (0.4%)

Table 13

Means and Standard Deviations for Initial Scores Estimate and Perceived Difficulty as a Function of Participant Gender in Study 4

	All (*N* = 241)	Women (*N* = 131)	Men (*N* = 110)
Estimated Score	20.90 (5.49)	20.65 (5.15)	21.21 (5.88)
Perceived Difficulty	4.65 (1.19)	4.56 (1.13)	4.75 (1.25)

Table 14

Means and Standard Deviations for Performance Satisfaction, Future Scores Estimates, and Performance State Self-Esteem as a Function of Participant Gender and Condition in Study 4

Variable	Condition Gender	All	Control	Percentile	Improvement
Satisfaction	All	3.12 (1.06)	3.02 (1.07)	3.24 (1.12)	3.11 (0.99)
		N = 241	*N* = 84	*N* =78	*N* =79
	Men	3.17 (1.11)	3.34 (1.10)	3.20 (1.09)	2.98 (1.13)
		N = 110	*N* = 36	*N* =37	*N* =37
	Women	3.08 (1.02)	2.78 (1.00)	3.28 (1.15)	3.22 (0.86)
		N = 131	*N* = 48	*N* =41	*N* =42
Future Estimate	All	4.73 (1.40)	4.68 (1.47)	4.76 (1.53)	4.76 (1.20)
		N = 241	*N* = 84	*N* =78	*N* =79
	Men	4.95 (1.31)	4.64 (1.22)	4.97 (1.48)	5.22 (1.16)
		N = 110	*N* = 36	*N* =37	*N* =37
	Women	4.55 (1.46)	4.71 (1.64)	4.56 (1.57)	4.36 (1.10)
		N = 131	*N* = 48	*N* =41	*N* =42
Performance State Self-Esteem	All	4.95 (1.05)	5.08 (1.08)	4.88 (1.02)	4.89 (1.05)
		N = 241	*N* = 84	*N* =78	*N* =79
	Men	5.23 (0.92)	5.41 (0.95)	5.15 (0.83)	5.12 (0.98)
		N = 110	*N* = 36	*N* =37	*N* =37
	Women	4.72 (1.10)	4.83 (1.11)	4.64 (1.13)	4.68 (1.09)
		N = 131	*N* = 48	*N* =41	*N* =42

Table 15

Means and Standard Deviations for Performance Attributions as a Function of Participant Gender and Condition in Study 4

Variable	Condition Gender	All	Control	Percentile	Improvement
Ability	All	14.56 (11.53) *N* = 241	12.18 (9.40) *N* = 84	18.13 (14.73) *N* =78	13.56 (8.98) *N* =79
	Men	15.11 (12.61) *N* = 110	13.00 (8.20) *N* = 36	19.76 (17.68) *N* =37	12.51 (8.36) *N* =37
	Women	14.09 (10.56) *N* = 131	11.56 (10.26) *N* = 48	16.66 (11.49) *N* =41	14.48 (9.50) *N* =42
Effort	All	21.81 (13.59) *N* = 241	24.82 (15.99) *N* = 84	20.40 (12.52) *N* =78	20.00 (11.22) *N* =79
	Men	19.77 (13.23) *N* = 110	22.03 (15.10) *N* = 36	19.38 (14.25) *N* =37	19.77 (13.23) *N* =37
	Women	23.52 (13.69) *N* = 131	26.92 (16.47) *N* = 48	21.32 (10.82) *N* =41	21.79 (12.15) *N* =42
Strategy	All	25.68 (13.93) *N* = 241	24.51 (15.60) *N* = 84	26.64 (13.32) *N* =78	25.96 (12.68) *N* =79
	Men	24.57 (13.36) *N* = 110	20.31 (15.16) *N* = 36	27.22 (13.24) *N* =37	26.08 (10.70) *N* =37
	Women	26.60 (14.38) *N* = 131	27.67 (15.32) *N* = 48	26.12 (13.54) *N* =41	25.86 (14.34) *N* =42
Task Difficulty	All	15.95 (10.83) *N* = 241	16.70 (11.38) *N* = 84	14.88 (8.72) *N* =78	16.19 (12.09) *N* =79
	Men	15.89 (11.79) *N* = 110	16.39 (11.96) *N* = 36	13.22 (8.63) *N* =37	18.08 (13.94) *N* =37
	Women	15.99 (10.00) *N* = 131	16.94 (11.05) *N* = 48	16.39 (8.62) *N* =41	14.52 (10.06) *N* =42
Experience	All	11.28 (11.70) *N* = 241	9.86 (11.43) *N* = 84	8.79 (9.43) *N* =78	15.24 (13.04) *N* =79
	Men	12.61 (11.98) *N* = 110	12.22 (12.33) *N* = 36	8.65 (9.58) *N* =37	16.95 (12.63) *N* =37
	Women	10.16 (11.38) *N* = 131	8.08 (10.49) *N* = 48	8.93 (9.42) *N* =41	13.74 (13.35) *N* =42
Luck/Random Chance	All	10.73 (11.74) *N* = 241	11.93 (12.81) *N* = 84	11.15 (12.64) *N* =78	9.05 (9.35) *N* =79
	Men	12.05 (12.74) *N* = 110	16.06 (15.17) *N* = 36	11.78 (14.11) *N* =37	8.41 (6.25) *N* =37
	Women	9.63 (1.10) *N* = 131	8.83 (9.77) *N* = 48	10.59 (11.29) *N* =41	9.62 (11.46) *N* =42

Table 16

Means and Standard Deviations for Task Reactions as a Function of Participant Gender and Condition in Study 4

Variable	Condition Gender	All	Control	Percentile	Improvement
Strong visual-spatial skills	All	3.78 (1.25) *N* = 238	3.73 (1.28) *N* = 83	3.79 (1.29) *N* =77	3.81 (1.18) *N* =78
	Men	4.12 (1.35) *N* = 107	4.03 (1.44) *N* = 35	4.00 (1.47) *N* =36	4.33 (1.12) *N* =36
	Women	3.50 (1.08) *N* = 131	3.52 (1.11) *N* = 48	3.61 (1.09) *N* =41	3.36 (1.06) *N* =42
Able to improve visual-spatial skills	All	5.77 (0.86) *N* = 238	5.74 (0.89) *N* = 82	5.69 (0.89) *N* =77	5.89 (0.80) *N* =79
	Men	5.77 (0.95) *N* = 107	5.65 (1.01) *N* = 34	5.58 (1.02) *N* =36	6.05 (0.74) *N* =37
	Women	5.78 (0.79) *N* = 131	5.81 (0.79) *N* = 48	5.78 (0.76) *N* =41	5.74 (0.83) *N* =42
Can succeed in STEM	All	5.43 (1.32) *N* = 239	5.54 (1.31) *N* = 83	5.32 (1.35) *N* =77	5.42 (1.31) *N* =79
	Men	5.70 (1.21) *N* = 108	5.71 (1.27) *N* = 35	5.58 (1.18) *N* =36	5.81 (1.20) *N* =37
	Women	5.21 (1.37) *N* = 131	5.42 (1.33) *N* = 48	5.10 (1.46) *N* =41	5.07 (1.31) *N* =42
Score tells nothing about ability to succeed in STEM	All	5.10 (1.45) *N* = 239	5.18 (1.34) *N* = 83	5.17 (1.46) *N* =77	4.96 (1.54) *N* =79
	Men	5.38 (1.49) *N* = 108	5.66 (1.26) *N* = 35	5.28 (1.52) *N* =36	5.22 (1.65) *N* =37
	Women	4.88 (1.38) *N* = 131	4.83 (1.31) *N* = 48	5.07 (1.42) *N* =41	4.74 (1.42) *N* =42

Table 17

Means and Standard Deviations for Persistence as a Function of Participant Gender and Condition in Study 4

Variable	Condition Gender	All	Control	Percentile	Improvement
Number of trials	All	2.39 (1.28) *N* = 234	2.57 (1.32) *N* = 82	2.20 (1.27) *N* =75	2.38 (1.23) *N* =77
	Men	2.34 (1.30) *N* = 107	2.66 (1.26) *N* = 35	1.97 (1.30) *N* =36	2.39 (1.29) *N* =36
	Women	2.43 (1.26) *N* = 127	2.51 (1.38) *N* = 47	2.41 (1.23) *N* =39	2.37 (1.18) *N* =41
Difficulty of trials	All	3.62 (0.88) *N* = 224	3.71 (0.76) *N* = 78	3.58 (0.71) *N* =71	3.56 (1.12) *N* =75
	Men	3.82 (0.96) *N* = 102	3.94 (0.84) *N* = 35	3.59 (0.71) *N* =32	3.91 (1.22) *N* =35
	Women	3.44 (0.77) *N* = 122	3.51 (0.63) *N* = 43	3.56 (0.72) *N* =39	3.25 (0.93) *N* =40
Time on trials	All	88.07 (65.39) *N* = 226	84.88 (59.97) *N* = 79	96.12 (76.22) *N* =71	83.88 (59.78) *N* =76
	Men	84.50 (61.96) *N* = 103	82.99 (64.36) *N* = 35	89.03 (63.25) *N* =32	81.95 (59.94) *N* =36
	Women	91.06 (68.23) *N* = 123	86.38 (59.95) *N* = 44	101.93 (85.79) *N* =39	85.61 (60.35) *N* =40

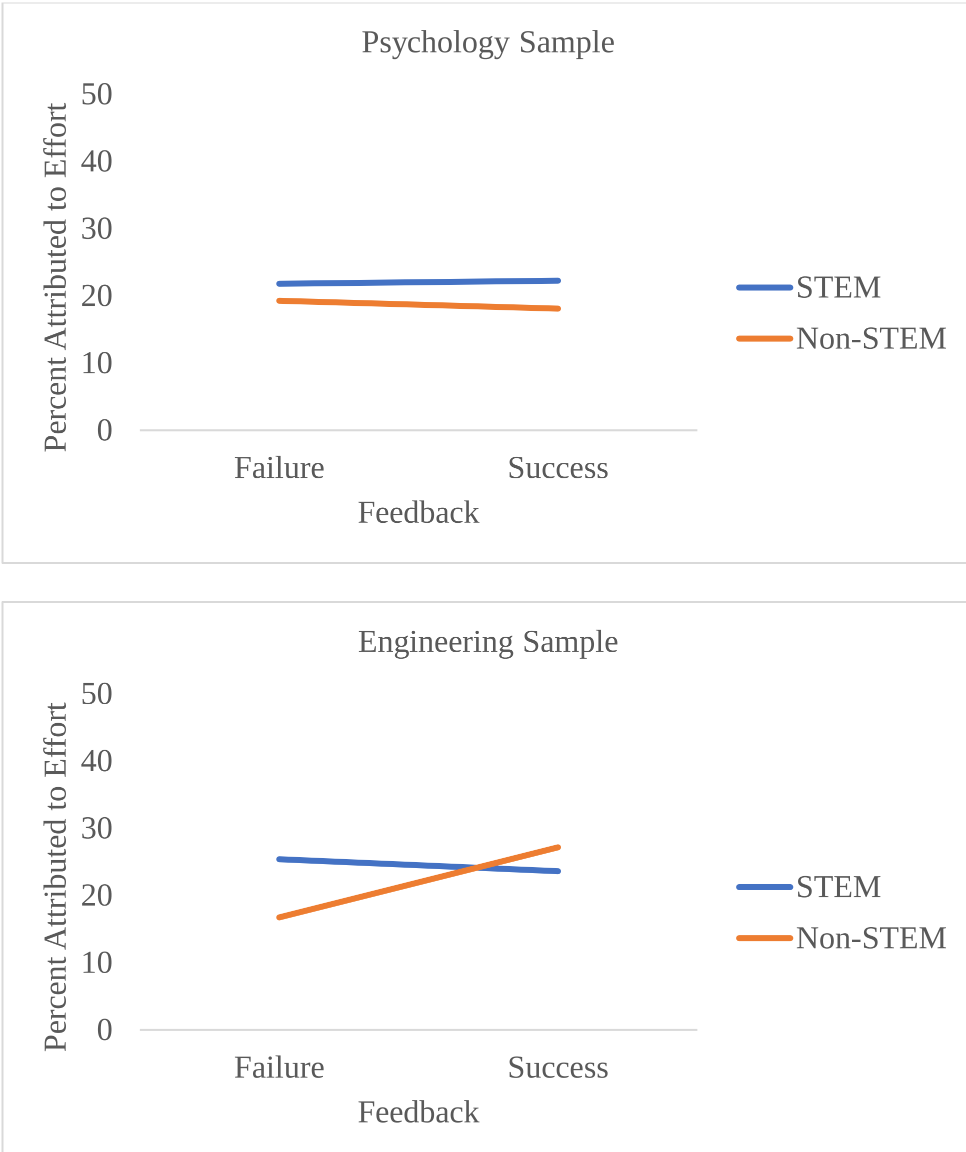

Figure 1. Predicted values for attributions to effort by sample, domain, and feedback, controlling for gender, growth mindset, and value of domain in Study 3. Predicted values are based on the third step of the hierarchical regression model, which includes the three-way interactions.

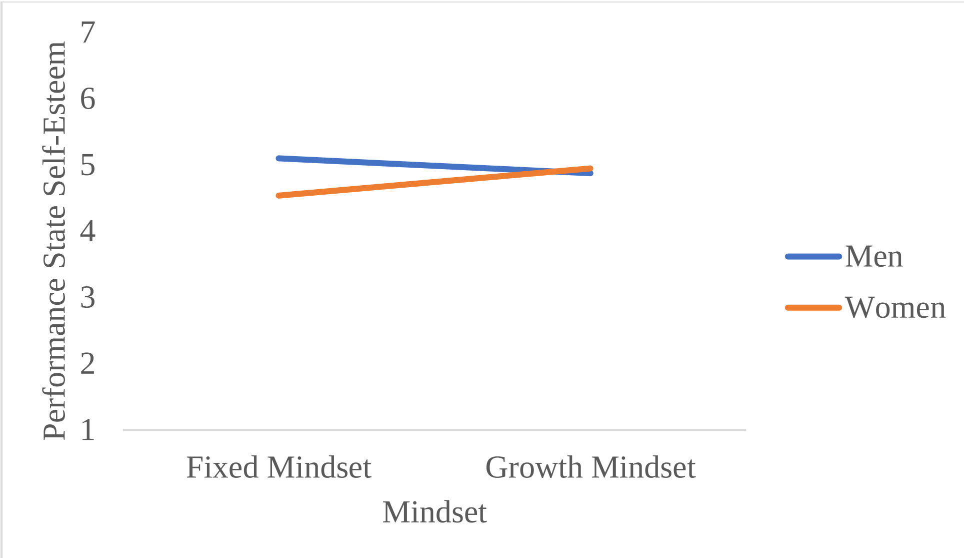

Figure 2. Predicted values for performance state self-esteem by growth mindset and gender, controlling for feedback, sample, domain, and value of domain in Study 3. Predicted values are based on the second step of the hierarchical regression model, which includes the two-way interactions. Fixed mindset is one standard deviation below the mean, and growth mindset is one standard deviation above the mean.

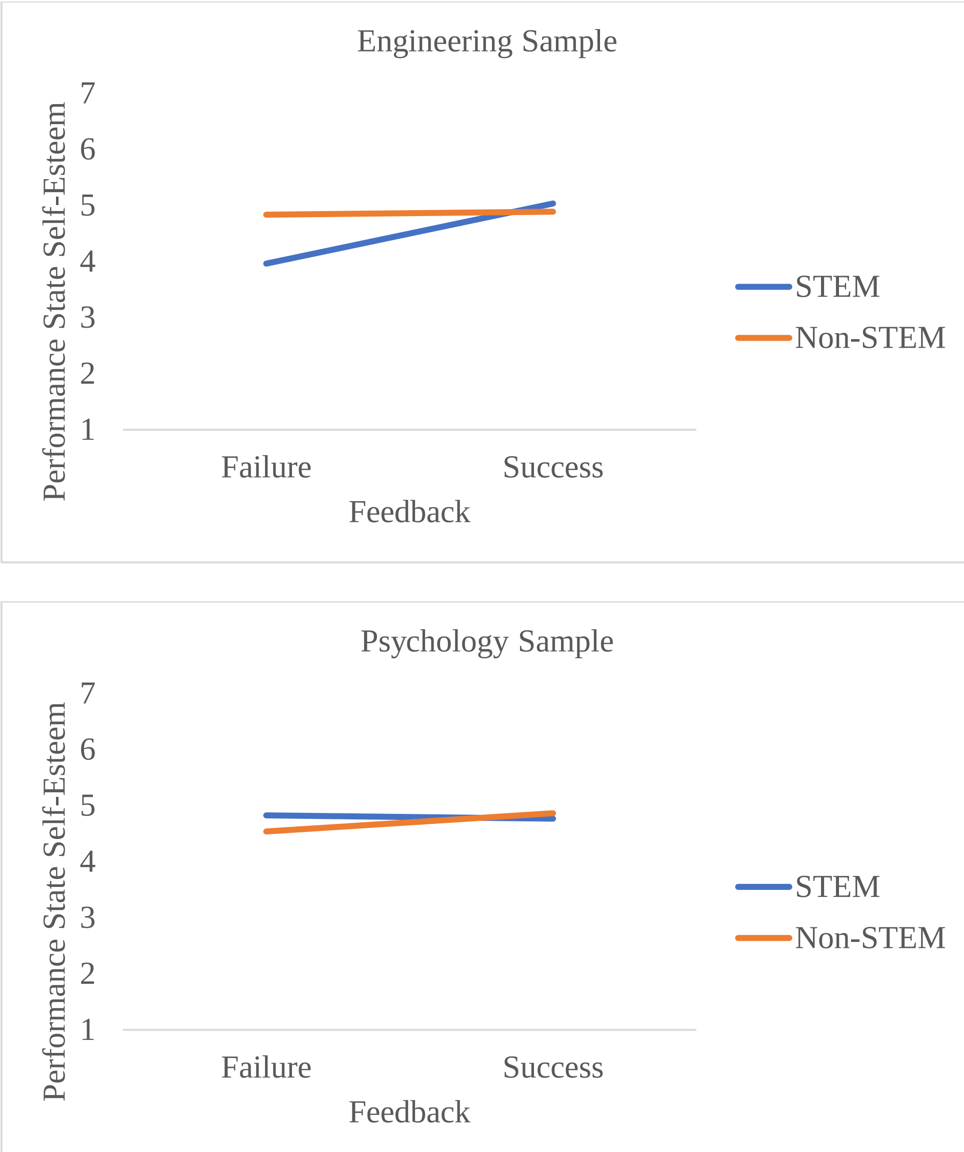

Figure 3. Predicted values for performance state self-esteem among women, by domain, feedback, and sample, controlling for growth mindset and value of domain in Study 3. Predicted values are based on the third step of the hierarchical regression model that was conducted solely on women and includes the three-way interaction and excludes gender and all interactions with gender as predictors.

Figure 4. Predicted values for requested difficulty of additional trials by feedback and gender, controlling for sample, domain, growth mindset, and value of domain in Study 3. Predicted values are based on the second step of the hierarchical regression model, which includes the two-way interactions.

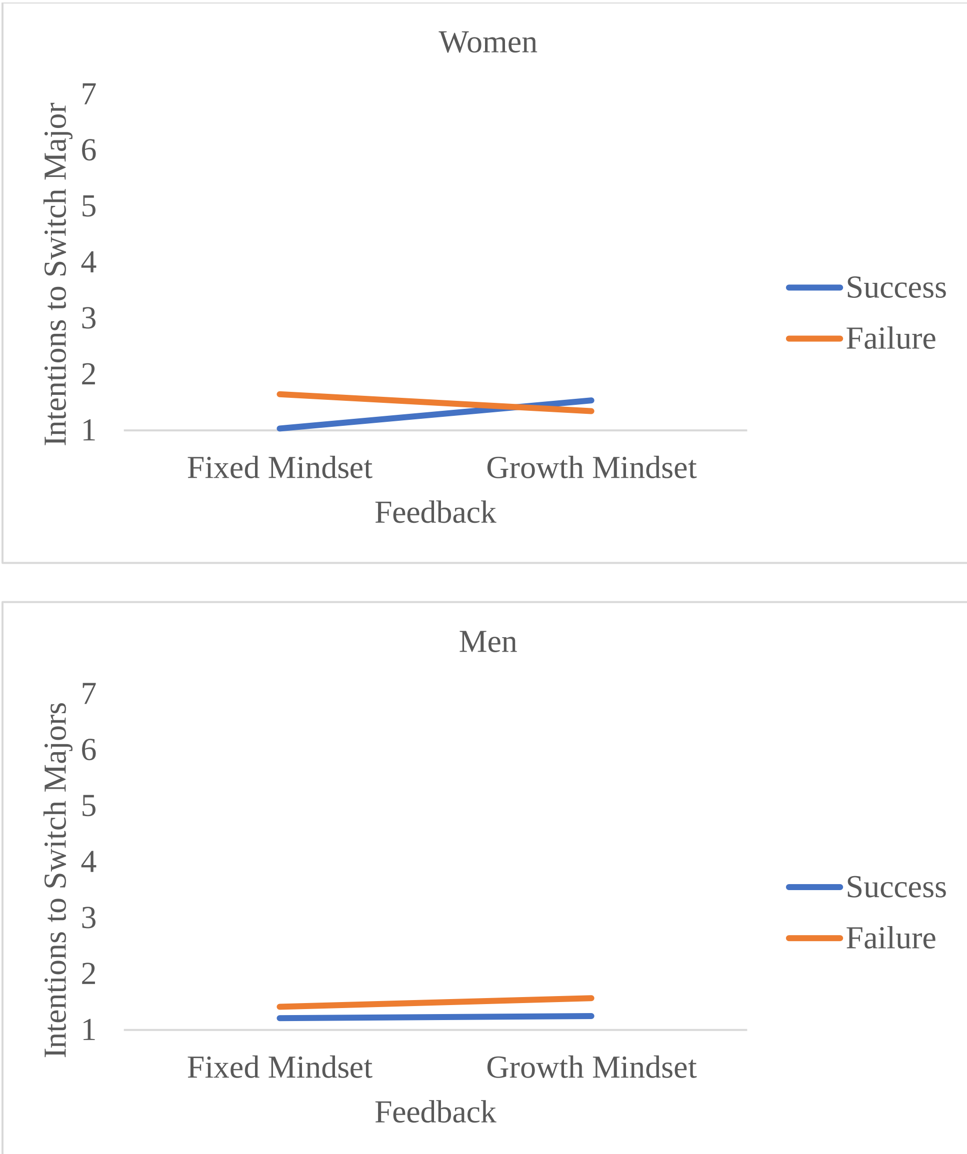

Figure 5. Predicted values for intentions to switch majors by gender, feedback, and growth mindset, controlling for domain and value of domain in Study 3. Predicted values are based on the third step of the hierarchical regression model, which includes the three-way interactions. Fixed mindset is one standard deviation below the mean, and growth mindset is one standard deviation above the mean. This analysis is specific to the engineering sample.

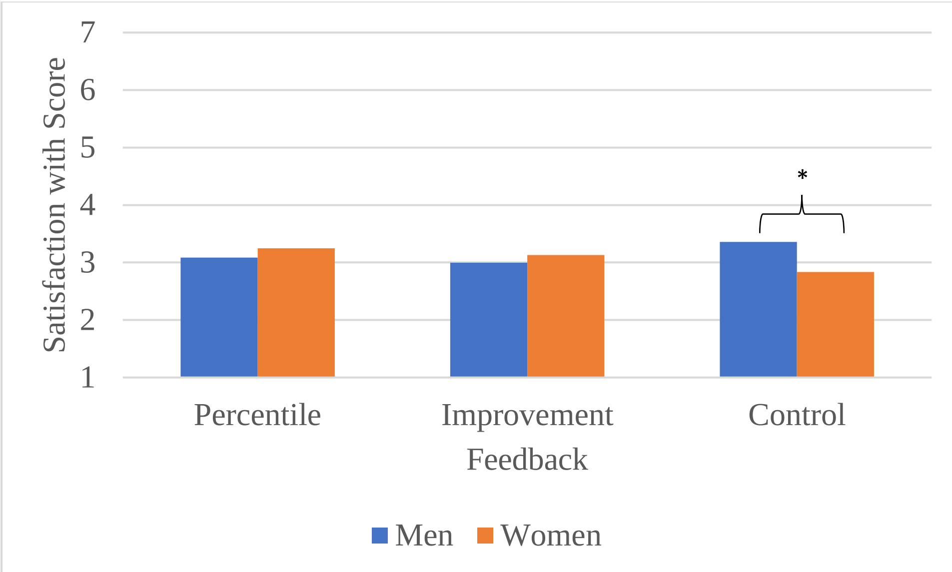

Figure 6. Predicted values for performance satisfaction by gender and feedback, controlling for growth mindset value of domain in Study 4. Predicted values are based on the second step of the hierarchical regression model, which includes the two-way interactions.

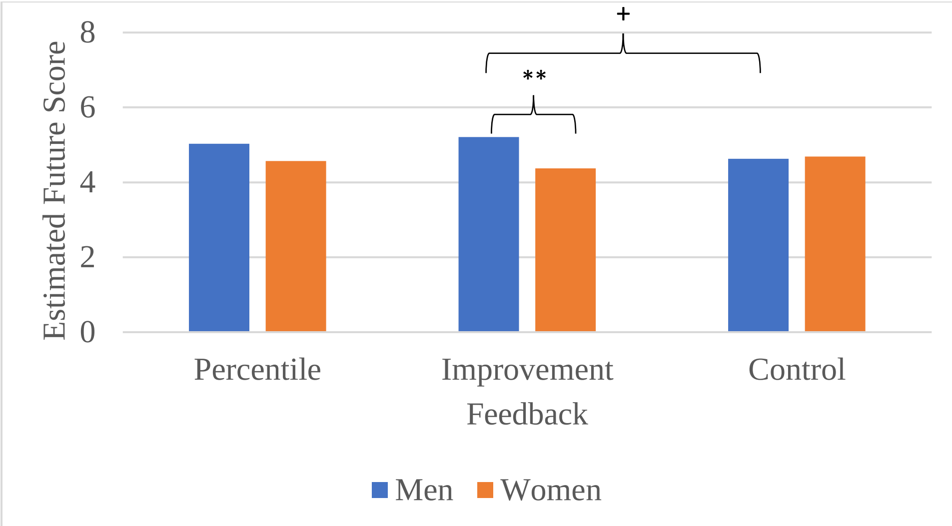

Figure 7. Predicted values for estimated future score by gender and feedback condition, controlling for growth mindset and value of domain in Study 4. Predicted values are based on the second step of the hierarchical regression model, which includes the two-way interactions.

Note: ***$p < .001$, **$p < .01$, *$p < .05$, +$p < .10$

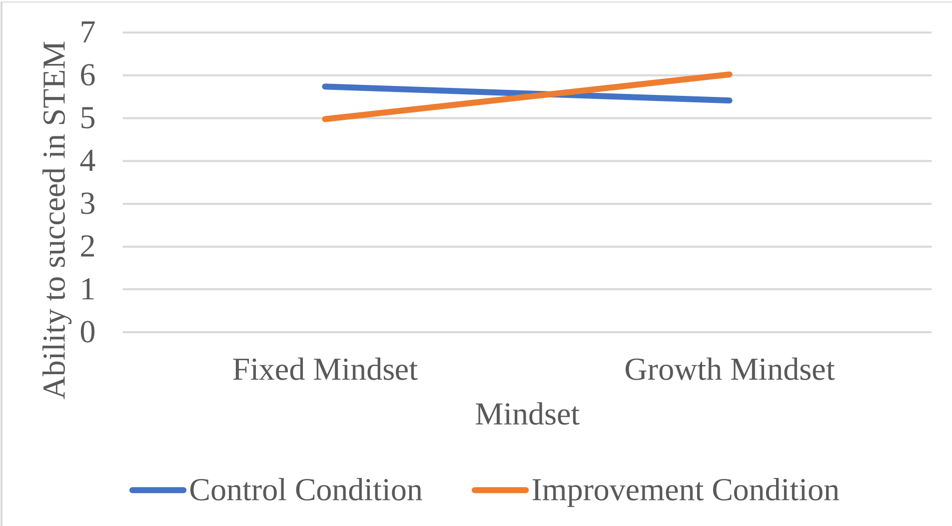

Figure 8. Predicted values for agreement with the statement “I have what it takes to succeed in STEM” by growth mindset and feedback condition, controlling for gender and value of domain in Study 4. Predicted values are based on the second step of the hierarchical regression model, which includes the two-way interactions.

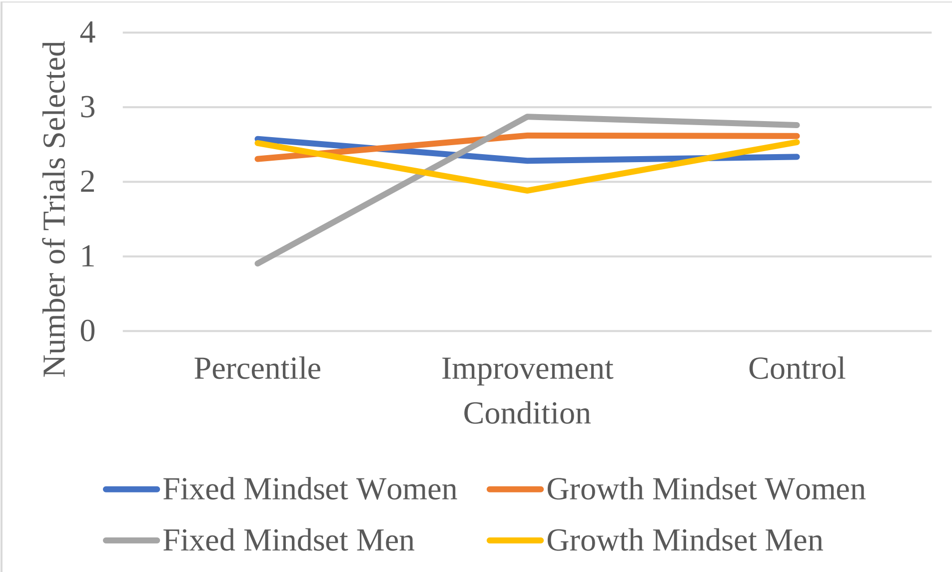

Figure 9. Predicted values for requested number of additional trials by gender, growth mindset, and feedback, controlling for value of domain in Study 4. Predicted values are based on the third step of the hierarchical regression model, which includes the three-way interactions. Fixed mindset is one standard deviation below the mean, and growth mindset is one standard deviation above the mean.