Female Success in STEM: How Self-Efficacy Drives Effort

A dissertation presented to

the faculty of

the College of Arts and Sciences of Ohio University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy

Leah R. Halper

August 2017

This dissertation titled

Female Success in STEM: How Self-Efficacy Drives Effort

by

LEAH R. HALPER

has been approved for

the Department of Psychology

and the College of Arts and Sciences by

Jeffrey B. Vancouver

Professor of Psychology

Robert Frank

Dean, College of Arts and Sciences

Abstract

HALPER, LEAH R., Ph.D., August 2017, Experimental Psychology

Female Success in STEM: How Self-Efficacy Drives Effort

Director of Dissertation: Jeffrey B. Vancouver

There is a shortage of STEM (science/technology/engineering/mathematics)-trained individuals in the United States, and entire groups of individuals (e.g., females) are underrepresented in the STEM fields. The majority of research focuses on why females do not pursue STEM as careers, with researchers agreeing that a key component is self-efficacy, which is one's belief in one's capability of achieving a goal. Moreover, research demonstrates that self-efficacy or self-confidence in one's abilities leads to an increased chance of pursuing a career in STEM. For this reason interventions have been developed to increase self-efficacy for STEM careers, and often these interventions are targeted towards females. However, there is disagreement about the effect of self-efficacy once one begins pursuing the goal of obtaining a career in STEM. Social cognitive theory proponents argue that self-efficacy will promote effort; whereas, control theory proponents argue that self-efficacy might undermine effort. STEM majors provide a unique population in which to study the relationships among the variables in a real world setting. In particular, this study investigates the relationship among gender, self-efficacy, effort, and performance in STEM-focused college courses.

To develop a set of hypotheses, the literatures on self-efficacy, which is a type of expectancy, were reviewed with a focus on the differences between social cognitive theory and control theory. Due to the theoretical and empirical evidence, control theory

was used to generate the hypotheses. In particular, the expected comparatively lower self-efficacy of females versus males for STEM-related classroom goals was hypothesized to encourage effort and enhance performance towards classroom goals such as exam grade goals. Additionally, alternative explanations and important contextual variables in regards to the hypotheses were considered due to the field study methodology in the current study.

Data were collected using a sample of engineering, mathematics, and physics majors in mathematics courses required for their major. Self-efficacy, effort, and performance were measured multiple times over the length of one semester. Performance measures were collected from instructors, and ability indices (i.e., standardized test scores) were collected from the university, both based on permission from participants.

As hypothesized, gender was found to relate to self-efficacy and effort such that females had lower self-efficacy for mathematics exams than males and also studied more hours for the exams (e.g., $M = 15.22$, $SD = 10.72$) than males studied (e.g., $M = 10.42$, $SD = 10.15$). Further, self-efficacy was found to fully mediate the relationship between gender and effort (95% CI: 0.081 to 6.270). Females also reported a lower goal on the exam than males. There was no difference between the genders on exam performance, possibly due to stereotype threat concerns undoing the advantages of extra studying. Indeed, there was no difference between males and females on ability as assessed via standardized test scores. Other possible reasons for the lack of a performance effect were explored in the discussion. For instance, differences in classroom networking or study habits between males and females (e.g., access to old tests), as well as issues of diminishing returns for additional studying were discussed. Also, the broader impacts of the results on future STEM interventions that include self-efficacy or confidence-building components were discussed. Overall, this study demonstrated that self-confidence (i.e., self-efficacy) does not always have a positive effect on motivation. Furthermore, females' lower self-confidence for science, technology, engineering, and math-related goals appeared to motivate studying behavior in mathematics courses.

Table of Contents

List of Tables

List of Figures

Page

Introduction

The number of domestic science, technology, engineering, and mathematics (STEM) degrees earned has been termed a “crisis”; there are currently not enough U.S. citizens with STEM expertise to meet U.S. labor and economic demands (Hira, 2010). On one hand, this lack of participation is surprising because STEM careers tend to pay well (Beede et al., 2011). On the other hand, STEM fields are seen as particularly challenging, and one has to do well in college to pursue such careers (Pascarella & Terenzini, 2005; Samson, Graue, Weinstein, & Walberg, 1984; Wai, Lubinski, & Benbow, 2009). Indeed, the challenging nature of STEM fields seems to be a major deterrent to women (Brainard & Carlin, 1997; Hill, Corbett, & St. Rose, 2010; Seymour & Hewitt, 1997), who in particular are underrepresented in the fields of engineering, mathematics, and physics (Baron-Cohen, 2002; Cvencek, Meltzoff, & Greenwald, 2011; Fredericks & Eccles, 2002; Herbert & Stipek, 2005).

In response to this STEM problem, research to date has focused on why females choose to pursue STEM fields less than males, and there is a general consensus in the psychology and education literatures on this issue (Brainard & Carlin, 1997; Seymour & Hewitt, 1997). A concept that is regarded as a key is self-efficacy (Campbell & Hackett, 1986; Pascarella & Terenzini, 2005), which is individuals’ belief in their capability to perform at a level necessary to achieve a goal (Bandura, 1999, 2001). Self-efficacy is thought to be an influential factor in the STEM gender gap primarily because of the influence of self-efficacy on choice of accepting or internalizing a goal (Bandura, 1997; Cheryan, Siy, Vichayapai, Drury, & Kim, 2011). Therefore, research is almost exclusively focused on how self-efficacy affects the goal choice process in STEM. More generally, self-efficacy theory (Bandura, 1997) predicts that individuals will likely choose goals or activities for which one's self-efficacy is high, provided the attractiveness of the outcomes associated with the option is positive (Bandura & Schunk, 1981; Colquitt, LePine, & Noe, 2000; Compeau, Higgins, & Huff, 1999; Locke, Frederick, Lee, & Bobko, 1984). For instance, self-efficacy of career-related abilities positively predicted what majors college students chose for themselves (Hackett & Betz, 1995), and individuals were more likely to consider possible careers when their self-efficacy for certain careers was high (Lent, Brown, & Larkin, 1986). In the case of STEM fields, females generally have lower self-efficacy than males for STEM careers (Campbell & Hackett, 1986; Correll, 2001; Ehrlinger & Dunning, 2003; Meece, Parsons, Kaczala, & Goff, 1982) and are thus less likely to attempt tasks that involve STEM topics, like college STEM classes, majors, or careers (Kay & Shipman, 2014). As a result of such research, scholars frequently advertise boosting self-efficacy to improve motivation and performance (Colquitt, LePine, & Noe, 2000). This focus on self-efficacy in regards to STEM has led to a concerted effort to include self-efficacy in STEM recruitment and retention interventions with a goal of encouraging participation from underrepresented groups (e.g., Stout, Dasgupta, Hunsinger, & McManus, 2011; Betz, 2004).

If self-efficacy components in STEM recruitment interventions are effective during the goal choice process, it is important to also consider how self-efficacy may affect motivation post-goal acceptance. However, there is some debate regarding how self-efficacy might affect individuals post-goal acceptance. According to many researchers, low self-efficacy can be detrimental during goal planning and goal striving (i.e., the processes by which resource planning and allocation for goal achievement occurs once a goal has been chosen), particularly in terms of effort toward goals (Bandura, 1977; Betz & Hackett, 1981; Schunk, 1991). Social cognitive theory (Bandura, 1986) and self-efficacy theory (Bandura, 1997) predict that individuals are less motivated to pursue goals that they have a lower expectancy of achieving. Therefore, individuals put less effort forth towards goals for which they have low self-efficacy (Bandura, 1989). However, other researchers suggest that low self-efficacy might instead be a motivator for individuals, increasing effort and possibly performance as well (Vancouver, More, & Yoder, 2008; Vancouver, 2012). These latter researchers suggest that beliefs in rates of goal progress per unit of resource applied, or the conceptualization of the self-efficacy belief, might negatively affect effort and performance. For example, an individual who believes goal progress will be difficult (i.e., the individual has lower self-efficacy for the goal) will feel the need to put more effort towards the goal than if the individual believed he or she could progress easily. Individuals will therefore *plan* to allocate more resources (i.e., in goal planning) and *will* allocate more resources (i.e., in goal striving) towards goals when self-efficacy is lower compared to higher (Vancouver, 2012). Thus, understanding the potential sign of self-efficacy's effect post-goal choice might be important when considering pursuing and striving for STEM careers. However, little to no research has examined this question or its implication regarding sex differences in self-efficacy post-goal acceptance (Kay & Shipman, 2014).

Indeed, among employed individuals committed to STEM domains, data show that females (e.g., in 2009, there were 1,199,000 college-educated STEM females in the United States; Beede et al., 2011) have lower domain-related self-efficacy than males (Post-Kammer & Smith, 1986). These differences also trace back to college (Besterfield-Sacre, Morena, Shuman, & Atman, 2001; Brainard & Carlin, 1998; Heilbronner, 2013; Hutchison, Follman, Sumpter, & Bodner, 2006), where there were roughly 103,450 female STEM majors in 2013 (Wang, 2013). That is, in a study measuring attitudes at 17 universities in the United States, females had lower self-efficacy than males regarding engineering ability one year after college-entry (Besterfield-Sacre et al., 2001). Yet, these individuals have apparently chosen to pursue a career in STEM. This gender difference in STEM-related self-efficacy during goal striving (i.e., post-goal acceptance) offers an opportunity to test the effects of self-efficacy on motivation during the goal-striving process. In addition, the gender difference in self-efficacy allows for a unique opportunity to study goal striving in a field setting, which is an area of research typically investigated in laboratory settings. Therefore, my research question involves examining how self-efficacy might impact the effort males and females apply towards STEM-related goals, such as obtaining a high grade on exams in a STEM-related class, which affects one's ability to enter a STEM career.

In this dissertation, I make the argument that low self-efficacy facilitates females' success in STEM courses. That is, I argue that although low self-efficacy decreases the probability that females will choose a STEM field to pursue, that same low self-efficacy may motivate a stronger work ethic for STEM-related classroom goals in those females who choose to pursue a STEM field subject. The basis of my argument is that during the goal-striving process, self-efficacy leads to the application of more effort in preparing for goals because individuals put more effort towards goals that seem harder to reach (Vancouver et al., 2008; Vancouver, 2012). If females have lower self-efficacy than males in relation to STEM courses during goal striving, females may put more effort into STEM-related goals than males. Moreover, this increased effort may lead females to outperform males in STEM-related classroom goals, on average. This difference in effort and performance may ultimately be attributed to the effect of lower self-efficacy.

The purpose of this paper is to demonstrate the negative effect of self-efficacy on motivation in goal striving by assessing the mechanisms that occur during the self-efficacy, effort, and performance processes for females and males. To do this I first briefly review the literature concerning motivation and self-efficacy. Following this review, I describe two broad views of self-efficacy: social cognitive theory, which contains self-efficacy theory (Bandura, 1989) and control theory, which describes a non-monotonic, discontinuous model of self-efficacy on effort (Carver & Scheier, 1998; Vancouver et al., 2008). In the process, I highlight the theoretical and empirical success of the non-monotonic, discontinuous model of self-efficacy, which serves as a basis for my hypotheses. Following the main hypotheses, I describe other relevant factors that may differentially affect the effort and performance of males and females in this particular context (i.e., STEM classrooms), which includes consideration of alternative explanations to the proposed effects as well as processes that may obscure the effects.

Theories of Motivation and Self-Efficacy

Motivation is the force that directs and sustains behavior over time (Diefendorff & Chandler, 2011). More specifically, motivation is defined as factors that impact the direction of behavior and effort (Pinder, 2008). The direction of behavior and effort is always towards some goal, or a desired state that an individual wants to obtain (Austin & Vancouver, 1996). Goals are central to motivational theories. Without goals, individuals would have few reasons to self-regulate, or direct, their behaviors. Goals are defined broadly in the literature and can range from biological (e.g., hunger or temperature thresholds that individuals want to reach) to academic (e.g., the desire to reach a particular skill level) (Austin & Vancouver, 1996). One related motivational factor is self-efficacy or the belief in one's capability to reach a particular goal. Self-efficacy is thought to impact both the direction of behavior (i.e., goal choice) and effort towards goals (i.e., goal striving) (Bandura, 1997; Vancouver et al., 2008).

Although researchers agree that there are many situations in which self-efficacy might impact goal choice and goal striving, there is some argument in regards to how self-efficacy impacts such behavior. Bandura (1977, 1997) posits that higher levels of self-efficacy increase (a) the chance that an individual will accept a goal (i.e., goal choice) and (b) the amount of effort allocated to the goal at hand (i.e., goal striving). That is, individuals who have high self-efficacy for a task have higher motivation for that task than individuals who have low self-efficacy. These roles for self-efficacy arise from the more comprehensive social cognitive theory (Bandura, 1986) described next.

Social cognitive theory. Social cognitive theory aims to predict and explain behavior based on the reciprocal relationships among person variables, environmental variables, and behaviors (Bandura, 1986; 1989). In social cognitive theory, the environment and behaviors shape person variables like beliefs, thoughts, and affect; both environment and person variables affect behavior; and person variables via behavior may affect the environment. This triage is called triadic reciprocal determinism, because all three variables (i.e., person variables, environmental variables, and behavior) influence and are influenced by the others. Therefore, individuals are products of their behaviors and environment as well as producers of their behavior and environment. The relationships among these variables contribute to self-regulated behavior, which is purposeful, self-monitored action (Bandura, 1991). Expressly, social cognitive theory credits self-efficacy (i.e., a person variable) as a particularly important variable in self-regulation. Self-efficacy is the belief in one's own agency, which is one's ability to produce change in the environment (Bandura, 1991). Due to self-efficacy's importance, self-efficacy theory is nested in social cognitive theory. Self-efficacy theory predicts the effects of capability beliefs on thoughts, motivation, and behavior.

Specifically, self-efficacy theory predicts a positive linear relationship between self-efficacy and motivation (Bandura & Locke, 2003). According to the theory, individuals are more motivated to work towards goals for which they have a higher self-efficacy. Therefore, individuals are more likely to accept goals that they believe they are likely to achieve (Bandura, 1991). When individuals accept a goal, they are more likely to allocate resources (i.e., put effort) towards the goal the higher their self-efficacy for that goal (Bandura, 1986, 1997). Specifically, Bandura credits discrepancy creation as a driver of effort for chosen goals. Discrepancy creation is a process where one sets a difficult goal for oneself, which then creates a discrepancy between one's current and desired states. Self-efficacy theory predicts that those with higher self-efficacy create larger discrepancies via setting more difficult goals. If those with high self-efficacy set more difficult goals than those with low self-efficacy, those individuals with higher self-efficacy create a bigger discrepancy between their current and desired states. Individuals put more effort towards these high-discrepancy goals for which they have high self-efficacy (Bandura & Locke, 2003).

Several meta-analyses corroborate such theorizing. Self-efficacy was positively related to behavior choice (r = 0.34) in one meta-analysis citing eight studies (Sadri & Robertson, 1993) and positively related to performance in more than one meta-analysis (r = 0.40, 16 studies, Sadri & Robertson, 1993; r = 0.38, 45 studies, Moritz, Feltz, Fahrbach, & Mack, 2000; r = 0.41, 67 studies, Gully, Incalcaterra, Joshi, & Beaubien, 2002). Most recently, Sitzmann and Yeo (2013) meta-analyzed 38 studies of self-efficacy and performance and found a positive effect for 93% of studies at the between-person level of analysis.

Yet, the research on the self-efficacy-to-performance relationship is not always so consistent. Using the same set of 38 studies noted above, Sitzmann and Yeo (2013) found no effect of self-efficacy on *subsequent* performance at the *within-person* level of analysis. They also found that performance's effect on subsequent self-efficacy was highly positive at the within-person level, suggesting that the between-person positive correlation is a function of performance's effect on self-efficacy, but not the other way around. Researchers generally agree that performance impacts subsequent self-efficacy positively. If an individual performs well on a goal, that individual will believe that he or she has the capacity to perform well on that goal in the future (Bandura, 1997; Vancouver & Kendall, 2006). Moreover, a control theory view of self-regulation (e.g., Carver & Scheier, 1998; Vancouver, 2008) expects no or slightly negative direct effect for self-efficacy on performance during goal striving, as well as a more pronounced negative effect on effort. The control theory view is reviewed next.

Control theory. In control theories of self-regulation, discrepancies are a main factor in determining action (Diefendorff & Chandler, 2011). Specifically, if an individual has not reached a desired goal, there is a discrepancy between where that individual wants to be and where the individual currently is, and the individual is likely motivated to act to reduce the discrepancy (Vancouver, Thompson, Tischner, & Putka, 2002). Indeed, this view of discrepancy as a driver of behavior is also found in social cognitive theory (Bandura, 1986). Also similar to social cognitive theory, control theory assumes individuals have many goals and that these are arranged hierarchically within the individual. However, somewhat departing from social cognitive theory, control theorists contend that other processes and complex behaviors emerge from these linked goals. For example, goals might be accepted, rejected, or abandoned via goal systems that compare two or more goals (Vancouver, Weinhardt, & Schmidt, 2010). Collectively, this process is referred to as a goal-choice process. Once chosen, goal-planning and goal-striving processes may ensue. In all of these processes, beliefs like self-efficacy are presumed to play a role. However, because the processes are not identical, the roles are more complex than implied by social cognitive theory predictions. In particular, control theory describes a non-monotonic, discontinuous empirical model of self-efficacy's effect on motivation (e.g., Vancouver, 2008).

The non-monotonic, discontinuous model corresponds with self-efficacy theory's indirect effect of self-efficacy on goal choice (i.e., there is a positive effect of self-efficacy on goal choice), but does not correspond with self-efficacy theory's direct self-efficacy effect on effort during goal striving. In terms of goal choice, control theory predicts that individuals with higher self-efficacy are more likely to adopt goals relevant to the task for which individuals have high self-efficacy (Vancouver, 2008). For example, an individual with high self-efficacy for learning STEM topics and low self-efficacy for learning humanities is more likely to declare a STEM major than a humanities major. This goal adoption creates a positive discontinuity in motivation toward the goal (i.e., no motivation if chosen not to pursue, some motivation if chosen to pursue) as a function of self-efficacy. However, when considering effort allocated while planning or striving under conditions of feedback ambiguity, control theory predicts self-efficacy may negatively influence effort allocated. This change from a positive discontinuity to a negative slope for self-efficacy creates the non-monotonic, discontinuous model of self-efficacy (Vancouver, 2008; 2012). See Figure 1.

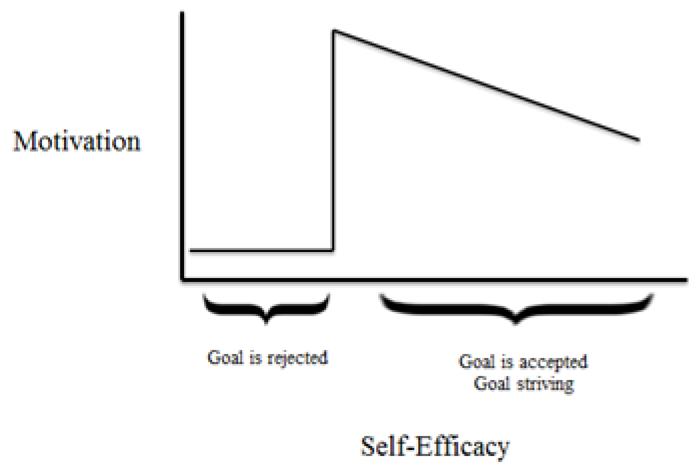

Figure 1. The non-monotonic, discontinuous model of self-efficacy. Adapted from “Self-efficacy and resource allocation: Support for a nonmonotonic, discontinuous model,” by J.B. Vancouver, K.M. More, and R.J. Yoder, *Journal of Applied Psychology, 93,* p. 36. Copyright 2008 by American Psychological Association.

In the non-monotonic, discontinuous model of self-efficacy, self-efficacy represents a weight for the individuals’ belief of their rate of goal progress per unit of effort. This notion is consistent with Bandura’s (1997) definition of self-efficacy as the belief in one’s capacity. Specifically, the higher one’s capacity, the faster progress towards the goal per unit of resource (e.g., time or effort) one should expect (Vancouver & Purl, 2017). That is, a highly self-efficacious individual believes he or she will make progress on the goal quickly per unit of effort. An individual with low self-efficacy believes he or she will make relatively slower progress on the goal per unit of effort. Those with high self-efficacy, who believe they will make progress towards their goal quickly per unit of effort, also perceive there are fewer resources needed to achieve some goal than those with lower self-efficacy. Conversely, those with lower self-efficacy believe more resources are needed to achieve some goal because they will make progress towards their goal more slowly per unit of effort (Vancouver, 2012).

Put differently, motivation (e.g., effort allocated) for a rejected goal is extremely low (i.e., zero). Motivation for an accepted goal is higher, but varies depending on the exact level of self-efficacy, or the belief in progress per unit of effort one has (Vancouver & Purl, 2017). An individual whose self-efficacy falls toward the high end believes he or she will make more progress on the goal per unit of effort than an individual whose self-efficacy falls toward the middle of the scale (Vancouver, 2012). Thus, if the goal is accepted, those with higher self-efficacy allocate less effort than those with lower self-efficacy. Of course, allocated effort is a matter of degree (i.e., not dichotomous), and more effort is allocated for progressively lower levels of self-efficacy given one does not fall off the "cliff" of the discontinuity in the above figure. Likewise, as one's self-efficacy increases, less and less effort is allocated.

There is empirical evidence for the theoretical underpinnings of control theory. In a study demonstrating the validity of the non-monotonic, discontinuous model of self-efficacy individuals were instructed to click on squares jumping around a computer screen (Vancouver et al., 2008). Self-efficacy was manipulated by the size of the squares (i.e., individuals had lower self-efficacy regarding hitting the smaller squares) and by enacted mastery (Bandura, 1986). High enacted mastery increases when individuals practice on easiest problems first and as compared to when individuals begin practicing on the most difficult problems. Those in the high enacted mastery condition had higher self-efficacy for the task, on average, because they started off with easier squares to hit. Those with low enacted mastery had lower self-efficacy for the task. Individuals were given the option to pass on any particular square given its size. If individuals chose not to pass, they were asked to record how many seconds they would like to allocate to attempting to hit that square by clicking on the square as it “jumped” around the screen. Vancouver et al. found that individuals were more likely to pass on the most difficult squares (i.e., allocate no resources). Thus, when self-efficacy was very low, individuals would not accept the goal. This trend produced a positive effect of self-efficacy on resources allocated. However, if individuals chose to accept the goal, individuals allocated on average 4 seconds more to the most difficult target in comparison to the easiest target. Likewise, those in the high enacted mastery condition allocated fewer resources (i.e., chose to try to win the game in fewer seconds) than those in the low enacted mastery condition, on average. These results demonstrate that higher self-efficacy leads to allocating fewer resources than lower self-efficacy in goal striving. This study used both a within-person protocol and random assignment to the between-person manipulation of self-efficacy, providing evidence that the non-monotonic, discontinuous model can be demonstrated at both the within and between-person levels of analysis. Indeed, in a study that manipulated self-efficacy between-person using manipulated feedback in an anagram-like task, there was a significant reduction in time spent searching for anagrams for those in the high self-efficacy condition (i.e., 42.44 seconds) compared with those in the low self-efficacy condition (i.e., 92.90 seconds; Vancouver, Gullekson, Morse, & Warren, 2014).

In another study focused more directly on the topic of this dissertation, Vancouver and Kendall (2006) employed a within-subject methodology to examine the relationships between self-efficacy and motivation for studying for exams among college students. The researchers measured self-efficacy in terms of anticipated grade if the individual were to take the exam at that moment. They found that for every anticipated grade increase (i.e., higher performance self-efficacy), planned study time dropped by 15 minutes, on average. Likewise, participants' actual study time reported just prior to the exam was lower when self-efficacy was higher. Moreover, some studied less than planned, and when asked why, they reported feeling that they knew the material better than they thought; whereas, those who studied more than planned noted having more trouble than expected with the material.

There were some criticisms of Vancouver and Kendall's construct validity and thus their conclusions. In 2012, Albert Bandura criticized the content validity of this self-efficacy measure. Self-efficacy was measured via one item, which asked participants what grade they expected to get on the upcoming exam. Bandura argued that the answer participants gave may have been a result of factors other than self-efficacy (e.g., participants may have taken into account not just their capacity, but the amount of time they could reasonably study over the next two days prior to the exam given other obligations). In his commentary, Bandura recommended the use of a self-efficacy measure that assesses individuals' belief in their capacity for a range of grade levels (i.e., A through F) (Bandura, 2012). Further, Bandura criticized the study for the use of a retrospective measure of effort, or time spent studying, which was not validated.

Given criticisms of past field studies, there were two primary goals of the current study. The first was to address the lack of gender-related research in regards to goal-striving in STEM. Overall, the field of psychology has paid much attention to how females' and males' self-efficacy may affect their goal choices. Lower self-efficacy for STEM prevents females from entering STEM fields; males' higher self-efficacy for STEM facilitates their pursuance of STEM fields (Campbell & Hackett, 1986; Correll, 2001; Ehrlinger & Dunning, 2003; Kay & Shipman, 2014). Little to no research has examined gender differences in self-efficacy during the goal-striving process. As a start, researchers have demonstrated that females have lower self-efficacy than males after declaring the major (Brainard & Carlin, 1998; Heilbronner, 2013), and that there may be differences in females' and males' perceptions of their study habits during goal striving (Besterfield-Sacre et al., 2001). To be specific, in one longitudinal study, researchers found that female engineering majors felt more comfortable with their study habits than male engineering majors (Besterfield-Sacre et al., 2001). Authors conjectured that female engineers believed that their studying habits for topics such as engineering and physics were more effective than male engineers (i.e., females had a higher self-efficacy for their ability to study). However, another way to interpret the result of females feeling more comfortable with their study habits is that these females in STEM believed their knowledge, skills, and abilities were poorer than the knowledge, skills, and abilities of males in STEM, but this difference in knowledge, skills, and abilities could have been compensated for by studying harder. Because few studies have examined how goal striving may differ between the genders, and the few that have studied such a question have unclear results, the main goal of the current study was to examine such potential differences in motivation.

The second goal was to examine the goal-striving process with past measurement criticism of self-efficacy and effort in mind. To get a true sense of capacity beliefs, self-efficacy should be measured over a range of grade levels (Bandura, 2012). The current study measured self-efficacy in that way. Additionally, the current study aimed to develop a valid measure of effort expended (i.e., time studied for examinations). There is a theoretical reason to use a valid retrospective effort in assessing past studying. Retrospective measures can capture all studying completed for exams (i.e., even the cramming completed in the hours or minutes leading up to an exam) whereas antecedent effort measures cannot. However, it is important to validate retrospective effort measured used. Thus, multiple measures of effort were part of the current study to assess the reliability and validity of retrospective effort. Consequently, self-efficacy and effort were measured in a way would could address the concerns in past commentaries on self-efficacy and goal striving (Bandura, 2012). In my hypotheses below, I translate the non-monotonic, discontinuous model in terms of the particular context of this dissertation (i.e., male and females in STEM) and provide predictions for both effort and performance outcome variables.

Self-Efficacy-Effort/Performance Relationship for Females and Males in STEM

According to evidence from past studies related to self-efficacy and gender, females should have lower STEM-related self-efficacy than males even after declaring a STEM major. Often, this self-efficacy is measured in terms of career-related goals (e.g., belief in ability to be hired as an engineer, etc.), because past literature has mostly studied females during the goal choice process. However, if females believe they can or cannot achieve superordinate STEM goals such as become engineers, it is also likely the case that they have a belief that their capabilities to achieve important subgoals that contribute to superordinate goals during the goal striving process are lesser as well. Examples of such subgoals include performing well on STEM course examinations or getting a STEM internship. Thus, the self-efficacy measured in this dissertation is self-efficacy for achieving grade goals on examinations to capture the goal striving context in which this dissertation is set. Additionally, the purpose of Hypothesis 1 is to confirm that gender differences in self-efficacy can be extrapolated to subgoals such as STEM exam performance. If it is the case that females have lower self-efficacy for STEM course-related goals (i.e., exam goals) than males, there should be differences in effort exerted towards these goals as well. According to empirical evidence from the non-monotonic, discontinuous model, if females have lower STEM-related self-efficacy than males, females should apply more effort to STEM-related goals than males, such as grade goals on exams in STEM courses. In the context of STEM training or STEM classes, effort can be indexed by studying. For example, female students, on average, should study more for examinations than male students because females have lower self-efficacy for STEM-related topics than males, on average.

Hypothesis 1: Females in STEM will have lower self-efficacy for STEM-related goals than males in STEM.

Hypothesis 2: Lower self-efficacy will be related to more effort applied in goal striving.

Hypothesis 3: Females in STEM will apply more effort in goal striving for STEM-related goals.

Hypothesis 4: Self-efficacy will mediate the gender-to-effort relationship.

The amount of effort applied to a goal is often important in the way that the effort relates directly to the performance of the individual, all else being equal. Therefore, this dissertation also investigated the relationships among gender, effort, and performance. The performance relevant in the current study is exam performance in a STEM-relevant college course. In the literature on self-efficacy theory, more effort is linked to better performance (Bandura, 1994; 1997), but only in limited circumstances. For example, individuals with higher abilities may not need to apply as much effort as those with lower abilities to achieve the same outcome. Thus, it is only when comparing two individuals with equal ability that one might expect to see a relationship between effort and performance. If there are two individuals with equal ability, the individual who spends more time and resources in preparing for a task (e.g., exam) will likely perform better on that task (Beck & Schmidt, 2012; Fisher & Ford, 1998; Johnson, Joyce, & Sen, 2002). Similarly, the efficiency of that time and resources applied to the task would need to be equal across participants. If not, different individuals would make differential progress, perhaps somewhat detectable by the individual, in preparing for the task. For these reasons, effort and performance may not be related. However, if the average ability of females and males for STEM exams is equal (see the next section for further elaboration on this topic) and strategies are evenly distributed across genders, then the extra effort females apply as a function of self-efficacy should lead to gains in performance. This logic leads to my fifth and sixth hypotheses.

Hypothesis 5: Females in STEM will demonstrate higher performance than males for STEM-related classroom goals.

Hypothesis 6: Effort will mediate the gender-to-performance relationship.

Self-efficacy is measured, not manipulated, in the current study. Therefore, it is important to consider other factors that may impact the outcomes of interest (i.e., effort and performance) in this specific context. Such factors include (a) expectancy beliefs other than self-efficacy that females have that might influence effort expenditure, (b) other constructs that factor into motivation besides expectancy, such as goals and valence, (c) knowledge/skills/abilities differences between males and females in STEM, and (d) the possible effects of gender-related stereotypes about male and female knowledge/skills/abilities in STEM on performance. These factors are reviewed in the following sections.

Confounding Factors in the Gender-Self-Efficacy-Effort Relationships

Factors that may explain (i.e., be the true cause) or obscure relationships among gender, self-efficacy, and effort include differences in values, outcome expectancies, and goals. The possible roles of these constructs are detailed in the next sections.

Value. One non-expectancy-related variable that may impact motivation is value (Vroom, 1964). More specifically, individuals are more motivated to perform actions that are linked to valued outcomes. Outcomes can be valued if they are intrinsically interesting or if they contribute to important goals for individuals (Eccles & Wigfield, 2002). Therefore, a task can have value if completing the task is a step towards a larger goal (e.g., a career goal). In a well-regarded theory of motivation, lower value placed in one's goal predicts less effort put toward obtaining the goal (Vroom, 1964). If females find mathematics or other STEM-related topics to be less valuable than males, this value difference might be predictive of effort applied. In one longitudinal study of high schoolers, females reported lower math intrinsic value (e.g., reported feeling less interested in learning mathematics) and lower math utility value (e.g., reported the belief that mathematics was not helpful for future goals). Math intrinsic value and math utility value were positively related to STEM course selection during the ten years the authors studied the participants (Guo, Parker, Marsh, & Morin, 2015).

To extrapolate to the context of the current study, it is possible that the lower value females place in STEM, and specifically mathematics, the less effort females may put towards their STEM courses. It is important to note that the outcomes for Guo et al.,'s (2015) longitudinal study were only goal choice-related outcomes. If value partially guides goal choice, then it is likely that females who decide to enter STEM majors place a high level of value in STEM. This value for STEM will then be high during goal-striving. In other words, once in the major, males and females may have equivalent levels of value for STEM topics and thus, there is little variance between the genders in value. In the current study, two types of value (i.e., intrinsic enjoyment or interest in mathematics and valuing mathematics because mathematics impacts important career goals) were measured to assess gender differences in value and the relationship between value and effort. Further, belongingness, or how similar individuals believe other individuals in their class or major are to them, was measured in the current study given some evidence that belongingness is positively related to the value individuals place in their major (Cheryan, Plaut, Davies, & Steele, 2009). If a null or reverse gender-to-motivation effect is found, the result may be due to variance in value. Thus, value can only be tested as an alternative explanation if there is either no difference between males and females in effort applied in the course or females applied significantly less effort in the course. Additionally, value can only be tested as an alternative explanation if there is a gender difference in value.

Outcome expectancies. Being female in a male-dominated profession affects more than self-efficacy beliefs. Females are often treated differently than males in male-dominated settings (Bell, McLaughlin, & Sequeira, 2002). For instance, females are often rated as not having the agentic characteristics to succeed in such positions (Heilman, 2012). Further, if they demonstrate such characteristics to succeed, females are rated as less likeable due to stereotypes regarding females and agentic traits (i.e., females should not act in agentic manners; Heilman, Wallen, Fuchs, & Tamkins, 2004). As a result, females may have a more difficult time maintaining leadership positions and are less likely to be afforded a promotion (Magee & Galinsky, 2008). This effect (i.e., females have a more difficult time advancing in organizations) has been called the glass ceiling (Bell, McLaughlin, & Sequiera, 2002). Importantly, females believe that there is discrimination against females in many male-dominated fields, especially in terms of climbing the leadership ladder (Bell et al., 2002). This belief might lead to two effects that undermine the tests of the above hypotheses. First, if females believe that they will have a harder time obtaining opportunities, such as promotions and higher pay in organizations, due to discrimination they may also feel the need to put more effort into their work to compensate for these discriminatory practices. This process is similar to the self-efficacy belief hypothesis, but in this case the compensatory effort is to overcome evaluation or other biases, not to overcome capacity insufficiencies. Second, because of the belief in the inequalities facing them in the real world, females may reduce effort because of the reduced expected utility of the effort (Vroom, 1964).

Of the two, the former is more likely. Indeed, research demonstrates that females believe that there is gender bias in the workplace and that they have to work harder to achieve similar outcomes to males due to such bias (Cooper Jackson, 2001; Ragins, Townsend, & Mattis, 1998; Wrigley, 2002). For example, in one survey-based study females reported beliefs that (a) organizations value women less than men, (b) women are offered fewer job rotation assignments, (c) women whose leadership styles are more in line with male leadership styles are viewed negatively, (d) women are offered fewer development opportunities, and (e) there are few policies that aid in reducing barriers to women's advancement (Cooper Jackson, 2001). Further, successful females in male-dominated occupations, whether STEM-related or not, (e.g., CEOs, managers, and professional athletes) report the need to work harder than, and outperform, males to achieve similar outcomes (Kay & Shipman, 2014). Thus, due to gender discrimination females might believe that to advance post-college (e.g., get a job, achieve a good starting salary) or to advance in an organization they need to work harder and perform better than males.

One alternative explanation to the gender-to-effort relationship hypothesis is that females allocate more resources than males to STEM-related goals because females believe that to receive the same opportunities as males they need to work harder and outperform males on STEM-related goals. Therefore, it is possible that greater motivation would not stem from lower self-efficacy in one's current knowledge, skills, and ability to achieve STEM-related goals, but rather in a belief in discriminatory practices for which females need to compensate by performing above and beyond males. Likewise, females might believe that they need to apply more effort to STEM-related goals than males in contexts such as school to battle such gender bias. That is, females in STEM majors might believe that professors and employees will only consider them for post-college opportunities if they outperform the males in the classroom. If so, this belief may be an alternative explanation to why females are more motivated than males in regards to STEM-related goals. That is, instead of the effect of gender on effort stemming from a lower capability belief the effect of gender on effort may stem from a belief in a need to outperform males to attain similar opportunities due to gender discrimination. This possible alternative explanation should tested as a control in the hypotheses if there is support for the effort hypotheses, particularly if there is a gender difference in such beliefs, to assess the legitimacy of the alternative.

However, it is important to note that in the current context, I am investigating exam-related effort, and exams are generally considered an objective indicator of ability.

When objective scores are used in hiring/promotion decisions there will be significantly less bias or discrimination than when subjective measures are used (Arvey & Renz, 1992; Schultz & Schultz, 2010). To the extent that individuals believe that exams scores are objective and thus free from bias, belief in how to respond to gender discrimination may not be of import. I assessed this prediction by measuring beliefs regarding gender discrimination in general and in the classroom. This alternative operationalization of the gender-to-effort alternative explanation should also be tested as a control in the effort-related hypotheses if there is support for the effort hypotheses and if there is a gender difference in such beliefs.

Goals. When examining differences in effort, it should be acknowledged that there are other variables that may impact motivation besides expectancy variables such as self-efficacy and value beliefs. One of those variable is goals. Self-efficacy is positively related to goal level (Locke & Latham, 1990), meaning that those who believe they have a greater capacity to achieve a goal often set more difficult goals for themselves (Zimmerman, Bandura, & Martinez-Pons, 1992). Goals are also positively related to effort (Locke & Latham, 1990). This effect can be explained by control theory. If individuals set easier goals for themselves, these individuals are closer to achieving their goals than those who set more difficult goals for themselves. Therefore, individuals who set easier goals do not need to put in as much effort to reach those goals and do not put in as much effort as individuals with more difficult goals (Vancouver et al., 2002). If females have lower self-efficacy for STEM-related goals, the goal levels they set for exams (e.g., a goal of getting an A versus a B- versus a C+) may be lower than that of males. If goals positively relate to effort, then females may put less effort into studying for exams because they have lower goals for such exams and thus do not feel the need to put as much effort in to achieve the goal compared to males and their goal.

Two qualifying issues regarding goals and exam-related behavior need to be mentioned. First, the question of goals is less of a concern if there is no variance in reported goals. Indeed, a past study looking at a context similar to this dissertation found very little variance in self-reported goal levels for exam (i.e., most individuals reported that their goal on exams was either an A or an A-) and thus, goals did not have a substantial impact on the relationship between self-efficacy and effort (Vancouver & Kendall, 2006). Second, despite the strong evidence of an association between goal level and performance, Vancouver et al. (2001) found that self-reported goal level was negatively related to performance over time. They conjectured that the goal level was somewhat a prediction of performance, which was largely determined by self-efficacy. That is, in this type of predictive context goal level may not play a very large motivational role. Therefore, it is possible that goal level will not be important in the current study. To confirm this, goals were measured to assess gender differences and the effect of goal level on subsequent motivation.

Confounding Factors in the Gender-Performance Relationship

In studying the relationships among self-efficacy, effort, and performance researchers have learned that the effort-to-performance relationship is not always straightforward because not only do motivational factors impact performance, but so do knowledge, skills, and abilities (KSAs; Cortina & Luchman, 2013). Knowledge is thought of as relevant information an individual holds in regards to performance tasks (Campbell, 1990). Skills pertain to processes relating to performance behavior. An individual with sufficient skills understands and can execute the behavioral processes necessary for high performance (Campbell, 1990). Abilities, while are less malleable than knowledge and skills, include concepts like general intelligence, perceptual speed, and psychomotor capabilities (Kanfer & Ackerman, 1989). KSA's, like motivation, are important determinants of performance (Cortina & Luchman, 2013). They are also related to motivation in complicated ways (Vancouver et al., 2008). For example, an individual might put extra effort (i.e., motivation) towards a goal knowing that extra effort is necessary because of lack of skill, but the individual might not put forth *enough* effort to make up for the deficiency. Here I consider potential gender differences in knowledge, skills, and abilities in STEM (i.e., both actual and perceived differences). This is particularly important because self-efficacy is essentially one's belief in one's KSAs thought relevant to the task. Thus, I consider gender differences in STEM knowledge, skills, and abilities, as well as differences in perceptions of knowledge, skills, and abilities.

Ability, knowledge, and skills. As noted, in addition to the link between effort expenditure and performance, high amounts of ability, knowledge, and skills are linked to better performance (Guion, 2011; Leach, Wall, Rogelberg, & Jackson, 2005; Schmidt, Hunter, & Outerbridge, 1986). There are possible gender differences in knowledge, skills, and abilities (KSAs) in STEM. This possibility likely arises not because of some inherent differences between males and females, but as a result of the process of choosing to enter a STEM field. In particular, if self-efficacy arises from biased perceptions of ability differences, females in STEM may have higher KSA levels than males because their relatively lower self-efficacy keeps all but the most capable from pursuing the field. For example, if individuals believe that they have to reach some threshold to strive for a STEM-related goal (e.g., "I believe I have to be able to get B's in calculus to declare an engineering major"), research would indicate that females are less likely to believe they will reach such a threshold than males, but not because they have lower abilities. Therefore, it is possible that females in STEM have higher KSAs on average than males in STEM, because low self-efficacy keeps more females from attempting STEM-related goals than males. Thus, the females that choose STEM-related goals may be higher, on average, in KSAs than males (see Figure 2). As Figure 2 illustrates, more males with lower KSA's might enter a STEM field than females do. As a result, the average KSA level for males in STEM may be lower than the average KSA level for females in STEM. If it is the case that females in STEM have higher STEM KSAs on average than males in STEM, it is possible that these high KSAs, and not necessarily more effort expenditure due to low self-efficacy post-goal acceptance, will lead females to perform better than males.

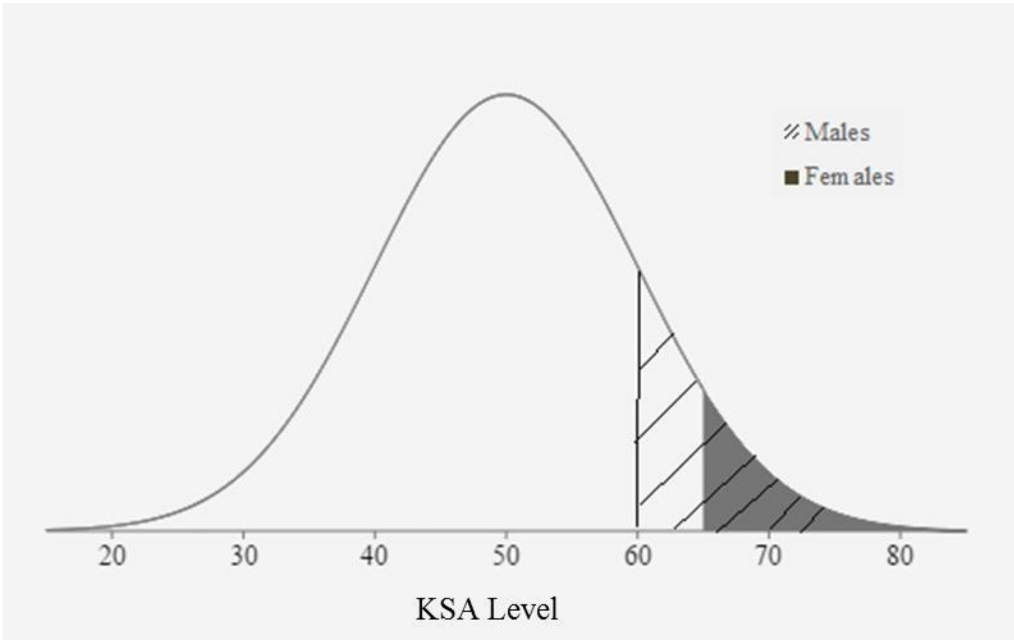

Figure 2. Illustration of KSA difference of males and females due to self-efficacy-determined likelihood of entering a STEM major.

The empirical evidence for differences in the KSA's levels of males and females in STEM fields is thin. Of the few studies that have examined female STEM KSAs and male STEM KSAs in college, little to no significant differences between the genders were found. For instance, in two studies researchers found no gender differences in grades or in grade point average (GPA) between male and female engineering majors or other STEM majors (Crocker, Karpinkski, Quinn, & Chase, 2003; MacPhee, Farro, & Cannetto, 2013). Likewise, Hackett, Betz, Casas, and Rocha-Singh (1992) found no difference in cumulative college GPA in male and female engineering/sciences majors. Another study found that there were no gender differences in STEM course grades (Miller & Halpern, 2013). One 16-year longitudinal, multi-institutional study found there to be significant GPA differences in biology, the physical sciences, and engineering in favor of females (i.e., the difference was approximately .089 GPA points). Authors found the difference to be more pronounced in smaller colleges (Sonnert & Fox, 2012). However, it should be noted that all five of these studies used performance as their measure of KSAs. Yet, differences in performance, when found, may have been related to motivational influences rather than KSA influences. Nonetheless, the current study measured performance with the purpose of controlling for initial (i.e., high school) knowledge, skills, and ability, and also current knowledge, skills, and ability. If there is a gender difference in ability, ability should be a control in all performance-related hypotheses, to serve as an alternative explanation for performance differences if performance differences are detected.

In considering the reason most research has not identified STEM KSA differences in college between males and females, it is important to consider other variables that might adversely affect the performance of females as compared to males in STEM. Indeed, perceptions of stereotypes have been shown to be related to lower performance in otherwise high-performing subgroups (e.g., females in mathematics). Therefore, I consider gender-related *perceptions* of ability. It is possible that these perceptions of ability mask gender KSA differences for STEM subjects in college. The next section will detail gender-related perceptions of ability via the concept of stereotype threat.

Stereotype threat. Whether or not there are true KSA differences between females and males regarding STEM-related goals, beliefs regarding the ability of males versus females in certain areas may impact performance such that females underperform relative to males or underperform relative to what they would have in the absence of a particular stereotype threat. The notion of stereotype threat is that the fear of endorsing a stereotype about one's group acts as an interference for performance (Nguyen & Ryan, 2008). Individuals who are aware of negative stereotypes regarding their ability become anxious and underperform relative to their ability (Steele & Aronson, 1995). Other researchers posit that stereotype threat leads to negative thinking, which impairs test-taking functions like working memory capacity (Cadinu, Maass, Rosabianca, & Kiesner, 2005).

There is empirical evidence that when an ability is stereotyped to be high for one group of individuals and low for another group of individuals, the anxiety related to confirming the stereotype might interfere with performance. For instance, males are stereotyped to be better at math than females (Steele, Spencer, & Aronson, 2002). Researchers found that when such gender stereotypes were made salient to individuals, females would regularly underperform in mathematics tasks relative to males (Steele & Aronson, 1995). However, when such gender stereotypes were not made salient, researchers found no differences in performance between males and females (Davies, Spencer, Quinn, & Gerhardstein, 2002; Lesko & Corpus, 2006; Schmader & Johns, 2003).

Additionally, stereotype threat is more pronounced when the individual stereotyped identifies with the subject area. For instance, females who identified with math, meaning math was important to them, and who also had high GPAs in calculus, were most affected by the stereotype threat (Steinberg, Okun, & Aiken, 2012). Therefore, if females in STEM are aware of stereotypes regarding male superior math and spatial ability, it is possible that even though lower self-efficacy will lead to more effort applied in studying for examinations, stereotype threat will undermine their performance on the exam, especially if they are math/science identified. Additionally, it is possible that females in STEM majors enter college with higher KSAs than males, but stereotype threat masks this effect on performance. These hypothetical paths might occur even if females are not explicitly reminded of such stereotypes. Indeed, one study demonstrated that stereotype threat affected the performance of individuals without explicit mention and endorsement of stereotypes in the study (Galdi, Cadinu, & Tomasetto, 2014). As such, the current study investigated performance effects by assessing the effects of stereotype threat and science/math identification on performance. If performance differences are not found, or if males perform significantly better on exams than females, stereotype threat should be tested as an explanatory variable for this null or reverse finding.

Additional relevant variables. I obtained measures of other potential differences between males and females that might confound effort and performance in a classroom setting, such as number of credit hours one is taking, number of hours one works at a full-time or part-time job a week, number of extracurricular organizations in which one is involved, and number of hours a week one participates in the extracurricular organizations. If certain students are more involved in outside-class activities than others and this involvement differs by gender, it is important to measure. For instance, if females do not study more than males, but report spending more time at jobs and in extracurricular activities than males, these extracurricular variables may be alternative explanations for null findings. Further, information about individuals' specific college and major were measured to exclude non-engineering, mathematics, and physics majors.

Data Analysis Plan

There were two types of control variables in this study. The first type of control variable was measured for the purpose of Type I error. These control variables may be alternative explanations for why effects were found. The aim of including these variables (e.g., outcome expectancies, goal level, knowledge/skills/abilities) was to test for alternative explanations if support for the hypotheses were found. Thus, these variables were added as controls in the appropriate analyses if support for the hypotheses were found. The second type of control variable was measured for the purpose of Type II error. These control variables may be alternative explanations for why effects were *not* found. The aim of including these variables (e.g., value, stereotype threat, math/science identification) was to test for alternative explanations if support for the hypotheses were *not* found. Thus, relevant analyses were conducted if it was possible these Type II-relevant variables were important. In other words, the hypotheses were investigated alone first, followed by adding control variables if deemed appropriate. As a note, all variables were included in correlation tables and descriptive tables to provide complete information, regardless of their use as control variables.

Method

Participants

The current study aimed to investigate the hypotheses using a sample of male and female engineering, mathematics, and physics majors from Ohio University's Russ College of Engineering, Department of Mathematics, and Department of Physics & Astronomy, respectively. The participants were volunteers from mathematics courses required for many engineering majors (i.e., chemical and biomolecular engineers, civil engineers, electrical engineers, industrial and systems engineers, and mechanical engineers), mathematics majors, and physics majors in their second year and beyond (e.g., courses included MATH 3200, MATH 3210, MATH 3300, MATH 3400). To expand the study sample, this study was also conducted in the Russ College of Engineering MATH 1300 and MATH 2301 courses taught by engineering mathematics instructors. Participants were recruited from a total of 17 different sections. All course instructors were male. On average, the gender composition of the classes was reported to be 70.82% male and 29.18% female. This proportion was roughly the same in every course (i.e., the proportions ranged from 67% male and 33% female to 75% male and 25% female).

All participants in each section were invited to take the first and final surveys for $5 each after consenting to the study. These surveys included measures such as demographic measures, effort measures, ability measures, self-efficacy measures, science/math identification measures, and value measures. Then, participants were matched based on class and gender to determine who would be eligible to sign up for the extended, daily diary portion of the study. For example, if sixteen students from one section signed up for the study and six were female then the six females and six randomly selected males from that class section were invited to be in the extended study. Participants were paid $1 per daily diary taken. There were a total of 200 students recruited as participants. A total of 64 (32%) of these 200 students were female; therefore 64 of the 136 males were randomly selected to take part in the extended study. All 64 females were invited to take part in the extended study.

To focus on STEM majors, only engineering majors, mathematics majors, and physics majors were included in the data analysis. Seven were dropped from the sample because they were in other majors (e.g., math education). Participants had the option of taking any or all of the surveys (i.e., the initial survey, any/all of the daily diaries, the final survey). Some of the surveys had partial responses (i.e., participants started, but did not complete them). Participants' daily diary measures (i.e., self-efficacy, goal, daily, and weekly effort) were included if there were at least two diaries completed. Therefore, four individuals were dropped from the diary analyses due to having completed only one daily diary. All but two instructors provided exam scores (i.e., I obtained 184 scores for the first exam and 176 scores for the second exam). Ability variables (e.g., ACT and SAT scores) were obtained from the University Registrar. The University Registrar could only provide scores that students had sent to Ohio University. Thus, *n*'s varied for the various analyses. *N*'s for each variable are detailed in tables in the Results section.

There were 109 males and 56 females in the sample. The mean age of the sample was 19.65 (SD = 2.57); 86.7% of the sample was White/Caucasian, 4.2% of the sample was African American, 3.0% of the sample was American Indian or Alaskan Native, 2.4% of the sample was Asian American, and 5.5% of the sample identified as “Other”; 2.4% of the sample identified themselves as international students; and 21.8% of the sample identified themselves as first generation college students. In terms of year in school, 38.8% were first years, 28.4% were second years, 19.4% were third years, 6.7% were fourth years, and 6.1% were fifth years or beyond.

On average, participants were enrolled in 15.70 credit hours (SD = 2.84), participated in 1.31 extracurricular organizations (SD = 1.31) for an average of 3.84 hours a week (SD = 6.19), and worked at a job and/or volunteered at an organization or in a laboratory for 5.80 hours a week (SD = 9.21). Females reported partaking in more extracurricular activities (M = 1.80, SD = 1.43) for more hours a week (M = 5.60, SD = 8.98) than males (activities: M = 1.06, SD = 1.18; hours a week: M = 2.95, SD = 3.90) (both ps < .01). These variables were not related to effort (all ps > .05), but number of weekly activity hours negatively related to performance on the second exam, r = -.174, p = .034. No gender differences were found on work hours.

Measures

Most constructs were measured in multiple ways to test for and improve construct validity. For instance, self-efficacy was measured using three different scales and value was measured using five different scales. The majority of the variables were measured via self-report. However, both exam performance and ability were objective measures obtained via course instructors or via the Ohio University Registrar.

Self-efficacy. Pre-exam self-efficacy was measured in three ways. The first two ways are commensurate with the way self-efficacy is conceptualized by the non-monotonic, discontinuous model and with the recommendations of Bandura in the 2012 commentary. That is, self-efficacy was measured by assessing perceived goal progress (e.g., grade on an exam across a range of grades) per unit of effort (e.g., time devoted to studying). Participants answered a series of questions regarding their perceived capacity to achieve varying grades on the test given various amounts of time, which adapts measures in Vancouver and Kendall (2006) and Halper and Vancouver (2016). One measure asked participants to anticipate the number of hours needed to study to obtain each grade (e.g., A+, B-, C, etc.) (i.e., "When you consider the kind of material you need to learn for this course, how many hours do believe you will need to study between now and the exam to obtain at least a C on the second exam in this class?") ($\alpha = .97$). Participants also recorded the likelihood of achieving certain grades after studying for some number of hours (e.g., "What is the grade you expect to get on the second exam if you studied for 5 hours?") ($\alpha = .92$). Grades were translated into GPA points (i.e., 0-4.0, where an A = 4.0) for the purposes of analysis.

To approximate self-efficacy related to a certain goal, an individual may take multiple pieces of information into account. Individuals assess their expectancy for accomplishing a particular goal using information about themselves, information about the task, and information about other people (Gist & Mitchell, 1992). In other words, individuals compare themselves to other individuals (Festinger, 1954) on a particular task in the process of developing self-efficacy. Therefore, self-efficacy was also measured relatively. This third measure asked participants how difficult participants believed the material was for them in comparison to other students in the class (1 – *Not at all difficult* to 7 – *Very difficult*).

Effort. Participants recorded time spent studying each evening leading up to an exam via a daily diary as well as time spent studying during the previous week. Participants answered questions indicating the length of time during which they studied that day and that week. These questions were adapted from Vancouver and Kendall (2006). Additionally, participants reported how many hours they studied total for the second exam on the final questionnaire. There was one particular course of 17 students in which the instructor and students gave permission to look at objective data via a computerized homework system. There were 12 sets of homework problems to complete online with a total of 141 math problems across the 12 sets. For each set, there was information regarding whether each math problem was attempted, how many times each math problem was attempted, and the score on each problem after the final attempt. Therefore, I created three variables from this homework data: (1) percentage of total problems attempted, (2) number of problems retried, and (3) number of problems eventually gotten correct. In essence, this last variable is persistence. This homework counted for points in the course.

Performance. Because individuals who perform poorly on exams tend to overreport scores (Kuncel, Credè, & Thomas, 2005), the class instructor provided the exam grades of students participating in the study. Exam scores were all scaled to percentages for the purpose of comparison.

Demographics. Participants indicated their gender (males = 0; females = 1), year in school, age, ethnicity, international student status (domestic = 0; international = 1), and first generation college student status (not first generation = 0; first generation = 1). In addition, participants indicated their specific college (e.g., Russ College of Engineering, College of Arts & Sciences, Honors Tutorial College) and specific major (i.e., mathematics major, electrical engineering major, physics major, etc.). Individuals indicated how many credit hours for which they were registered in that particular semester. Participants recorded the number of extracurricular organizations of which they were a part. Participants listed how many hours a week they spent participating in extracurricular organizations. Participants also recorded work experience (i.e., how many hours a week they work), which included work in research labs for credit hours.

Value. Overall value for STEM-related disciplines was measured via 5 scales, including the math/science identification scale (see section below stereotype threat; Rios et al., 2015). The other four scales included a math enjoyment scale (Wang, Degol, & Ye, 2015; e.g., "I enjoy math", *1 = strongly disagree* to *7 = strongly agree*, αs > .85 for both the initial and final survey assessments of the measure), personal importance scale (Cech, 2015; e.g., "What is the personal importance to you of making scientific discoveries?", *1 = very unimportant* to *7 = very important*, αs > .73 for both the initial and final survey assessments of the measure), belongingness scale (Master, Cheryan, & Meltzoff, 2016; e.g., "How similar do you believe you are to other individuals taking this math class?", *1 = not at all* to *7 = very much,* αs > .85 for both the initial and final survey assessments of the measure), and a calling scale (Dik, Eldridge, Steger, & Duffy, 2008; e.g., "My area of study helps me live out my life's purpose" , *1 = not at all true of me* to *7 = absolutely true of me,* αs > .82 for both the initial and final survey assessments of the measure).

Gender discrimination beliefs (outcome expectancies). To measure participants' beliefs in gender bias in STEM classes, participants answered questions about the extent to which they believed there was gender bias and perceptions regarding how to handle such bias in their particular major classes after the second exam of the semester. Each participant indicated agreement to the following statements on a 1 (*strongly disagree*) to 7 (*strongly agree*) scale: "I feel that I need to work twice as hard compared to the majority of my classmates to get noticed by professors and employers, because of unconscious or conscious gender biases in evaluating grades," "I believe exams are graded objectively," and "The grade I obtain on an exam in this course reflects how well I actually did on the exam and does not reflect a bias in grading." The Cronbach alpha for this scale was .65.

Because it is also possible that perceptions of gender bias in the workplace more generally could affect work ethic in college, participants' perceptions of gender discrimination more generally were measured. Participants completed the Career Pathways Survey after the second exam. The Career Pathways Survey is a 38-item scale with Likert response items, ranging from 1 (*strongly agree*) to 7 (*strongly disagree*) (Smith, Crittenden, & Caputi, 2012). In the current study, anchors were 1 (*strongly disagree*) to 7 (*strongly agree*). The scale has demonstrated Cronbach alphas that range from 0.70 to 0.81 (Smith et al., 2012; Smith, Caputi, & Crittenden, 2012). There are four factors in the Career Pathways Scale: denial, resignation, resilience, and acceptance. The denial scale was the focus of the current study, as the denial subscale had items related to beliefs in gender discrimination at work. An example of a denial scale item is "Women starting a career today will face sexist barriers (reverse scored)." The Cronbach alpha for this subscale was .89.

Goal. Goal was assessed with the following question: "What is the lowest grade that you would be satisfied with on the upcoming mathematics exam?" (Mento, Locke, & Klein, 1992). Participants chose from a list of grades (i.e., A, B+, C-, etc.) in response to the question. Grades were translated into points (i.e., 0-4) for the purposes of analysis. Participants answered this goal question each evening during the daily diaries. Responses were averaged for the purpose of analysis.

Knowledge/skills/ability. To assess KSAs, participants answered a series of questions concerning their experiences in high school, which served as a measure of knowledge/skill/ability pre-college. Participants recorded their high school GPA and their high school class rank if remembered. The college ability measure included standardized test scores frequently used as a means of assessing ability in college admissions processes (Coyle, Purcell, Snyder, & Richmond, 2014). College GPAs, math placement scores, ACT scores, SAT scores, and AP scores were provided by the Registrar's office at the end of the semester.

Stereotype threat. A measure of the participants' perceptions of experiencing stereotype threat was adapted from Marx and Goff (2005). Participants responded to four statements ($\alpha = .71$) on a 1 (*strongly disagree*) to 7 (*strongly agree*) scale (e.g., "I worry that my ability to perform well on mathematics exams this semester is affected by my gender”). The answers to the four questions were averaged, with a greater score indicating greater perceptions of stereotype threat. Some of the questions were edited so that they reflected tests taken during the semester leading up to answering the questions instead of reflecting perceptions regarding future, hypothetical exams. All participants answered the following question indicating belief in a particular gender stereotype: “Though it may not be their fault, females cannot perform as well as males in science and mathematics classes” on a 1 (*strongly disagree*) to 7 (*strongly agree*) scale.

Math/science identification. Participants responded to a 25-item scale about their identification with science and math. The scale was adapted from Marsh and O’Neill (1984) and Rios, Cheng, Totton, and Shariff (2015). The original scales were focused solely on science identification (e.g., “I have never been very excited about science”, *1 = strongly disagree* to *7 = strongly agree*). The scale adapted for the current study included similarly-worded items about mathematics as well. Example questions from the current scale are: “I am quite good at math”, and “I have generally done better in mathematics courses than other courses” (αs > .83 for both the initial and final survey assessments of the measure). For all items from all scales in the Method section, see Appendix A.

Procedure

Initial data was collected with the Russ College of Engineering STEP UP (Summer Transition and Engineering Preview for Undergraduate Preparation) Program under the supervision of Dr. Jody Markley (n = 9) during the summer. The STEP UP Program is a six-week summer program with a goal of assisting incoming first-year engineering students from underrepresented groups with the transition from high school to college. However, the majority of the data were collected in multiple sections of mathematics courses during the Fall 2016-2017 semester. Participants, who were engineering, mathematics, and physics majors in a variety mathematics classes required for each major, were asked to complete a number of self-reported surveys during the first or second week of class, every day between the first and second exams of the semester, and within one week following the second exam of the semester. All individuals in each targeted class were recruited during the first two weeks of class. During the recruitment class period, participants had an opportunity to sign up for the study. Participants who signed up were compensated per survey completed (i.e., \$5 for the initial and final surveys and \$1 for every daily diary survey). Participants were told that they may not be selected for the entire study. Consent included permission for the university and/or university colleges to provide the experimenter with the participants' college GPA and various test scores. Consent also included permission for instructors to provide the experimenter with the participants' exam and homework scores, as applicable. After consent, participants were sent a questionnaire, which included questions regarding specific major, specific college, registered credit hours, extracurricular activities, high school GPA, college GPA, self-efficacy for the first exam, math/science identification, and the demographics instrument. The instructor provided the dates of the first two exams during the semester.

After the first exam, participants were given instructions for completing the online daily diary surveys. The daily diary methods can provide researchers with information about daily stressors, sleep habits, or study habits, because participants record information once or multiple times a day at scheduled intervals (Almeida, 2005). Often, researchers cannot assure that participants completed the measures at specific times when using paper-and-pencil methods. Therefore, daily diary measures via Internet web pages are now popular (Almeida, 2005). Beginning three weekdays after the first exam, participants were allowed to complete the daily diary effort measures before bedtime every evening leading up to the second exam of the semester. The questionnaires also included measures of self-efficacy and goal level each night. Participants were emailed a link each night at 9:30 pm requesting that they complete the daily diary. If participants did not complete the survey by 11:30 pm, they were sent a reminder. Participants were sent one more reminder each morning at 1:00 am. Participants were also texted once each night, if they provided the researcher with their cell phone number, to remind them to complete the survey. Participants responded to the questionnaires before bed each day with a cutoff of 4:00 a.m (Gartland, O'Connor, Lawton, & Ferguson, 2013).

Within one week following the second exam, participants were provided with an online questionnaire asking for perceptions of gender discrimination (i.e., outcome expectancies), perceptions of stereotype threat, value, and self-reported second exam effort. At the end of the semester, the instructors of each course provided the actual exam grades of the individuals who were participating in the study. The University Registrar provided college GPA, SAT scores, ACT scores, math placement scores, and AP scores. The dissertation was funded by the Graduate Student Senate Original Works Grant, the Department of Psychology Graduate Student Competitive Research Fund, the American

Psychological Foundation COGDOP Graduate Research Scholarship, and minimal advisor funds.

Results

Data Reduction and Construct Validation

Four constructs – self-efficacy, effort, ability, and value – were measured in multiple ways. Because the sample size was not sufficient for structural equation modeling, I used more traditional methods for reducing the data into scores for each construct. These methods and their results are described next.

Self-efficacy. Recall that self-efficacy was measured in three ways. One measure simply asked participants how difficult participants believed the material was for them in comparison to other students in the class. The two other self-efficacy measures from the daily diary included multiple items; thus, I first assessed the internal consistency of each of these two measures. The questions assessing how many hours one needed to study to achieve eight different grade levels displayed high reliability ($\alpha = .97$). Likewise, the five questions assessing what grade was anticipated based on increasing amounts of studying also displayed high reliability ($\alpha = .92$). Next, I examined the stability of the ratings relative to individual differences in the ratings using an intraclass correlation coefficient (McGraw & Wong, 1996). The ICC (2, k) was .80 for the three measures. Given this information, I averaged ratings over time for each participant such that each participant had one score per self-efficacy scale. For intercorrelations of the raw variables, see Table 1. I then calculated the reliability of the three measures. Given the high reliability coefficient ($\alpha = .93$), the measures were averaged and standardized. A higher z-score indicates higher self-efficacy.

An alternative method for translating the ratings from the multi-items scales into self-efficacy beliefs is to calculate the rate of change in grade expected given amount of hours studied. This alternative method provides a meaningful scale (i.e., a rate belief), but is not typically used (Bandura, 2012). Rather, the averaging method, which provides a scaled "area under the curve" index is more common. Nonetheless, a description of the rate of change method and its use to test Hypothesis 1 is provided in Appendix B.

Table 1

Correlations among self-efficacy variables.

	M	SD	1	2	3
1. Reported Hours to Obtain Grade	8.89	5.69			
2. Reported Grade Given Hours	2.94	0.75	.86**		
3. Reported Relative Difficulty	3.64	1.41	.70**	.60**	
4. Gender	0.34	0.47	-.15	-.21*	-.22*

Note. * $p < .05$, ** $p < .01$. Gender: Male = 0, Female = 1.

Effort. Effort was measured via reports of effort applied daily, in the past week, and since the previous exam. I also had objective data (i.e., effort applied on homework) for 17 individuals. Because individuals often save the bulk of studying for the last few hours before an exam (Bjork, Dunlosky, & Kornell, 2013), which was not captured through the daily or weekly questions on the daily diary (i.e., participants did not report hours crammed between the final daily diary and the exam), total effort was used in the analyses of the hypotheses. The other measures were used to validate the total effort measure. In particular, the total effort self-report is a retrospective measure of effort. To confirm that the retrospective nature of the total effort variable was not a problem, the daily, weekly, and homework variables were used to assess the convergent validity of the total effort measure.

For the homework data, there were three variables indexing effort: (1) percentage of total problems attempted, (2) number of problems retried, and (3) number of problems eventually gotten correct (i.e., persistence). Each of the 17 individuals had one score for each of these variables. The correlations among these variables are reported in Table 2. All three items had strong correlations (r's > .75) with each of the other three variables. The three variables also had high reliability when combined ($\alpha = .92$). Thus, they were combined and standardized.

Table 2

Correlations among homework effort variables.

	M	SD	1	2	3
1. Percentage Total Problems Attempted	69.63	26.09			
2. Number of Retries	183.24	105.94	.82**		
3. How Many Total Correct After Retries	43.00	20.29	.75**	.86**	
4. Gender	0.42	0.51	.24	.46	.49*

Note: * $p < .05$, ** $p < .01$. Gender: Male = 0, Female = 1.

Table 3 presents the interrcorrelations among all the effort measures. The highest response rate was on the total effort measure ($n = 99$). The daily diary measure had a strong correlation with the total effort measure ($r = .59$, $p < .001$), the weekly measure had a moderately strong correlation with the total effort measure ($r = .53$, $p < .001$), and the objective effort measure had a strong correlation with the total effort measure ($r =$

.70, p = .02). Given the strong correlations, the total effort measure had strong convergent validity.

Importantly, differences between the genders were in the hypothesized direction for each of the four effort measures, but not surprisingly these effects were smaller the smaller the time scale. That is, based on the self-reported total, females studied roughly five hours more than males, which was significant (p = .03, d = .46). Based on the self-report weekly time studying, females studied roughly one hour more than males, which was marginally significant (p = .10, d = .43). Based on the self-report daily time studying, females studied roughly 12 minutes more than males, which was not significant (p = .48, d = .15), but the difference was in the hypothesized direction for the daily measure as well. Finally, the genders differed on the objective effort score as well. That is, females put in more effort into their homework (M = 0.46, SD = 0.81) than males (M = -0.32, SD = 0.92), though the difference was only marginally significant, $t(15)$ = 1.80, p = .09, d = 0.90. Of course, the power for this analysis was low.

Table 3

Correlations among effort variables.

	M	SD	N	1	2	3	4
1. Daily Reported Effort Exam 2	0.93	0.83	85				
2. Weekly Reported Effort Exam 2	3.46	2.80	85	.72**			
3. Total Reported Effort Exam 2	12.17	10.57	99	.59**	.53**		
4. Objective Homework Effort	0.00	1.00	17	.75*	.60	.70*	
5. Gender	0.34	0.47	165	.08	.18	.22*	.42

Note: * p < .05, ** p < .01. Gender: Male = 0, Female = 1.

Value. There were several measures of value. Each scale was measured at two time points during the study (i.e., at the beginning of the semester and after the second exam). Test-retest reliabilities were .72 for the belongingness scale, .73 for the calling scale, .78 for the math value scale, .67 for the math and science identification scale, and .63 for the personal career importance scale. For the purpose of data reduction, the scores on each of the individual value measures at Time 1 were averaged together with their corresponding measure at Time 2 (e.g., belongingness ratings at Time 1 and Time 2 were averaged together). These averaged measures were then correlated with each other (see Table 4). Except for the correlations between belongingness and calling and belongingness and value for mathematics, all correlations were significantly positively related. The five value measures demonstrated strong internal consistency ($\alpha = .71$) and thus, all five scales across the two time points were standardized and averaged to create a value score per individual, though the math/science identification scale was also used alone when testing the hypothesis about stereotype threat. Of interest, the only gender difference on any of the value scale was that females ($M = 4.65$, $SD = 1.22$) reported a significantly lower personal importance for STEM career objectives (e.g., the importance of making important scientific discoveries, being a leader in one's field, managing future technologies, etc.) than males ($M = 5.31$, $SD = 1.10$), $t(95) = -2.75$, $p = .01$, $d = .57$.

Table 4

Correlations among value variables.

	M	SD	N	1	2	3	4	5
1. Belonging	5.00	1.03	97					
2. Calling	3.08	0.59	97	.17				
3. Math Value	5.60	1.06	97	.19	.32**			
4. Math/Science Identification	5.20	0.63	97	.45**	.27**	.62**		
5. Personal Importance	5.07	1.18	97	.34**	.33**	.23*	.38**	
6. Gender	0.34	0.47	165	.07	.12	.07	.05	-.27**

Note: * $p < .05$, ** $p < .01$. Gender: Male = 0, Female = 1.

Ability. There were several measures of ability included in the surveys. Table 5 presents the interrcorrelations among the ten raw ability variables, both exam score variables, and gender. There is theoretical reason to focus on the math/science ability measures in the current project because individuals in the study were in mathematics courses taking mathematics examinations. However, when including only math and scientific reasoning subsections of the standardized tests as well as the math placement, the reliability was substantially poorer ($\alpha = .70$) than when with all scores were included ($\alpha = .81$). Thus, an ability composite was created out of all ACT subsections (i.e., English, math, reading, science reasoning, and writing), all SAT subsections (i.e., math, critical reading, and writing), AP Calculus AB, and math placement data. Participants had to have data on at least half of the scores for their ability mean to be calculated. The scores were standardized and then averaged. Math placement scores include a range of 0-3. Given 3 is the highest math placement score at Ohio University, the math placement mean of the sample (M = 2.40, SD = 0.60) was fairly high. The ACT composite mean (M = 27.11, SD = 3.40) was about 5 points higher than the state-adjusted average in 2016 (M

= 22.47; Zhang, 2016), which is roughly 1.36 standard deviations above the state-adjusted average. The SAT composite mean (*M* = 1700.79, *SD* = 187.90) was about 50 points higher than the state-adjusted average in 2016 (*M* = 1652; Zhang, 2016), which is about a quarter of a standard deviation higher than the state-adjusted average. The AP Calculus AB score range was 1-5. In 2015, the AP Calculus AP national average was 2.86 (Edwards, 2015), making the sample's average (*M* = 2.90, *SD* = 1.59) close to the national average. Exam 1 and Exam 2 scores were in percentages. Average exam scores (~70%) were around a C average.

There were no significant gender differences on any of the individual ACT, SAT, math placement, or AP Calculus scores, except that males (*M* = 27.77, *SD* = 4.37) performed significantly better, $t(124) = -2.18$, $p = .03$, $d = .41$, on the science reasoning portion of the ACT than females (*M* = 26.02, *SD* = 4.01). There was no significant gender difference in standardized ability score (females: *M* = -0.13, *SD* = 0.68; males: *M* = 0.02, *SD* = 0.77), $t(124) = -1.06$, $p = .29$, $d = .21$.

GPA scores and high school class rank were also collected and examined. Given that GPA and class rank may contain motivation as well as ability effects (Cheng & Ickes, 2009; Richardson et al., 2012), these measures were not used to index ability. Moreover, including the GPA-related variables did not increase the internal reliability of the ability measure ($\alpha = .80$). Interestingly, females' GPA was significantly higher in high school ($p = .03$, $d = .35$, with a difference of .18 points) and college ($p = .002$, $d = .48$, with a difference of .32 points) than males. In addition, females (*M* = 23.32, *SD* =

33.39) reported a significantly higher ($t(81) = -2.00$, $p = .049$, $d = .42$) high school class rank than males ($M = 43.16$, $SD = 57.69$).

Table 5

Correlations among ability variables.

	M	SD	N	1	2	3	4	5	6	7	8	9	10	11	12
1. Math Placement	2.40	0.60	144												
2. ACT English	25.88	4.35	128	.30**											
3. ACT Math	27.02	3.31	128	.58**	.52**										
4. ACT Reading	27.70	4.81	128	.31**	.69**	.42**									
5. ACT Science	27.16	4.30	128	.43**	.54**	.54**	.58**								
6. ACT Writing	7.42	1.23	53	.26	.12	.22	.01	.05							
7. SAT Math	611.05	71.01	38	51**	.39*	.76**	.36	.27	.12						
8. SAT Reading	563.68	84.42	38	.34**	.57**	.28	.62**	.27	.26	.32*					
9. SAT Writing	526.05	75.36	38	.49**	.61**	.41*	.55**	.52**	.59*	.47**	.67**				
10. AP Calculus	2.90	1.59	41	.50**	.52**	.60**	.31	.39*	.16	73*	.77*	.53			
11. Exam 1	68.36	21.30	184	.21*	.26**	.26**	.11	.01	.17	.40**	.10	.27	.42**		
12. Exam 2	71.48	19.97	176	.32**	.15	.13	-.05	-.02	.32*	.27	.22	.27	.08	.45**	
13. Gender	0.34	0.47	165	-.04	.03	-.17	-.02	-.20*	.15	-.13	.11	.28	-.01	.07	-.03

Note: * *p* < .05, ** *p* < .01. Gender: Male = 0, Female = 1.

Descriptives

Table 6 presents means, standard deviations, correlations, and Cronbach's α on all key variables in the project. Looking at the table, some relationships stand out. For example, there was a moderately strong positive correlation between the score on the first exam and the score on the second exam ($r = .45$, $p < .001$). Additionally, the correlation between performance on the first exam and self-efficacy, assessed after the exam, was stronger ($r = .75$, $p < .001$) than the correlation between self-efficacy and performance on the second exam ($r = .44$, $p < .001$). This corroborates the theory that self-efficacy is influenced by past performance. Given that the first exam score was individuals' most proximal measure of their past performance, or ability in that course, it logically follows that self-efficacy would be more highly correlated with the first exam score than the second exam score. Despite this, the females in the sample, who did slightly better on the first exam ($r = .07$, $p = .39$), had significantly lower self-efficacy ($r = -.23$, $p = .03$). This implies that though the previous exam influences self-efficacy to a large extent, gender appears to create an offset.

Another interesting correlation was between ability and the exam scores. Specifically, the ability composite was significantly related to the score on the first exam ($r = .24$, $p = .01$), but not the second exam ($r = .14$, $p = .13$). This may be because the first mathematics exam included more review material that was on standardized tests and thus reflected in the composite ability scores more than the second exam. Similarly, effort in preparing for the second exam was significantly negatively related to the first exam score ($r = -.30$, $p = .003$), but unrelated to second exam score ($r = -.13$, $p = .21$). This result implies that individuals may have used their knowledge of their ability (i.e., which could be partially indexed in self-efficacy) to determine how much to study. Those who had higher ability may have studied less, but still performed better on the exam than those with lower ability who studied more. To assess this possibility I controlled for ability, performance on the first exam, or both together. Controlling for both together caused the regression weight of effort predicting performance to be positive for both males and females (*bs* = .11 for both genders), but effort applied toward preparing for the second exam was still not significantly related to performance on the second exam with these controls in the analysis (ps > .70). Controlling for the variables separately did not make a difference in the result; effort did not significantly predict performance regardless. Finally, value was moderately positively related to self-efficacy ($r = .30$, $p = .02$) and performance on the first exam ($r = .25$, $p = .02$), but unrelated to effort ($r = .02$, $p = .88$) or performance on the second exam ($r = .12$, $p = .27$).

Table 6

Correlations among key variables.

	M	SD	N	1	2	3	4	5	6	7	8	9	10	11	12
1. Self-Efficacy	0.00	1.00	87	**.93**											
2. Gender	0.34	0.47	165	-.23*											
3. Goal	3.46	2.80	87	.71**	-.25*										
4. Effort Exam 2	12.17	10.57	99	-.58**	.22*	-.20									
5. Exam 1	68.36	21.30	184	.75**	.07	.44**	-.30**								
6. Exam 2	71.48	19.97	176	.44**	-.03	.35**	-.13	.45**							
7. Exam Bias Beliefs	2.34	1.21	98	-.21	.06	-.28*	.10	-.28**	-.35**	**.65**					
8. Workplace Bias Beliefs	4.09	1.25	98	.07	.43**	-.01	.02	.20*	.09	.04	**.89**				
9. Ability	0.00	1.00	128	.36**	-.10	.42**	-.13	.24**	.14	-.33**	-.02	**.81**			
10. Stereotype Threat	2.04	1.44	97	-.24	.41**	-.32*	.28**	-.11	-.32**	.46**	.25**	-.31**	**.71**		
11. Math/Science ID	5.20	0.63	97	.44**	.05	.31*	-.05	.32**	.08	-.39**	-.10	.20	-.13	**>.83**	
12. Value	0.00	1.00	97	.30*	-.03	.18	.02	.25*	.12	-.29**	-.18	.01	-.03	.75**	**.71**

Note: * $p < .05$, ** $p < .01$. Cronbach's α is presented in bold on the diagonal. Gender: Male = 0, Female = 1.

To get a better handle on gender differences, Table 7 presents descriptive statistics on all predictor, mediator, outcome, and control variables separated by gender. Recall that self-efficacy, ability, and value were all transformed into z-scores. Females reported having a lower goal by about a half a grade level ($M = 2.81$, $SD = 0.61$) than males ($M = 3.09$, $SD = 0.46$), $t(83) = -2.41$, $p = .02$, $d = 0.52$. Females ($M = 2.45$, $SD = 1.37$) did not rate the exams as more or less biased against females than males ($M = 2.29$, $SD = 1.12$), $t(96) = 0.63$, $p = .53$, $d = .13$. However, females did report that they believed females were likely discriminated against at work ($M = 4.81$, $SD = 0.90$) more than males reported this ($M = 3.69$, $SD = 1.24$), $t(96) = 4.71$, $p < .001$, $d = 1.02$. Females also endorsed experiencing more stereotype threat ($M = 2.83$, $SD = 1.49$) than males ($M = 1.59$, $SD = 1.22$), $t(95) = 4.43$, $p < .001$, $d = 0.91$. However, females did not report significantly lower math/science identification ($M = 5.24$, $SD = 0.63$) than males ($M = 5.18$, $SD = 0.63$), $t(95) = 0.20$, $p = .84$, $d = 0.10$, nor a lower value for STEM ($M = -0.04$, $SD = 0.93$) than males ($M = 0.02$, $SD = 1.04$), $t(95) = -0.28$, $p = .78$, $d = 0.06$.

Table 7

Descriptive statistics by gender and accompanying inferential statistics

Variable	Female Mean (SD)	Male Mean (SD)	*t* (*df*)	*p*
Self-Efficacy Measure				
Self-Efficacy	-0.24 (0.97)	0.19 (0.80)	-2.18 (*83*)	.03
Goal Measure				
Goal	2.81 (0.61)	3.09 (0.46)	-2.41 (*83*)	.02
Effort Measure				
Total Effort (i.e., Hours)	15.22 (10.72)	10.42 (10.15)	2.22 (*97*)	.03
Performance Measures				
Exam 1 Performance	73.03 (21.52)	70.02 (19.92)	0.86 (*151*)	.39
Exam 2 Performance	72.66 (18.33)	73.70 (20.81)	-0.30 (*146*)	.77
Gender Bias Measures				
Exam Gender Bias Beliefs	2.45 (1.37)	2.29 (1.12)	0.63 (*96*)	.53
Workplace Gender Bias Beliefs	4.81 (0.90)	3.69 (1.24)	4.71 (*96*)	<.001
Ability Measure				
Ability	-0.13 (0.68)	0.02 (0.77)	-1.06 (*124*)	.29
Stereotype Threat-Related Measures				
Stereotype Threat	2.83 (1.49)	1.59 (1.22)	4.43 (*95*)	<.001
Math and Science ID	5.24 (0.63)	5.18 (0.63)	0.20 (*95*)	.84
Value Measure				
Value	-0.04 (0.93)	0.02 (1.04)	-0.28 (*95*)	.78

Hypothesis Tests

The first hypothesis stated that gender would be related to self-efficacy. As noted above, this hypothesis was supported. That is, females had significantly lower self-efficacy (M = -0.24, SD = 0.97) than males, (M = 0.19, SD = 0.80), $t(83) = -2.18$, $p = .03$, $d = .48$ (see also, Appendix B). These results also provide evidence that females have lower self-efficacy for certain subgoals relevant to a STEM career, such as STEM-related exam goals, in addition to superordinate goals (e.g., self-efficacy for being hired as an engineer).

Hypothesis 2 predicted that lower self-efficacy for STEM-related goals was related to more effort applied in goal striving. Indeed, the higher one's self-efficacy the less effort individuals applied, $b = -7.63$, $SE = 1.43$, $t(57) = 5.34$, $p < .001$, $R^2 = .337$. Therefore, Hypothesis 2 was supported. I also confirmed that gender was not a moderator in this relationship ($p = .26$).

Hypothesis 3 predicted that there would be a gender difference in effort. There was a gender difference in total reported hours studying, $t(97) = 2.22$, $p = .03$, $d = .46$. Females reported studying significantly more hours for the second exam total (M = 15.22, SD = 10.72) than males (M = 10.42, SD = 10.15). Indeed, the females reported studying about 50% more than the males, on average. Thus, Hypothesis 3 was supported.

As noted above, there was a gender difference in self-efficacy, a gender difference in total number of hours studied for the second exam, and self-efficacy predicted effort (i.e., hours studied) on the second exam. Thus, the first three criteria for mediation were met. When effort was regressed onto gender (male = 0; female = 1) controlling for self-efficacy, the relationship between self-efficacy and effort remained significant, b = -7.98, SE = 1.46, t(55) = 5.48 p < .001, whereas the relationship between gender and effort dropped to b = -2.88, SE = 2.52, t(55) = -1.14, p = .26, total R^2 = .35. This pattern of results is consistent with a full mediation model. Moreover, a bootstrapping analysis with 1,000 estimates revealed a significant indirect effect of gender, through self-efficacy, on effort (*estimate of indirect effect* = 2.83, SE = 1.58, 95% CI: 0.081 to 6.270), and no direct effect on effort (*estimate of direct effect* = -2.88, SE = 2.52, 95% CI: -7.928 to 2.165), consistent with full mediation (Preacher & Hayes, 2004). Figure 3 shows this mediation model.

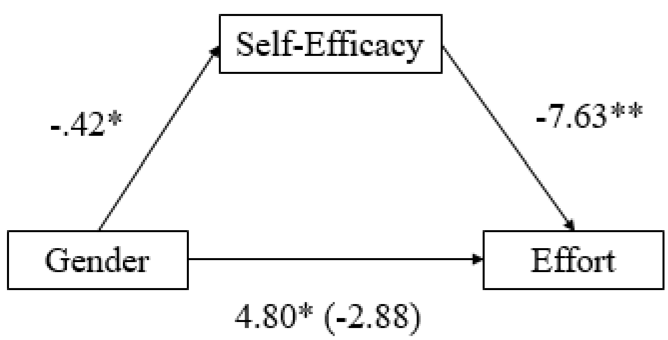

Figure 3. Mediation model of gender and self-efficacy on total effort or total number of hours studied (unstandardized coefficients). Indirect effect of gender, via self-efficacy, on effort in parentheses. Gender: Male = 0; Female = 1
*p< .05.**p< .001.

Hypothesis 5 predicted a gender difference on performance and Hypothesis 6 predicted a mediating effect of effort between gender and performance. There was no difference between females (M = 73.03, SD = 21.52) and males (M = 70.02, SD = 19.92) on the first exam performance, t(151) = 0.86, p = .391, d = .15, and no difference between females (M = 72.66, SD = 18.33) and males (M = 73.70, SD = 20.81) on the second exam performance, $t(146) = -0.30$, $p = .766$, $d = .05$. Thus, Hypothesis 5 was not supported. Hypothesis 6, which predicted that effort would mediate the gender-performance relationship, was not tested because conditions for mediation were not met. As a note, the effort-performance effect is likely very small because effort is applied to compensate for ability (Vancouver et al., 2008).

Assessing Confounding and Limiting Effects

Support for adding control variables (i.e., values, outcome expectancies for future exams and careers, and goals) to the analyses was assessed. Though females had slightly lower values for STEM, the difference was not significant ($p = .78$) and females put more effort into STEM exam preparation than males ($p = .03$). Thus, value was not tested as an alternative explanation. In terms of outcome expectancies and beliefs in exam biases, there was no difference between the genders on perceptions of exam bias. Further, feelings of exam bias were not related to overall studying (i.e., effort), $r = .10$, $p = .35$, or self-efficacy, $r = .21$, $p = .11$. Therefore, given that exam bias had questionable internal consistency and was not related to other variables, perceptions of exam bias was not tested as a control in the effort mediation analyses. There was a gender difference on the workplace bias scale such that females reported a stronger perception that females were discriminated against at work. Feelings of gender bias in the workplace were not related to overall studying (i.e., effort), $r = .02$, $p = .89$, or self-efficacy, $r = .07$, $p = .61$. Because the belief in discrimination at work variable was related to gender, the denial subscale score on the Career Belief Scale was entered as a control variable in the mediation analyses. Finally, there was a difference in reported goal level in that females reported being satisfied with a lower grade on the exam than males. Goal was not related to effort, $r = -.20$, $p = .14$, but was strongly related to self-efficacy, $r = .71$, $p < .001$. Further, gender did not moderate the effect of goal level on effort, $b = 2.89$, $p = .71$. Goal was also included as a control in the gender-effort mediational analyses given the differences in goal level between the genders and given its relationship with self-efficacy. As a note, work hours and extracurricular hours were not included as controls because their purpose was to serve as a possible alternative explanation if there were no gender differences in effort.

When effort was regressed onto gender (male = 0; female = 1) controlling for self-efficacy, goal, and beliefs in gender bias in the workplace, the relationship between self-efficacy and effort remained significant, $b = -9.62$, $SE = 1.83$, $t(55) = 5.27$ $p < .001$, whereas the relationship between gender and effort dropped to $b = -0.12$, $SE = 2.87$, $t(55) = -0.04$, $p = .97$, total $R^2 = .43$. A bootstrapping analysis with 1,000 estimates revealed a significant indirect effect of gender, through self-efficacy, on effort (*estimate of indirect effect* = 3.35, *SE* = 1.63, 95% CI: 0.741 to 7.308), and no direct effect on effort (*estimate of direct effect* = -0.12, *SE* = 2.87, 95% CI: -5.865 to 5.635), consistent with full mediation (Preacher & Hayes, 2004). Thus, there is evidence that even when controlling for goal level and outcome expectancies, gender has an indirect effect on effort through self-efficacy. Figure 4 shows this mediation model.

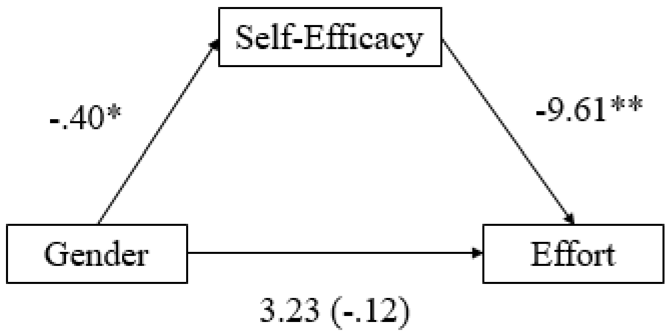

Figure 4. Mediation model of gender and self-efficacy on total effort or total number of hours studied with controls (unstandardized coefficients). Indirect effect of gender, via self-efficacy, on effort in parentheses. Gender: Male = 0; Female = 1
$*p < .05$. $**p < .001$.

Although Hypothesis 5 and 6 were not supported, recall that stereotype threat was hypothesized as a possible explanation for null performance results. There was a gender difference in stereotype threat. Females reported experiencing more stereotype threat than males. More stereotype threat experience predicted a lower exam score on the second exam, $b = -4.60$, $SE = 1.42$, $t(91) = -3.24$, $p = .002$, $R^2 = .104$, but did not predict the exam score on the first exam, $b = -1.81$, $SE = 1.72$, $t(91) = -1.05$, $p = .296$, $R^2 = .012$. There was not an interaction between gender and stereotype threat in the prediction of the second exam score ($p = .628$). However, there was evidence of influential cases, so this regression was performed with and without influential cases according to standard diagnostics. Specifically, when four influential cases, which were identified cases with leverage values (leverage value > (2k+2/n) or Cook's distance values (Cook's d > 4/n) that exceeded the rules of thumb in a diagnostic test, were removed, there was a significant interaction between gender and stereotype threat, $p = .026$, such that females performed significantly worse on the second exam when stereotype threat was greater $b = -7.10$, $SE = 3.55$, $t(87) = -1.20$, $p = .049$, $R^2 = .30$, but males did not, $b = 7.99$, $SE = 5.67$, $t(87) = 1.41$, $p = .163$, $R^2 = .30$. Further, there was not a three-way interaction including gender, stereotype threat, and math/science identification on performance on the second exam, $p = .434$. Neither gender was more highly identified with math and science.

Stereotype threat was also tested as a control in the gender-performance relationship to further determine its possible impact on performance. When controlling for stereotype threat in a regression where gender predicts performance, gender marginally predicted performance on Exam 1 such that females performed marginally better than males, $b = 8.50$, $SE = 5.42$, $t(90) = 1.57$, $p = .12$, $R^2 = .038$. This regression was performed with and without influential cases according to standard diagnostics. When four influential cases were removed, gender predicted performance such that females performed significantly better on the first exam than males, $b = 9.82$, $SE = 5.10$, $t(86) = 1.94$, $p = .05$, $R^2 = .217$. Controlling for stereotype threat on the second exam changed the sign of the beta (i.e., rather than gender predicting performance in favor of males, gender predicted performance in favor of females), but gender did not significantly predict performance, $b = 2.98$, $SE = 4.52$, $t(90) = 0.66$, $p = .511$, $R^2 = .108$.

Discussion

The purpose of this study was to examine gender differences in goal striving for STEM-related goals. In the current study, I predicted that females would have lower self-efficacy for mathematics exams for courses related to their major. I predicted that this low self-efficacy would encourage effort applied towards studying for such examinations and perhaps facilitate performance on these examinations as well. I found support for the effort hypotheses, but not the performance hypotheses. Multiple theoretical, methodological, and practical implications are discussed below. I also address limitations and ideas for future research in each section.

Theoretical Implications

Perhaps the most important theoretical contribution of this study is that it highlights a theoretical gap in the STEM education literature. Namely, the focus of the STEM education literature has been on goal choice processes, which includes investigations into why fewer females declare STEM majors and enter STEM fields. STEM education researchers have found a gender difference in the level of self-efficacy for STEM-related goals and a positive effect of self-efficacy on goal choice for STEM. Specifically, females are less likely to enter STEM fields because their self-efficacy for STEM careers is lower (i.e., which is the positive effect) than that of males (i.e., the gender difference). However, as ascertained in the broader psychological research literature, self-efficacy can play multiple roles depending on the exact motivational process examined. Very few studies have investigated how self-efficacy functions in individuals who have *already* committed to a STEM field (e.g., declared a STEM major).

These STEM majors are striving for their chosen goal, which is a process that has not been studied. This paucity of research inspired the current study. In particular, given the evidence that during the post-goal choice (i.e., goal striving) process, individuals with higher self-efficacy put less effort towards their goals than those with lower self-efficacy, I predicted that a) females would have lower self-efficacy than males for course-related goals such as exam goals and b) the low self-efficacy of females would facilitate effort (e.g., studying for exams). Hypothesized subgroup differences were found for both self-efficacy and effort. Females had lower self-efficacy than males and studied more for examinations in their mathematics courses. Male and female engineering, physics, and mathematics majors demonstrated the hypothesized negative relationship between self-efficacy and effort post-goal choice. Therefore, the current study discovered a gender difference and a negative effect of self-efficacy in STEM education during goal striving. This represents a novel prediction and theoretical contribution to the STEM education field.

The current study also speaks to the broader motivational literature and, in particular, speaks to the multiple viewpoints regarding the role of self-efficacy in motivation. One view holds that high self-efficacy is motivating, regardless of the goal process considered. This is the view expressed in social cognitive (Bandura, 1986) and self-efficacy theory (Bandura, 1997). The other view holds that self-efficacy can have multiple effects on motivation, depending on the goal process considered. This is the view expressed in self-regulation theories based on control theory (e.g., Carver & Scheier, 1998; Vancouver, 2005). Both views agree that during goal choice high self-efficacy is positively motivating such that individuals are more likely to attempt goals when self-efficacy is high. During goal striving, however, the control theory view predicts that self-efficacy informs individuals about their current state and can therefore affect the discrepancy between current states and desired states, or it represents their belief in the rate of discrepancy reduction based on efforts. Specifically, individuals with higher self-efficacy will believe they are closer to their goal or can reach it with fewer resources and therefore can afford to put in less effort toward the goal than those with lower self-efficacy (Vancouver et al., 2008). Due to the universal agreement regarding the effect of high self-efficacy during goal choice, as well as current evidence in the STEM education literature, the current study did not investigate goal choice. The current study did investigate goal striving, and the results indicate even more support for the already substantiated control theory viewpoint of self-efficacy: regardless of sex, there was a negative effect of self-efficacy on behavioral effort during goal striving. However, this study contributes a unique finding to the debate regarding the role of self-efficacy during goal striving, because it is the only between-person field study to demonstrate such effects.

Indeed, although there are many empirical studies that have examined the competing predictions of self-efficacy's effect on motivation that have shown overwhelming support for the control theory point of view, this empirical research was criticized for being found only in unrealistic laboratory situations by researchers holding the social cognitive theory view (e.g., Bandura & Locke, 2003). In response, researchers have begun investigating the relationships in field settings (Vancouver & Kendall, 2006).

For example, the Vancouver and Kendall studied students taking exams and found the negative effect of self-efficacy on planned and reported study time across the exams (i.e., covariance was within-person). The current study is another example of a field study that can counter criticisms against laboratory-based control theory research. Similar to the Vancouver and Kendall, the current study contributed support for the nonmonotonic, discontinuous model of self-efficacy (Vancouver et al., 2008) and the viewpoint that low self-efficacy facilitates effort by demonstrating a negative effect of self-efficacy on effort with students taking actual exams. Additionally, by demonstrating these relationships in different majors (i.e., engineering, mathematics, and physics), different courses (i.e., mathematics courses), and different subgroups (i.e., males and females), the current study provides some additional external validity for the model beyond that which the Vancouver and Kendall (2006) paper provided (i.e., that paper investigated psychology majors in a psychology course without making any distinctions between particular subgroups such as gender).

The current study also theoretically extends the Vancouver and Kendall study by way of the conceptualization of self-efficacy. In the Vancouver and Kendall paper, researchers measured self-efficacy for current capacity (e.g., “If you were to take the exam right now, what grade would you expect to get?”). The current study measured one’s belief in one’s capacity to learn (e.g., “How many more hours of studying will you need to obtain an A?” or “What grade would you obtain on the exam with 10 more hours of studying?”). Belief in one’s ability to learn (i.e., learning self-efficacy) is likely more important than belief in one’s current perceived capacity for any related task (i.e., performance self-efficacy) such as an exam when change in capacity is possible. This learning-focused conceptualization of self-efficacy seemed more appropriate for this particular investigation because individuals are in training for their careers. Therefore, individuals in the study were in the process of learning and increasing knowledge and skills to eventually obtain positions in their chosen career paths. In other words, individuals were in the process of *changing*, or *expanding*, their capabilities for engineering, mathematics, and physics. The implications of using this particular conceptualization of self-efficacy are twofold. First, by investigating and finding a negative effect of *learning* self-efficacy on motivation rather than *performance* self-efficacy, the current paper provides a theoretical extension of past field studies in addition to replication (i.e., Vancouver & Kendall, 2006). Indeed, this learning versus performance distinction is also important because there may be specific situations in which a high performance self-efficacy may be beneficial in goal striving in STEM courses particularly (e.g., high self-efficacy may facilitate a female voicing her ideas in a group project with all male colleagues, or during a project presentation to instructors and classmates). Second, and more broadly, the current study increases the generalizability of the non-monotonic, discontinuous model of self-efficacy to be applicable to a learning-based self-efficacy in addition to performance-based self-efficacy.

The conceptualization of self-efficacy included an element of relative comparison in addition to belief in capacity to learn. That is, not only did individuals with low self-efficacy believe they had a more restricted capacity to learn, but also these individuals reported that the material was more difficult for them compared to their peers. This relative comparison component has theoretical implications for literature on the academic self-concept and the big-fish-little-pond (BFLP) effect. The BFLP effect describes a particular situation where social comparison is key in developing beliefs about the self. An academic self-concept is the self-perception of abilities related to school (Marsh, Byrne, & Shavelson, 1988). The academic self-concept is domain-specific, meaning one can have, for example, an engineering self-concept, a linguistics self-concept, and a biological sciences self-concept (Marsh et al., 2008). These self-concepts are similar to self-efficacy (i.e., one has a certain belief in capacity for engineering and a certain level for linguistics). According to the BFLP theory, academic self-concepts are heavily influenced by the context of the individual's academic environment or the average abilities of other students (Marsh & Parker, 1984), along with objective indicators of one's abilities such as exam and standardized test scores (Marsh et al., 2008). For instance, two individuals with comparable academic biological sciences capabilities might have vastly different biological sciences self-concepts if one individual is attending a biological sciences high-ability school (i.e., a school where the average student achievement in biological sciences is relatively high) and one individual is attending a biological sciences low-ability school (i.e., a school where the average student achievement in biological sciences is relatively low; Marsh, 1987). The individual attending the biological sciences high-ability school will have a lower academic self-concept than the individual attending the biological sciences low-ability school, even though both individuals' biological sciences abilities are similar. This effect also works in the reverse. If the same individual with the low academic self-concept at the high ability school were to transfer to a lower ability school, this individual would then feel like a "big fish"; the individual would perceive his or her ability to be relatively high in comparison with the average student at the low ability school. Thus, the BFLP effect (i.e., or how one stands up to the perceived abilities of others) contributes to the creation of one's academic self-concept.

There are two theoretical implications of the current study to literature regarding the academic self-concept. The first is that, given the high correlations between one's perceived capacities to learn and the variable that measured relative difficulty of learning material compared to peers, the current study supports the research claiming the two constructs are related, although cannot necessarily verify a causal direction or, more likely, reciprocal influences. The second theoretical contribution is that the current study may expand the academic self-concept research to apply to goal striving outcomes other than goal choice outcomes, which has been the focus of academic self-concept researchers' attention. Specifically, the academic self-concept is related to class and academic major selection (Marsh, 1991; Marsh, Kuyper, Morin, Parker, & Seaton, 2014). Although there is substantial evidence that having a lower academic self-concept may reduce the desire for and choice of educational attainment (e.g., an individual with a low biological sciences academic self-concept will not choose to take AP Biology in high school nor major in biology in college; Guay, Larose, & Biovin, 2004), little research has been conducted on the effect of a low academic self-concept (i.e., or being a little fish) on one's goal-directed behavior post-goal choice. Perhaps, similar to what the current study found, individuals who feel like "little fish in a big pond" are more motivated for the topic in which they have relatively low academic self-concept than those who have higher academic self-concepts. If so, there could be advantages to having a lower academic self-concept. For example, these individuals may learn more about the topic (e.g., biological sciences) from putting in more effort. This idea needs to be systematically studied before drawing further conclusions, but if substantiated, this research could contribute to broadening the scope of the BFLP effect.

Like many studies conducted in academic settings with actual students, the current study obtained a high degree of ecological validity (i.e., generalizability) to a real-life setting (Chaytor & Schmitter-Edgecombe, 2003). This type of generalizability is an advantage in that the conclusions of the study have more implications for individuals in STEM fields than those conducted in less ecologically valid laboratory settings. However, the ecological validity comes at a methodological cost in that many variables that could typically be manipulated for better control in laboratory studies needed to be measured. The next section describes the strengths and weaknesses of these measurements, as well as details implications for future methodology.

Methodological Implications and Limitations

There were many strengths in the measures used in the current study. However, as with any study, there is room for improvement in future measurement and methodology. Additionally, given the lack of manipulated variables, there were multiple third variables that needed to be measured and controlled to make more accurate inferences about STEM majors. The purpose of this section is to detail the successes and limitations of the measurements of the predictor variables, the criterion variables, and the control variables.

Arguably the most important variable in the current study was self-efficacy. Self-efficacy was measured in a way to counter past measurement criticism from skeptics of the control theory view of self-efficacy (e.g., Bandura, 2012) as well as to adhere to general self-efficacy measurement standards (i.e., Bandura, 1997). The main methodological criticism of the past field studies demonstrating negative effects of self-efficacy was that researchers used a single item to measure self-efficacy (Vancouver & Kendall, 2006). Further, this single item asked individuals to estimate their grades, which may have been influenced by factors other than one's self-efficacy, such as life circumstances outside one's control (Bandura, 2012). In the current study, self-efficacy was measured with multiple multi-item scales that took into account a range of grade levels, as per suggestions in the 2012 Bandura commentary (e.g., students rated likelihood of obtaining certain grades). Further, the current study's measurement focused on direct perceptions of one's learning capabilities for exam performance so that the rating yielded less noise, or variability due to other factors besides self-efficacy. For example, an individual answering the self-efficacy question, "How many more hours would you need to study to obtain a B on the upcoming exam in your mathematics course?" would not take into account work schedules or the fact that he or she needed to take a friend to the emergency room in the answer. In contrast, it is possible that the individual could have taken these unrelated factors into account when answering the one item grade prediction measure of self-efficacy, which introduces noise to the measurement. Additionally, as per self-efficacy measurement recommendations, the measure was consistent with the target (i.e., STEM-related classroom exam self-efficacy was operationalized as grade levels on an exam; Bandura, 1997). Thus, self-efficacy was measured in a way that was consistent with the measurement standards of highly regarded researchers in the area.

There were also some unique features of the way self-efficacy was measured in the current study, and these unique features further improved the quality of the measurement. The first feature was that the units of measurement were meaningful. In particular, two of the measures of self-efficacy assessed how much goal progress one believed one could make per unit of effort (i.e. unit of time spent studying) applied. By operationalizing self-efficacy as time studying to achieve grades, the current study measured self-efficacy in meaningful units commensurate with the dependent variable (i.e., effort). Self-efficacy is rarely measured in meaningful units.

The second unique feature of the measurement was that it identified the self-perceived discrepancy between one's current state and one's desired state (e.g., "I am 55 minutes of studying away from getting an A"), which is also rarely captured in self-efficacy measurements (i.e., often, self-efficacy strength is measured on a 0-10 scale; Vancouver, Alicke, & Halper, in press). These unique features are important because they assist in measuring self-efficacy in a way that matches the definition. That is, the method of measurement in the current study takes into account one's belief in one's current capacity relative to a desired state in meaningful effort (i.e., time) units. As a note, self-efficacy was not measured as meaningfully in the item regarding one's belief in one's ability compared to peers. The semantic differential scale of the relative comparison self-efficacy item is a good example of the units of measurement typically used in measurements of self-efficacy.

The third unique feature of the current methodology is that the self-efficacy measures were collected multiple times over the course of the semester. In particular, the daily measurements of self-efficacy capture changing self-efficacy as one studies better than one measure of self-efficacy. Self-efficacy is often measured just once during the course of one study. By adhering to and improving upon self-efficacy measurement standards, one methodological strength of the current study is that the key mediator was measured in such a way that was uniquely consistent with the definition of self-efficacy and able to better capture the dynamics the construct implies.

Another variable that possessed some methodological rigor was effort. Effort was measured in four ways with the intention of validating what was the most encompassing measure of effort: total effort (i.e., hours) put towards studying for the exam. As a reminder, it is common for students at any level to apply the strategy of cramming, or putting off studying until the last few days or hours before an exam (Bjork et al., 2013; Taraban, Maki, & Rynearson, 1999), which was best measured by the item inquiring about total number of hours studied. However, to assure that the total effort variable demonstrated strong construct validity I also measured effort daily, weekly, and with a small sample, objectively via homework completed. There were strong positive relationships among these other measures of effort and total effort, although not as strong as the correlations among the self-efficacy measures, which provides further evidence that cramming likely occurred. Further, females put more effort into their mathematics exam preparation according to each self-reported measure. The difference was marginal for the weekly measures and in the direction expected, but not significant for the daily measures. This pattern held for the objective measure of effort as well. More so than males, females attempted more homework problems, retried more homework problems when a solution was incorrect the first time, and persisted on these problems until getting them completely correct. The effect for the objective effort analysis was large, but given the small sample size (n = 17), not reliable.

Despite successfully validating the total effort measure, there were some methodological lessons learned working with both daily diary data and objective data. Specifically, one of the goals of the daily diary data was to capture all studying completed during the day (i.e., before bedtime). To take into account varying bedtimes, I sent the daily surveys out at 9:30 p.m., 11:30 p.m., and 1:00 a.m. each night. Despite multiple reminders, one third of individuals reported to the researcher at the end of the study that they studied after taking the daily survey. Further, although the majority of individuals invited to be part of the extended study took at least one daily diary, on average only 12 out of 24 daily diaries were completed at all during the course of the study. Completing only half of the daily diaries reduces the validity of the daily measure, especially if participants did not study the same number of hours each evening, which according to the literature on learning and cramming, is likely (Bjork et al., 2013; Taraban et al., 1999). Any follow-up studies using such daily diary measures could correct these issues with equipment. For instance, actigraphs have been used for daily diary studies involving sleep (Chan, 2016). Participants could be given actigraphs to wear during the study, which could a) remind them to take the daily diaries and b) monitor actual bedtimes. If participants are told that the researcher will know when they go to bed from the actigraph's information, they may be more likely to adhere to the study protocol. An alternative to this idea is to ask individuals to report the previous day's studying the next morning.

There are also ways to improve objective effort data collection in the future. In terms of construct validity advantages, the objective measures provide an element of construct validity that the subjective measures do not. The biggest disadvantage of the objective effort data in the current study was that objective data was only obtained for 17 individuals in one course. The homework measure could be used in future studies as an objective measure of effort, instead of a measure used purely for criterion-related validity, with a greater sample size. Also, given advanced technology, a useful variable to add to an objective homework measure would be recorded time spent on each problem (i.e., especially if the problem was not gotten correct on the first attempt). One other disadvantage of the objective effort data was that the homework counted as a grade for students. Thus, it is hard to parse out if more homework attempts indicated greater motivation for learning the course material, a greater motivation for high exam grades, or if more homework attempts indicated a greater value placed in the grade associated with homework completion, or some combination of the three explanations. One solution is to measure homework effort on homework problem sets that are optional for students and do not count towards students' grades. Other useful objective indicators of STEM-related effort might include number of internships applied for or number of hours spent on projects (i.e., either alone or with a group, but logged by swiping in and out of a project lab).

Low power and poor response rates were not the only possible weaknesses of the effort measures. There are some weaknesses of the effort measures that may explain why the effects of gender and self-efficacy on effort did not translate to exam performance. Despite equivalent ability levels indexed by standardized test scores, the increased effort of female STEM majors did not convert to gains in performance relative to males. One possible methodological explanation is that the effort variable captured time studying, but not other elements of effort that could relate to performance. For example, females may have estimated that more studying was needed, but were unsuccessful in estimating how to best study the material for the exam. Studying was defined to participants as engaging in at least one type of method of studying (e.g., highlighting/underlining, rereading, practice testing, self-explanation), but different methods of studying are more or less effective than others. For instance, practice testing is significantly more effective for long-term learning than highlighting, rereading, or underlining (Bjork et al., 2013). The effort items used in the current study did not distinguish what type of studying they did each day, each week, or in total. The items were set up this way because it would have greatly lengthened each daily diary to ask individuals dozens of questions regarding exact studying techniques and percentage of time engaging in each technique. Thus, it is possible that males studied less overall, but used more effective methods such as practice testing, than females, although this seems unlikely given the similar relationships between effort and performance for males and females and given that females completed marginally more homework problems than males in the small subsample of participants with data on objective homework. Nonetheless, investigating the question directly by asking about specific study strategies and length of time using each strategy to confirm there is no gender difference in study strategy is a good idea for a follow-up study. If males are not employing more effective study strategies than females, it would be useful to collect more information on how effective certain study strategies are for both genders. For instance, it may be the case that males and females are using a similar mix of study strategies (e.g., ~50% practice testing, ~20% rereading, ~10% underlining/highlighting, and ~10% mixture of the remaining strategies), but the effectiveness of this mix of strategies has diminishing returns after 10 hours of study time. If it were the case that the study strategy of the average STEM student is only positively related to performance for a certain range of study hours (e.g., 1-10 hours) but not outside of this range, it would be promising to investigate the mix of study strategies employed and the general effectiveness of this mix in the hopes of uncovering a better mix of study strategies to teach students. Results of these follow-up studies could inform teaching methodology of STEM professors and study strategies for STEM learners.

Another possible methodological limitation related to the null performance finding is that the effort measures did not capture other ways individuals might absorb the material besides studying. There is research that states that individuals who attend class and, more importantly, pay attention in class learn more (Credé, Roch, & Kieszcaynka, 2010; Wei, Wang, & Klausner, 2012) and perform better on assessments in class (Handelsman, Briggs, Sullivan, & Towler, 2010). Paying attention during class is a measure of effort (Wei et al., 2012). The effort measures in this study did not measure how much individuals absorbed lecture material in class or in other situations such as office hours. Although presumably difficult to measure (i.e., such measurement would likely involve behavioral measures coded by raters), future research might benefit from tackling this concept to determine if any gender differences exist and to determine if material absorption outside of, or even during, studying impacts criterion variables such as performance.

Besides possible differences in study strategy effectiveness or material absorption, it is also possible that there are gender differences in classroom networking. In other words, males may have had more access to study group networks and past exams than females because of the male-dominated culture of STEM areas. This idea is called the "Old Boys' Club" (Welde & Laursen, 2011). In STEM culture, males are more likely to study together and pass down materials from previous courses to males in younger generations. If this phenomena were happening during the current study, males might have studied less but not performed worse because they were part of the "Old Boys' Club" and had networking opportunities to obtain exam material that females did not.

Of course, the relationships among self-efficacy, effort, and performance are complicated. Many theorists contend that there is a spurious positive relationship between self-efficacy and performance because self-efficacy is caused by past performance and thus, it is ability that influences both self-efficacy and performance. These researchers recommend controlling for past performance in analyses relating self-efficacy and performance (Sitzmann & Yeo, 2013; Vancouver & Kendall, 2006). When controlling for past performance, the relationship between self-efficacy and performance becomes null or negative (e.g., Sitzmann & Yeo, 2013). However, other researchers argue that by controlling for past performance one is taking away variance from self-efficacy that should not be removed (i.e., overcontrolling; Bandura, 2012). One methodological way to address this question is to measure the variables over time within-person. This allows for a more powerful test of the self-efficacy-performance relationship by controlling for past performance within person (Vancouver et al., 2008). To fully examine these relationships in the current context, within-person methodologies would likely be useful. It is possible that self-efficacy's positive relationship with performance would reverse if examined within-person.

All of these possibilities for the null performance effects relate to one other possible explanation for why there was a gender difference in effort but not in performance. It is possible that females needed to study more than males to compensate for a slower rate of learning. In other words, it was knowledge of actual gender-related capacity differences that prompted extra effort, but females only compensated for this slower rate of learning enough to match the performance of males. As theorists who study self-efficacy note, if individuals have knowledge of their actual ability, it is this knowledge that drives motivation rather than belief in capacity. It is possible that as individuals studied, they became aware of their true rate of learning and adjusted studying appropriately. If females had a slower rate of learning than males, this actual capacity difference and subsequent effortful compensation would explain a difference in effort and would explain a null effect in performance. This possible explanation cannot be tested without having measured students' true rates of learning. Despite not being able to eliminate this possibility, the explanation is unlikely given that females had higher GPAs in high school and college and equivalent standardized test scores. Additionally, there is no evidence for such a gender difference in college according to past research on the topic (e.g., Crocker et al., 2003; MacPhee et al., 2013).

In sum, there were both strengths and weaknesses in the measurement of the predictor, mediator, and criterion variables. To draw conclusions about the relationships among these key measured variables, I also measured some variables that might explain the results for both the effort-related and performance-related hypotheses. That is, there were multiple measures of variables that were considered to be control variables, or alternative explanations, for the various hypotheses. For the effort hypotheses, these variables included value, outcome expectancies, and goal level. Each of these variables, and the implications of the results associated with each variable, are reviewed in turn.

Although it was predicted that females would study more for exams than males, value might undermine this relationship. In particular, females might have valued STEM less than males and such value differences may have negatively related to motivation. However, males and females did not differ on their report of value for STEM topics, or, in particular, for mathematics. As mean values mostly ranged between a 5 and a 6 on 7-point scales (with 7 indicating the most value) on all value scales, both males and females who are STEM majors likely have high value regarding STEM. Further, value was unrelated to effort or performance on the second exam. These results paint a picture in which self-efficacy has more of an impact on motivation during goal striving than value, because effort varied with self-efficacy levels, but not with value.

Beliefs in gender discrimination at work and belief in gender biases in exam grading were measured as possible control variables for the effort hypotheses as well. These particular control variables were measured as an alternative explanation for the hypothesized proposed gender on effort effects. Beliefs in gender biases might affect outcome expectancies (i.e., expectancies other than that of self-efficacy). Because beliefs in gender discrimination at work related to one of the predictors (i.e., gender), it was included as a control. However, the relationships between gender, self-efficacy, and effort remained the same even when controlling for such beliefs. This result provides evidence that beliefs regarding the need to put in more effort to achieve similar opportunities due to sexist workplace practices is not an alternative explanation, or an alternative mediator, to self-efficacy, at least when it comes to studying for exams in a school setting.

Additionally, there were some measures developed to capture beliefs in interpreting exam scores in a sexist way (e.g., artificially deflating females' exam scores and inflating males' when making decisions about good performance), because these beliefs seemed more relevant to college-aged participants. However, this variable did not relate to the predictor or criterion variables. Additionally, the items had poor internal consistency, perhaps because of a floor effect. That is, few endorse any items suggesting bias in exam scores. Of note, one of the questions had a complicated stem: "I feel that I need to work twice as hard compared to the majority of my classmates to get noticed by professors and employers, because of unconscious or conscious gender biases in evaluating grades." It would be useful to simplify this item for future research (e.g., "Professors and employers are biased against certain genders, so I believe I must work harder than my classmates to be noticed"). Even with the simplification, the item is still double-barreled, which often confuses participants (Morling, 2015). Efforts to remove the double-barreled nature of the item may improve internal consistency. One solution is to create two subscales: one subscale to measure perceived gender biases from professors and advisors and one subscale to measure working harder due to such perceived gender biases. Such subscales and the associated construct validity should be rigorously tested before use. Another solution might be to measure stigma consciousness, or how aware individuals' are of their status relative to others in society and any stigmas associated with that status (Brown & Pinel, 2003).

Desired goal level was also used as a control in the effort analyses in the current study due to its relationship with the predictor variables, but surprisingly, had very little effect on the effort mediation analysis. Additionally, the relationship between goal level and effort was negative for both genders. This result is inconsistent with predictions made in the goal-setting literature (Locke & Latham, 1990) and self-efficacy literature (Bandura, 2012), although not the first time a negative effect of goal has been found (i.e., lower goals corresponded with higher effort in other past studies; Vancouver et al., 2001). There are multiple possible explanations for this inconsistency. The goal setting literature predicts that when goals are more difficult, individuals exert more effort and when goals are lower, individuals exert less effort, given that all else is equal. In the current study, females had lower goals, but studied more for the examinations than males. Females might have studied more despite having lower goals, because their perceived ability to achieve the exam-related goals were lower (i.e., they believed that they were making slower progress on their goals and that the material was more difficult for them than males). This explanation seems likely. Females might have studied more despite having lower goals because they needed to study more to compensate for slower learning rates, although given that females and males had equivalent ability levels, this explanation seems unlikely. Last, the measurement might have been inadequate. To avoid obtaining low variance on goal level, which occurred in past studies (Vancouver & Kendall, 2006), I used a particular measure of goal level which asked for the lowest grade one would be satisfied with on the upcoming exam. Even though this measure yielded gender differences, the standard deviations were still small for both genders (i.e., one quarter of a grade level). Additionally, despite the argument for measuring goal in this manner (Mento, Locke, & Klein, 1992), the content validity for the measure is lacking. A goal is a representation of an individual's desired state, not a representation of the worst scenario an individual will accept. To adequately measure goal, one's true desired state should be measured, despite concerns about a ceiling effect. It might have been the case that when measuring goal as desired state, females would have reported a higher goal level or desired state. It would be useful in the future to investigate how much of the inconsistency is measurement-based and how much of the inconsistency is theory-based. In other words, researchers might investigate why manipulations of goal often yield results consistent with goal setting theory (Locke, Shaw, Saari, & Latham, 1981), but when goal is measured, inconsistencies with theory are found (Vancouver et al., 2001).

To summarize the effects of the control variables on the effort hypotheses, the results of the control variables strengthen the conclusions of this study regarding gender, self-efficacy, and effort. Although there were some limitations of the measurements, the way the majority of the controls were measured had theoretical and methodological support. The mediating effect of self-efficacy in explaining the relationship between gender and effort remained even after controlling for third variable explanations.

There were also three variables measured either to reduce threats to the conclusions drawn from the performance hypotheses or to explain possible null effects. These three variables were KSAs (i.e., knowledge, skills, and abilities), stereotype threat, and in conjunction with stereotype threat, math/science identification. These variables are reviewed in turn.

There was a possibility that there may be a gender difference in KSAs, which could explain performance differences between the genders. One possible source of gender differences could arise because females have a higher threshold for entering STEM than males, or at the very least are less likely to believe they meet the self-chosen threshold to enter STEM majors or STEM fields. This could lead to females having higher KSAs than males. KSAs were measured via a variety of standardized test scores. However, females did not perform better on either exam nor did they demonstrate higher KSAs via the standardized tests. These results substantiate most of the past research that found no gender differences in such scores. One exception to this typical finding was that females did have a significantly higher GPA in both high school and college than males, which corroborated one study that found a similar but much smaller difference in GPA (i.e., .089 points; Sonnert & Fox, 2012) than the current study found (i.e., .18-.30 points depending on if the GPA reported was high school or college). There are multiple implications of the gender difference in GPA in contrast to the lack of a gender difference in KSAs. First, it is possible that females perform better than males at non-exam related assignments (e.g., group projects, papers, or homework – the objective homework measure in the current study provides some evidence of this possibility), which would account for the GPA difference when there were no test score differences. Second, there may be something distinctive happening during the test-taking process that is not happening during other performance assessment processes.

A well-studied factor that is particularly important during examinations and standardized test-taking situations is stereotype threat. Because stereotype threat was measured in this dissertation, the results have some methodological implications. Belief in stereotype threat was measured to assess if it might relate to performance and thus reduce females' performance scores. Indeed, there was a gender difference in stereotype threat endorsement. Females endorsed experiencing more stereotype threat than males. Moreover, stereotype threat beliefs were also negatively related to performance on the second exam as well as some of the ability measures (i.e., standardized test scores). Although stereotype threat was not negatively related to scores on the first exam, females performed either marginally (i.e., with influential cases included) or significantly (i.e., without influential cases included) better on the first exam when taking into account feelings of stereotype threat. After removing influential cases in the regression predicting the second exam score, there was a significant negative relationship between stereotype threat experience and exam performance for females. These results credit the idea that when females were more fearful of endorsing stereotypes involving females and mathematics abilities in a performance situation, females underperformed relative to males (Steele & Aronson, 1995). In other words, it is possible that stereotype threat eliminated the performance enhancements that females in STEM would otherwise have had due to their increased efforts.

There are two limitations of the stereotype threat results in the current study. The first is that some of these effects were not found when influential cases were included in the analyses. There is much debate about when to include and when not to include influential cases in research. It seemed prudent to report results both with and without high-impact outliers and allow readers to draw conclusions. Given the results do not tell the same story when influential cases are or are not included, the effect of stereotype threat in this particular context remains questionable. The second limitation involves the scope of the results. Even though females endorsed feeling more stereotype threat than males, and stereotype threat endorsement predicted exam score, it is not possible to determine a) how many female individuals were affected by stereotype threat and b) how much exam or standardized test scores suffered from stereotype threat (i.e., what the performance enhancements of females in STEM would have really been without the stereotype threat present). Methodologically, it is hard to tease out the true effect of stereotype threat when there is no participant data without the stereotype threat present and when there is more noise due to measurement rather than manipulation. A recent meta-analysis involving stereotype threat for females on 47 different math, science, and spatial skills tests provided an effect size of -0.22, which means performance was affected negatively when stereotype threat was present (Flore & Wicherts, 2015). However, even with the reduction in noise and a number that can be directly attributable to the effect of stereotype threat, it is still a struggle to interpret what the arguably weak effect size really means (Cohen, 1992). This type of practical implication of stereotype threat is discussed along with other implications in the following practical implications section.

Practical Implications

There are two primary practical implications of the current study. The first implication is related to self-efficacy interventions for STEM recruitment. This section of the discussion highlights the need to consider the results of the current study when deciding how to structure future STEM interventions that include self-efficacy components. The second implication is related to stereotype threat. Specifically, this discussion focuses on how the stereotype threat results might relate to school and work-related outcomes.

Intervention efforts are often focused on recruiting and retaining underrepresented groups in STEM majors and careers (Stout et al., 2011). Due to research on goal choice and self-efficacy, these interventions often include a self-efficacy component (e.g., “You can make it in STEM!”) (Betz, 2004; Rincón & George-Jackson, 2016; Rittmayer & Beier, 2009) or, at the very least, a self-efficacy component is recommended for STEM interventions (Betz & Schifano, 2000; Vogt, 2008; Zeldin & Pajares, 2000). When individuals have artificially low self-efficacy compared to their ability, and this artificially low self-efficacy prevents them from attempting a STEM goal (e.g., taking a mathematics course, declaring a physics major, applying for industry engineering jobs), a self-efficacy component in an intervention could successfully increase STEM recruitment. However, as reviewed extensively in this paper, although low self-efficacy negatively impacts goal choice, high self-efficacy negatively impacts effort applied to goal striving. By including self-efficacy components in interventions, it is possible that raising self-efficacy artificially could also have deleterious effects on individuals. In particular, one might raise an individual's self-efficacy to levels that substantially reduce the effort they apply during goal striving. Thus, there is a dilemma in regards to the effectiveness of self-efficacy components in STEM interventions.

One solution to the above dilemma is to focus not on the level of self-efficacy, but rather on calibrating individuals to their true capacity. Calibration is the degree to which an individual can accurately align their beliefs about capacity with their actual capacity such that they understand the amount of effort necessary to achieve a certain goal given their current capacity (Chen, 2002; Linnenbrink & Pintrich, 2003). Miscalibration can be a serious problem because it can lead individuals to underestimate their learning needs and allocate resources poorly (Halper, Hall, Vancouver, & Purl, 2016; Vancouver, Halper, & Bayes, 2017). Often, self-efficacy components in interventions to recruit or retain individuals in STEM provide blanket self-efficacy-raising sentiments without taking into account the importance of calibration. Thus, the interventions may instead have the unintended effect of miscalibrating individuals by artificially raising their self-efficacy. Individuals may be misled to believe that they can accomplish a difficult goal without much effort allocated on their part. This issue may be solved with important additional clauses (e.g., “You can make it in STEM even though you do not believe you can *if you work hard*”). Although these phrases are general, their goal is to inform the individual about the effort necessary for a career choice such as engineering. The effect of self-efficacy components in recruitment or retention interventions warrants further research to determine if focusing on calibration may improve effort expenditure for either gender.

Importantly, these calibration-included phrases would be useful to impress upon anyone interested in STEM (i.e., both males and females). Recalibration interventions should not just be focused on females. Past studies have found that there are certain areas where males are overconfident (i.e., miscalibrated such that they overestimate their ability). These areas include cognitive tests (Stankov & Lee, 2008), reading and listening (Stankov, Lee, & Paek, 2009), and mathematical addition problems (Mertins & Hoffeld, 2015). In the current study, it is possible that females are under confident, but it is also possible that males are overconfident. Females and males scored similarly on the ability composite, demonstrated similar exam performance, but males had higher self-efficacy than females in STEM. In calibration/self-efficacy interventions, interventionists should target males and females and emphasize to everyone that effort needs to be applied to succeed in a STEM career. These interventions should also mention that the hard work put in will be worthwhile; not only do STEM careers pay well (Beede et al., 2011), but because few individuals put in the work to graduate with a STEM degree, there is a high demand for STEM graduates in the workforce (Hira, 2010).

Another practical implication involves stereotype threat and STEM. Stereotype threat may be practically important at school for STEM majors, given the results of this study and given the results of many other studies examining females, mathematics ability beliefs, and test performance. If it is the case that females face performance decrements in STEM due to fear of endorsing stereotypes, females in STEM face even more hurdles in college than previously realized. Future studies should aim to investigate the true level of performance decrements in STEM majors attributable to stereotype threat, which includes a thorough analysis of what an effect size of -0.22 means practically for females. If there are substantial decrements (i.e., or the enhancements that would have otherwise been there are substantial), there may be sufficient reason to implement stereotype threat interventions among females STEM majors. The stereotype threat interventions might emphasize statistics from studies cited in this paper along with the results of this paper, which demonstrate that when looking at STEM majors, females' ability is no different than that of males or possibly slightly above that of males. The intervention might also provide statistics that the motivation levels of females are higher than that of males. This type of intervention is called a reconstrual intervention (Spencer, Logel, & Davies, 2016). If female STEM majors learn that their capacity for STEM-related goals is equivalent to males, they may be less afraid of confirming the gender stereotype. If females are less afraid of confirming a stereotype, stereotype threat will affect their performance on exams less.

Beyond school, stereotype threat may be practically important at work, but a review of past and recent research does not provide a clear answer. Past and recent stereotype threat research has mostly focused on how tests in college are negatively impacted by fear of confirming stereotypes (e.g., Boucher, Rydell, & Murphy, 2015; Drazkowski et al., 2015; John-Henderson, Rheinschimidt, & Mendoza-Denton, 2014; Spencer et al., 2016; Tempel & Neumann, 2016). Fewer studies have examined real-life work situations and stereotype threat (Block, Koch, Liberman, Merriweather, & Roberson, 2011) and even fewer have examined employed females in STEM and stereotype threat at work. Of those articles that have examined the effect of stereotype threat on work performance, individuals who experienced some form of stereotype threat were more likely to disengage from the task (Kulik, Perera, Cregan, 2016) and more likely to indirectly seek feedback at work (e.g., pay attention to unsolicited feedback cues from supervisors and peers; Roberson, Deitch, Brief, & Block, 2003). Results from the above studies have contradictory implications for performance. Disengagement is negatively related to performance (Christian, Garza, & Slaughter, 2011), and feedback-seeking is positively related to performance (Ashford & Tsui, 1991; Seijts, Taylor, & Latham, 1998). It may be possible that individuals under stereotype threat behave more passively at work, which can be either beneficial or harmful, depending on one's tasks at work.

One exception to the ignorance of the stereotype threat literature of workplace implications was a study measuring training performance in a college sample. After a multi-day self-guided training program, females functioning under stereotype threat performed worse on a declarative knowledge test than females in the control condition (Grand, 2017). One can draw the conclusion that stereotype threat may impact transfer-of-training for females in work situations. However, the performance measure used was a declarative knowledge test, similar to a college exam. Also, participants were college students, and not organizational employees. Importantly, investigations into if stereotype threat may impact STEM employees' day-to-day job performance (i.e., team projects, client interactions, data analyses) is lacking. Before drawing conclusions regarding if stereotype threat may interfere with day-to-day work performance, and thus should or should not be a factor in hiring, more research needs to be done. Researchers could examine project performance of STEM majors in college or project performance of STEM employees after being exposed to stereotype threat to begin to answer the question. Indeed, if stereotype threat is unique to proctored testing situations, the results of this study speak to an advantage of hiring females for STEM jobs due to their motivation levels.

Conclusion

The current study contributed to the field's big-picture understanding of STEM education. There is considerable research concerning motivational elements in goal choice processes. Overall, researchers understand factors that keep individuals from beginning their STEM education and thus keep individuals from STEM careers. There is much less research concerning how similar motivational elements function in individuals pursuing their STEM education and later, STEM careers. Control theory and the empirical non-monotonic model of self-efficacy were used to hypothesize how males and females study for their STEM-related course goals. Females in STEM were found to be more motivated to study for STEM-related exams than males in STEM. Self-efficacy was found to be a mediator in the current study insomuch as females had lower self-efficacy than males for STEM-related classroom goals and this low self-efficacy was related to effort applied. Even though females put in more time studying for exams and there were no substantive ability differences between females and males, performance did not differ between the genders. It is possible that the performance result could be explained by stereotype threat, although more research is needed to make this claim and any practical implications of the claim. It is also possible that there are diminishing returns of the effect of studying on performance. Additionally, future research should examine the performance hypotheses within person over a longer period of time to increase the power to detect an effect. The conclusions from this study have implications for theories of self-efficacy and theories of academic self-concepts. There are also implications for interventions that call on self-efficacy to manipulate goal choice for STEM. Improved interventions for recruitment and retention in STEM might focus on promoting accurate beliefs to increase the number of successful future STEM majors. In all, this study points a way forward that involves calibrating individuals to their capacity, so that they make good choices in the future, work harder on what is difficult to learn, and do not worry about what their performance may say about their group.

References

Almeida, D. M. (2005). Resilience and vulnerability to daily stressors assessed via diary methods. *Current Directions in Psychological Science, 14*, 64-68.

Arvey, R. D., & Renz, G. L. (1992). Fairness in the selection of employees. *Journal of Business Ethics*, *11*, 331-340.

Ashford, S. J., & Tsui, A. S. (1991). Self-regulation for managerial effectiveness: The role of active feedback seeking. *Academy of Management Journal, 34*, 251-280.

Austin, J. T., & Vancouver, J. B. (1996). Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin*, *120*, 338-375.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191–215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Inc.

Bandura, A. (1989). Social cognitive theory. In R. Vasta (Ed.), Annals of child development. Vol.6. *Six theories of child development* (pp. 1-60). Greenwich, CT: JAI Press.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*, 248-287.

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71-81). New York: Academic Press. (Reprinted in H. Friedman [Ed.], *Encyclopedia of mental health*. San Diego: Academic Press, 1998).

Bandura, A. (1997). *Self-efficacy: the exercise of control.* New York: Freeman.

Bandura, A. (1999). Social cognitive theory: An agentic perspective. *Asian Journal of Social Psychology, 2*, 21-41.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*, 1-26.

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*, 87-99.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, *41*, 586-598.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences, 6*, 248-254.

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief,* (04-11).

Bell, M. P., McLaughlin, M. E., & Sequeira, J. M. (2002). Discrimination, harassment, and the glass ceiling: Women executives as change agents. *Journal of Business Ethics, 37*, 65-76.

Besterfield-Sacre, M., Moreno, M., Shuman, L. J., & Atman, C. J. (2001). Gender and ethnicity differences in freshmen engineering student attitudes: A cross-institutional study. *Journal of Engineering Education, 90*, 477-489.

Betz, N. E., & Schifano, R. S. (2000). Evaluation of an intervention to increase Realistic self-efficacy and interests in college women. *Journal of Vocational Behavior, 56,* 35–52.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444.

Block, C. J., Koch, S. M., Liberman, B. E., Merriweather, T. J., & Roberson, L. (2011). Contending with stereotype threat at work: A model of long-term responses. *The Counseling Psychologist, 39*, 570-600.

Boucher, K. L., Rydell, R. J., & Murphy, M. C. (2015). Forecasting the experience of stereotype threat for others. *Journal of Experimental Social Psychology, 58*, 56-62.

Brainard, S. G., & Carlin, L. (1997, November). A longitudinal study of undergraduate women in engineering and science. In *Frontiers in Education Conference, 1997. 27th Annual Conference. Teaching and Learning in an Era of Change. Proceedings*. (Vol. 1, pp. 134-143). IEEE.

Brainard, S. G., & Carlin, L. (1998). A six-year longitudinal study of undergraduate women in engineering and science*. *Journal of Engineering Education, 87*, 369-375.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology, 39*, 626-633.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, *16*, 572-578.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M.D. Dunnette & L.M. Hough (Eds.) *Handbook of industrial and organizational psychology* (Vol. 1, 2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

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

Cech, E. (2015). Engineers and engineeresses? Self-conceptions and the development of gendered professional identities. *Sociological Perspectives*, *58*, 56-77.

Chan, W. S. (2016). Delay discounting and response disinhibition moderate associations between actigraphically measured sleep parameters and body mass index. *Journal of Sleep Research, 26,* 21-29.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*, 181-197.

Cheng, W., & Ickes, W. (2009). Conscientiousness and self-motivation as mutually compensatory predictors of university-level GPA. *Personality and Individual Differences, 47*, 817-822.

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

Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM?. *Social Psychological and Personality Science, 2*, 656-664.

Christian, M. S., Garza, A. S., & Slaughter, J. E. (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology*, *64*, 89-136.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85,* 678–707.

Compeau, D., Higgins, C. A., & Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly*, 145-158.

Cooper Jackson, J. (2001). Women middle managers' perception of the glass ceiling. *Women in Management Review*, *16*, 30-41.

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

Cortina, J.M., & Luchman, J.N. (2013). Personnel selection and employee performance. In N.W. Schmitt & S. Highhouse (Eds.) *Handbook of psychology volume 12: Industrial and organizational psychology.* Hoboken, NJ: John Wiley & Sons, Inc.

Coyle, T. R., Purcell, J. M., Snyder, A. C., & Richmond, M. C. (2014). Ability tilt on the SAT and ACT predicts specific abilities and college majors. *Intelligence*, *46*, 18-24.

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

Crocker, J., Karpinski, A., Quinn, D. M., & Chase, S. K. (2003). When grades determine self-worth: Consequences of contingent self-worth for male and female engineering and psychology majors. *Journal of Personality and Social Psychology*, *85*, 507-516.

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

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, *28*, 1615-1628.

Dik, B. J., Eldridge, B. M., Steger, M. F., & Duffy, R. D. (2012). Development and validation of the calling and vocation questionnaire (CVQ) and brief calling scale (BCS). *Journal of Career Assessment*, *20*, 242-263.

Diefendorff, J. M., & Chandler, M. M. (2011). Motivating employees. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, volume 3: Maintaining, expanding, and contracting the organization* (pp. 65-135). Washington D.C., US: American Psychological Association.

Drążkowski, D., Szwedo, J., Krajczewska, A., Adamczuk, A., Piątkowski, K., Jadwiżyc, M., & Rakowski, A. (2015). Women are not less field independent than men—the role of stereotype threat. *International Journal of Psychology*, 1-5.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109-132.

Edwards, H. (2015, March 13). Average AP scores for every AP exam. Retrieved from http://blog.prepscholar.com/average-ap-scores-for-every-ap-exam. Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5-17.

Fisher, S. L., & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51*, 397-420.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*, 25-44.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*, 519-533.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233-239.

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development*, *85*, 250-263.

Gartland, N., O'Connor, D. B., Lawton, R., & Ferguson, E. (2014). Investigating the effects of conscientiousness on daily stress, affect and physical symptom processes: A daily diary study. *British Journal of Health Psychology, 19*, 311-328.

Grand, J. A. (2016). Brain drain? An examination of stereotype threat effects during training on knowledge acquisition and organizational effectiveness. *Journal of Applied Psychology, 102*, 115-150.

Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity*, *3*, 53-68.

Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions* (2nd ed.). Routledge.

Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beaubien, J. M. (2002). A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships. *Journal of Applied Psychology, 87*, 819-832.

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, *51*, 1163-1176.

Hackett, G., & Betz, N. E. (1995). Self-efficacy and career choice and development. In J.E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment* (pp. 249-280). New York: Plenum Press.

Hackett, G., Betz, N. E., Casas, J. M., & Rocha-Singh, I. A. (1992). Gender, ethnicity, and social cognitive factors predicting the academic achievement of students in engineering. *Journal of Counseling Psychology, 39*, 527-538.

Halper, L.R., Hall, K.E., Vancouver, J.B., & Purl, J. (2016, May). *Learning self-efficacy is positively and negatively related to effort allocated to learning.* Poster session presented at the 28th annual meeting of the Association for Psychological Science (APS) Convention, Chicago, IL.

Halper, L.R., & Vancouver, J.B. (2016). Self-efficacy's influence on persistence on a physical task: Moderating effect of performance feedback ambiguity. *Psychology of Sport & Exercise, 22,* 170-177.

Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *The Journal of Educational Research, 98*, 184-192.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*, 408-420.

Heilbronner, N. N. (2013). The STEM pathway for women: What has changed?. *Gifted Child Quarterly*, *57*, 39-55.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior, 32*, 113-135.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology, 89*, 416.

Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology, 26*, 276-295.

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in Science, Technology, Engineering, and Mathematics*. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.

Hira, R. (2010). US Policy and the STEM workforce system. *American Behavioral Scientist*, *53*, 949-961.

Hutchison, M. A., Follman, D. K., Sumpter, M., & Bodner, G. M. (2006). Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education, 95*, 39-47.

John-Henderson, N. A., Rheinschmidt, M. L., Mendoza-Denton, R., & Francis, D. D. (2014). Performance and inflammation outcomes are predicted by different facets of SES under stereotype threat. *Social Psychological and Personality Science, 5*, 301-309.

Johnson, D. L., Joyce, B. P., & Sen, S. (2002). An analysis of student effort and performance in the finance principles course. *Journal of Applied Finance, 12*, 67-72.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*, 657-690.

Kay, K., & Shipman, C. (2014). *The confidence code: The science and art of self-assurance---what women should know.* New York, NY: Harper Collins.

Kulik, C. T., Perera, S., & Cregan, C. (2016). Engage me: The mature-age worker and stereotype threat. *Academy of Management Journal, 59*, 2132-2156.

Kuncel, N. R., Credè, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*, 63-82.

Leach, D. J., Wall, T. D., Rogelberg, S. G., & Jackson, P. R. (2005). Team autonomy, performance, and member job strain: Uncovering the teamwork KSA link. *Applied Psychology*, *54*, 1-24.

Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of Counseling Psychology*, *33*, 265-269.

Lesko, A. C., & Corpus, J. H. (2006). Discounting the difficult: How high math-identified women respond to stereotype threat. *Sex Roles*, *54*, 113-125.

Locke, E. A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task performance. *Journal of Applied Psychology*, *69*, 241-251.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Englewoody Cliffs, NJ: Prentice-Hall, Inc.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin*, *90*, 125-152.

MacPhee, D., Farro, S., & Canetto, S. S. (2013). Academic self-efficacy and performance of underrepresented STEM majors: Gender, ethnic, and social class patterns. *Analyses of Social Issues and Public Policy, 13*, 347-369.

Magee, J. C., & Galinsky, A. D. (2008). 8 social hierarchy: The self-reinforcing nature of power and status. *The Academy of Management Annals, 2*, 351-398.

Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology*, *108*, 424-437.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-388.

Marsh, H. W. (1991). Multidimensional students' evaluations of teaching effective-ness: A test of alternative higher-order structure. *Journal of Educational Psychology, 83*, 285-296.

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80*, 366-380.

Marsh, H. W., Kuyper, H., Morin, A. J., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction, 33*, 50-66.

Marsh, H. W., & O'Neill, R. (1984). Self-Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement, 21*, 153-174.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20*, 319-350.

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

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30-46.

Meece, J. L., Parsons, J. E., Kaczala, C. M., & Goff, S. B. (1982). Sex differences in math achievement: Toward a model of academic choice. *Psychological Bulletin, 91*, 324-348.

Mento, A. J., Locke, E. A., & Klein, H. J. (1992). Relationship of goal level to valence and instrumentality. *Journal of Applied Psychology*, *77*, 395-405.

Mertins, V., & Hoffeld, W. (2015). Do overconfident workers cooperate less? The relationship between overconfidence and cooperation in team production. *Managerial and Decision Economics, 36*, 265-274.

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and Individual Differences*, *26*, 141-152.

Moritz, S. E., Feltz, D. L., Fahrbach, K. R., & Mack, D. E. (2000). The relation of self-efficacy measures to sport performance: A meta-analytic review. *Research Quarterly for Exercise and Sport, 71*, 280-294.

Morling, B. (2015). *Research methods in psychology: Evaluating a world of information*. W.W. Norton: New York.

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314-1334.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, *9*, 1-12.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research.* Jossey-Bass: San Francisco.

Post-Kammer, P., & Smith, P. L. (1986). Sex differences in math and science career self-efficacy among disadvantaged students. *Journal of Vocational Behavior*, *29*, 89-101.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*, 717-731.

Ragins, B. R., Townsend, B., & Mattis, M. (1998). Gender gap in the executive suite: CEOs and female executives report on breaking the glass ceiling. *The Academy of Management Executive, 12*, 28-42.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*, 353-387.

Rincón, B. E., & George-Jackson, C. E. (2016). Examining department climate for women in engineering: The role of STEM interventions. *Journal of College Student Development*, *57*, 742-747.

Rios, K., Cheng, Z. H., Totton, R. R., & Shariff, A. F. (2015). Negative stereotypes cause Christians to underperform in and disidentify with science. *Social Psychological and Personality Science*, 1-9.

Rittmayer, M.A. & Beier, M.E. (2009). Self-Efficacy in STEM. In B. Bogue & E. Cady (Eds.). *Applying Research to Practice (ARP) Resources*. Retrieved <Mar 3, 2017> from http://www.engr.psu.edu/AWE/ARPresources.aspx.

Roberson, L., Deitch, E. A., Brief, A. P., & Block, C. J. (2003). Stereotype threat and feedback seeking in the workplace. *Journal of Vocational Behavior, 62*, 176-188.

Sadri, G., & Robertson, I. T. (1993). Self-efficacy and work-related behaviour: A review and meta-analysis. *Applied Psychology, 42*, 139-152.

Samson, G. E., Graue, M. E., Weinstein, T., & Walberg, H. J. (1984). Academic and occupational performance: A quantitative synthesis. *American Educational Research Journal, 21*, 311-321.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440-452.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, *71*, 432-439.

Schultz, D. P., & Schultz, S. E. (2010). *Psychology and work today: An introduction to industrial and organizational psychology*. Prentice Hall.

Schunk, D. H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, *25*, 71-86.

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26,* 207-231.

Seijts, G., Taylor, L., & Latham, G. (1998). Enhancing teaching performance through goal setting, implementation and seeking feedback. *The International Journal for Academic Development, 3*, 156-168.

Seymour, E., & Hewitt, N. (1997). *Talking about leaving.* Boulder, CO: Westview Press.

Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance?. *Personnel Psychology, 66*, 531-568.

Smith, P., Crittenden, N., & Caputi, P. (2012). Measuring women's beliefs about glass ceilings: Development of the Career Pathways Survey. *Gender in Management: An International Journal, 27*, 68-80.

Smith, P., Caputi, P., & Crittenden, N. (2012). How are women's glass ceiling beliefs related to career success?. *Career Development International, 17*, 458-474.

Sonnert, G., & Fox, M. F. (2012). Women, men, and academic performance in science and engineering: The gender difference in undergraduate grade point averages. *The Journal of Higher Education, 83*, 73-101.

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology, 67*, 415-437.

Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology, 100*, 961-976.

Stankov, L., Lee, J., & Paek, I. (2009). Realism of confidence judgments. *European Journal of Psychological Assessment*, *25*, 123-130.

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

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34*, 379-440.

Steinberg, J. R., Okun, M. A., & Aiken, L. S. (2012). Calculus GPA and math identification as moderators of stereotype threat in highly persistent women. *Basic and Applied Social Psychology*, *34*, 534-543.

Taraban, R., Maki, W. S., & Rynearson, K. (1999). Measuring study time distributions: Implications for designing computer-based courses. *Behavior Research Methods, Instruments, & Computers*, *31*, 263–269.

Tempel, T., & Neumann, R. (2016). Taming test anxiety: The activation of failure-related concepts enhances cognitive test performance of test-anxious students. *The Journal of Experimental Education, 84*, 702-722.

Vancouver, J. B. (2008). Integrating self-regulation theories of work motivation into a dynamic process theory. *Human Resource Management Review, 18*, 1-18.

Vancouver, J. B. (2012). Rhetorical reckoning: A response to Bandura. *Journal of Management, 38*, 465-474.

Vancouver, J.B., Alicke, M., & Halper, L.R. (in press). Self-efficacy. In D.L. Ferris, R. Johnson, & C. Sedikides (Eds.) *SIOP Organizational Frontiers: The Self at Work*, New York, NY: Routledge.

Vancouver, J. B., Gullekson, N. L., Morse, B. J., & Warren, M. A. (2014). Finding a between-person negative effect of self-efficacy on performance: Not just a within-person effect anymore. *Human Performance*, *27*, 243-261.

Vancouver, J.B., Halper, L.R., & Bayes, K.A. (2017). Regulating our own learning: Stuff you did not realize you needed to know. In R. Noe & J. Ellingson (Eds.) *Autonomous Learning in the Workplace*, Florence, KY: Taylor & Francis.

Vancouver, J. B., & Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology*, *91*, 1146-1153.

Vancouver, J. B., More, K. M., & Yoder, R. J. (2008). Self-efficacy and resource allocation: Support for a nonmonotonic, discontinuous model. *Journal of Applied Psychology, 93*, 35-47.

Vancouver, J. B., & Purl, J. D. (2017). A computational model of self-efficacy's various effects on performance: Moving the debate forward. *Journal of Applied Psychology, 102,* 599-616.

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87*, 506-516.

Vancouver, J. B., Thompson, C. M., & Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology*, *86*, 605-620.

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

Vroom, V. H. (1964). Work and motivation. San Francisco, CA: Jossey-Bass.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817-835.

Wang, X. (2013). Modeling entrance into STEM fields of study among students beginning at community colleges and four-year institutions. *Research in Higher Education, 54*, 664-692.

Wang, M. T., Degol, J., & Ye, F. (2015). Math achievement is important, but task values are critical, too: Examining the intellectual and motivational factors leading to gender disparities in STEM careers. *Frontiers in Psychology*, *6,* 1-9.

Wei, F. Y. F., Wang, Y. K., & Klausner, M. (2012). Rethinking college students' self-regulation and sustained attention: Does text messaging during class influence cognitive learning?. *Communication Education, 61*, 185-204.

Wrigley, B. J. (2002). Glass ceiling? What glass ceiling? A qualitative study of how women view the glass ceiling in public relations and communications management. *Journal of Public Relations Research, 14*, 27-55.

Zeldin, A. L., & Pajares, F. (2000). Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal, 37*, 215–246.

Zhang, F. (2016, March 6). Average SAT and ACT scores by state (participation adjusted). Retrieved from http://blog.prepscholar.com/average-sat-and-act-scores-by-stated-adjusted-for-participation-rate.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, *29*, 663-676.

Appendix A: Measures

Predictor variables.

Self-efficacy.

How many hours do you believe you need to study to obtain <u>at least</u> a D on the next exam in this class? __________________

How many hours do you believe you need to study to obtain <u>at least</u> a C on the next exam in this class? __________________

How many hours do you believe you need to study to obtain <u>at least</u> a C+ on the next exam in this class? __________________

How many hours do you believe you need to study to obtain <u>at least</u> a B- on the next exam in this class? __________________

How many hours do you believe you need to study to obtain <u>at least</u> a B on the next exam in this class? __________________

How many hours do you believe you need to study to obtain <u>at least</u> a B+ on the next exam in this class? __________________

How many hours do you believe you need to study to obtain <u>at least</u> an A- on the next exam in this class? __________________

How many hours do you believe you need to study to obtain <u>at least</u> an A on the next exam in this class? __________________

What grade would you expect to get on the <u>next</u> exam if you studied for <u>1 hour</u>?
D
C
C+
B-
B
B+
A-
A

What grade would you expect to get on the <u>next</u> exam if you studied for <u>5 hours</u>?
D
C
C+
B-
B
B+
A-
A

What grade would you expect to get on the <u>next</u> exam if you studied for <u>10 hours</u>?
D
C
C+
B-
B
B+
A-
A

What grade would you expect to get on the <u>next</u> exam if you studied for <u>15 hours</u>?
D
C
C+
B-
B
B+
A-
A

What grade would you expect to get on the <u>next</u> exam if you studied for <u>20 hours</u>?
D
C
C+
B-
B
B+
A-
A

Compared to other students in this class, how difficult is this material for you?

1	2	3	4	5	6	7
Not at all difficult						Very difficult

Criterion variables.
Effort.

Please answer each of the following questions **only** in relation to the mathematics class in which you are enrolled that is connected to this research project.

Please indicate the length of time during which you studied today for your upcoming mathematics exam.
Please indicate for what length of time you employed any of the following studying techniques today in studying for your upcoming mathematics exam. The definition of each technique is next to each label.

<u>Elaborative interrogation</u>: Generating an explanation for why an explicitly stated fact or concept is true.

<u>Self-explanation</u>: Explaining how new information is related to known information/explaining steps taken during problem solving.

<u>Summarization</u>: Writing summaries of to-be-learned texts.

<u>Highlighting/underlining</u>: Marking potentially important portions of to-be-learned materials.

<u>Keyword mnemonic</u>: Using keywords and mental imagery to associate verbal materials.

<u>Imagery for text</u>: Attempting to form mental images of text materials while reading or listening.

<u>Rereading</u>: Restudying text material again after an initial reading.

<u>Practice testing</u>: Self-testing or taking practice tests over to-be-learned material. This would include working on practice problems in homework or beyond.
0 minutes
30 minutes
1 hour
1 hour and 30 minutes
2 hours
2 hours and 30 minutes
3 hours
3 hours and 30 minutes
4 hours
4 hours and 30 minutes or more

Please indicate the length of time you have studied total so far this week since this past Sunday?
(NOTE: If today is a Sunday, please record how much time total you have studied since last Sunday).
I have not studied at all this week
30 minutes
1 hour
1 hour and 30 minutes
2 hours
2 hours and 30 minutes
3 hours
3 hours and 30 minutes
4 hours
4 hours and 30 minutes
5 hours
5 hours and 30 minutes
6 hours
6 hours and 30 minutes
7 hours
7 hours and 30 minutes
8 hours
8 hours and 30 minutes
9 hours
9 hours and 30 minutes
10 hours
10 hours and 30 minutes
11 hours
11 hours and 30 minutes
12 hours or more

How many hours total did you study for the <u>second</u> exam of the semester in this course? Please record the number of hours.

Control/demographic variables.

Demographics.

What gender do you most identify with? (pick which applies)
Male
Female

Are you of Hispanic, Latino, or Spanish origin?
Yes
No

What is your ethnic/racial background? (pick all that apply)
American Indian or Alaskan Native
Asian American
Pacific Islander
Black/African American
White/European American
Other
Decline to answer

Are you an international student?
Yes
  If so, what is your continent of origin?
    North America
    South America
    Africa
    Europe
    Asia
    Australia
    Other
    Decline to answer
No

What is your age in years? ___________

What is your year in school? (pick which applies)
High school
First year
Second year
Third year
Fourth year
Fifth year or beyond

Are you a first generation college student (i.e., are you the first of your immediate family to attend a four year degree program in college)?
Yes
No

College and major.

What is your major?
Aviation – Engineering
Chemical and Biomolecular Engineering
Civil Engineering
Computer Science Engineering
Electrical Engineering
Energy Engineering
Engineering Technology and Management
Industrial and Systems Engineering
Mechanical Engineering
Undecided Engineering
Actuarial Sciences
Mathematics
Applied Mathematics
Mathematical Statistics
Mathematics – Meteorology
Physics
Applied Physics
Astrophysics
Physics – Meteorology
Other: fill in

Which is your college? Be specific (e.g., Russ College of Engineering, College of Arts & Sciences). If you are in the Honors Tutorial College, please select "Honors Tutorial College" in answer to this question.
Russ College of Engineering
College of Arts & Sciences
Honors Tutorial College
Other: fill in ____

Hours registered.

How many credit hours are you enrolled in this semester?
____

Extracurricular organizations and job-related activities.

Please record the number of extracurricular organizations of which you are a part:

____

Please estimate the number of hours a week you spend participating in all of the organizations listed above combined (e.g., if you spend 2 hours a week participating in one organization and 4 hours a week participating in another, please write 6):

____

Please estimate the number of hours a week you spend working at a part-time or full-time job. This estimate should include hours a week you work in a research lab, even if you are getting college credit for the research lab:

____

***Gender discrimination beliefs*.**

Exam-Focused Beliefs in Bias

I feel that I need to work twice as hard compared to the majority of my classmates to get noticed by professors and employers, because of unconscious or conscious gender biases in evaluating grades.

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

I believe exams are graded objectively.

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

The grade I obtain on an exam in this course reflects how well I actually did on the exam and does not reflect a bias in grading.

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

General Workplace Beliefs in Bias (split up into four subscales; item number followed by item)

Denial

30 Women starting careers today will face sexist barriers.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

9 Women and men have to overcome the same problems at the workplace.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

39 It will take decades for women to reach equality with men in high level management positions.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

10 Even women with many skills and qualifications fail to be recognized for promotions.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

13 Women have reached the top in all areas of business and politics. REVERSE CODED

1 2 3 4 5 6 7
Strongly disagree Strongly agree

1 Women face no barriers to promotions in most organizations. REVERSE CODED

1 2 3 4 5 6 7
Strongly disagree Strongly agree

11 Women leaders are seldom given full credit for their successes.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

15 Women in senior positions face frequent putdowns of being too soft or too hard.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

7 Women who have a strong commitment to their careers can go right to the top. REVERSE CODED

1 2 3 4 5 6 7
Strongly disagree Strongly agree

4 Talented women are able to overcome sexist discrimination. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

Resignation

36 Women executives are very uncomfortable when they have to criticize members of their teams.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

26 Women leaders suffer more emotional pain than men when there is a crisis within their teams.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

37 Being in the limelight creates many problems for women.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

20 Women are more likely to be hurt than men when they take big risks necessary for corporate success.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

31 Women believe they have to make too many compromises to gain highly paid positions.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

8 Jealousy from co-workers prevents women from seeking promotions.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

34 Even very successful women can quickly lose their confidence.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

35 Women know that work does not provide the best source of happiness in life.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

18 If women achieve promotions they might be accused of offering sexual favors.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

5 Smart women avoid careers that involve intense competition with colleagues.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

Resilience

38 The more women seek senior positions, the easier it will be for those who follow.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

33 Higher education qualifications will help women overcome discrimination.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

27 Women have the strength to overcome discrimination.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

40 When women are given opportunities to lead they do effective jobs.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

24 Daughters of successful mothers are inspired to overcome sexist hurdles.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

6 Women are capable of making critical leadership decisions.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

21 A supportive spouse/partner or close friend makes it easier for a woman to achieve success in her career.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

32 Successful organizations seek and want to retain talented female staff.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

16 The support of a mentor greatly increases the success of a woman in any organization.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

25 Women's nurturing skills help them to be successful leaders.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

3 Networking is a smart way for women to increase the chances of career success.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

Acceptance

19 Women are just as ambitious in their careers as men.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

12 Women have the same desire for power as men do.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

23 Motherhood is more important to most women than career development.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

22 Women are less concerned about promotions than men are.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

2 Women prefer a balance life more than gaining highly paid careers.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

28 Women reject the need to work incredibly long hours.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

14 Women commonly reject career advancement as they are keener to maintain a role raising children.

1 2 3 4 5 6 7

Strongly disagree Strongly agree

Goal

What is the lowest grade that you would be satisfied with on the upcoming mathematics exam?

F
D-
D
D+
C-
C
C+
B-
B
B+
A-
A

***Knowledge/skills/ability*.**

What was your high school GPA? _____________

What was the scale of your high school GPA? (e.g., 0.0 - 4.0, 0.0 - 5.0)
0.0 - 4.0
0.0 - 4.5
0.0 - 5.0

What was your high school academic rank? If you do not remember, please skip this question.

Stereotype threat.

I worry that my ability to perform well on mathematics tests is affected by my gender.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

I worry that if I performed poorly on mathematics exams this semester, my professor attributed my poor performance to my gender.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

I worry that people's evaluations of me will be affected by my gender.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

I worry that, because I know the gender stereotypes about mathematics achievement, my anxiety about confirming that stereotype will negatively influence how I perform on mathematics tests.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

Though it may not be their fault, females cannot perform as well as males in science and mathematics classes.

1 2 3 4 5 6 7
Strongly disagree Strongly agree

***Math/science identification*.**

I find many scientific problems interesting and challenging.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I have hesitated to take courses that involve math. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I have a lot of intellectual curiosity.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I would have no interest in being an inventor. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I have never been very excited science and mathematics topics. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I am never able to think up answers to problems that haven't already been figured out. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I am good at combining ideas in ways that others have not tried.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I have generally done better in mathematics courses than other courses.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I wish I had more imagination and originality. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

Science makes me feel inadequate. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

Math makes me feel inadequate. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I am quite good at science.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I am quite good at math.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I do not do well on tests that require a lot of reasoning ability. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I'm not much good at problem solving. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I have trouble understanding anything that is based on science and mathematics. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I have always done well in math-related classes.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I am not very original in my ideas, thoughts, and actions. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I never do well on tests that require scientific reasoning. REVERSE CODED
1 2 3 4 5 6 7
Strongly disagree Strongly agree

At school, my friends always come to me for help in science or math.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I have never been very excited about the mathematics. REVERSE CODED

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

I can often see better ways of doing routine tasks.

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

Additional Value

Please answer the following questions in regards to how you feel about math.
I enjoy math.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

Math is useful in everyday problems.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

Math interests me.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

It is important to me to study math.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

I enjoy solving math problems even when I do not have to do them for a class.
1 2 3 4 5 6 7
Strongly disagree Strongly agree

Please answer the following questions.

What is the personal importance to you of making important scientific discoveries?
1 2 3 4 5 6 7
Very unimportant Very important

What is the personal importance to you of managing future technologies?
1 2 3 4 5 6 7
Very unimportant Very important

What is the personal importance to you of inventing new technologies?
1 2 3 4 5 6 7
Very unimportant Very important

What is the personal importance to you of being a leader in your field?
1 2 3 4 5 6 7
Very unimportant Very important

Please answer the following questions.

How similar do you believe you are to individuals who are majoring in STEM (science/technology/engineering/math) topics currently?

1	2	3	4	5	6	7
Not at all much						Very

How similar do you believe you are to other students taking this class?

1	2	3	4	5	6	7
Not at all much						Very

How well do you believe you fit in with the general environment of STEM (science/technology/engineering/math) classes?

1	2	3	4	5	6	7
Not at all much						Very

To what extent do you believe you belong in STEM (science/technology/engineering/math) courses?

1	2	3	4	5	6	7
Not at all much						Very

To what extent do you believe that you fit in with the students in STEM (science/technology/engineering/math) courses?

1	2	3	4	5	6	7
Not at all much						Very

Instructions: Please indicate the degree to which you believe the following statements describe you, using the following scale. Try not to respond merely as you think you ‘‘should’’ respond; rather, try to be as accurate and as objective as possible in evaluating yourself. If any of the questions simply do not seem relevant to you, ‘‘1’’ may be the most appropriate answer
1 = Not at all true of me
2 = Somewhat true of me
3 = Mostly true of me
4 = Absolutely true of me

My area of study helps me live out my life’s purpose.

1	2	3	4
Not at all true of me	Somewhat true of me	Mostly true of me	Absolutely true of

I see my area of study as a path to purpose in life.

1	2	3	4
Not at all true of me	Somewhat true of me	Mostly true of me	Absolutely true of

My area of study is an important part of my life’s meaning.

1	2	3	4
Not at all true of me	Somewhat true of me	Mostly true of me	Absolutely true of

I try to live out my life purpose when I am at school.

1	2	3	4
Not at all true of me	Somewhat true of me	Mostly true of me	Absolutely true of me

Appendix B: Gender Differences on Slope (or Rate) of Self-Efficacy Variables

Using the questions assessing how many hours one needed to study to achieve eight different grade levels and the questions assessing what grade was anticipated based on increasing amounts of studying, I used change in responses across levels (e.g., what grade one believed they could achieve after 5 hours of studying, 10 hours of study, 15 hours of studying, etc.) to assess average differences between the genders. This analysis was performed on self-efficacy for the first exam given substantially more individuals completed these measures (i.e., 165 responses as opposed to 87 responses in the daily diary) than the daily dairy self-efficacy measure. If a gender difference exists, gender at level-2 should relate to one's predicted grade given hours studied at level-1. This was found. That is, the interaction between gender and hours studied using the 8-item measure of self-efficacy was significant, $b_{11} = -0.04$, $SE = 0.02$, $p < .001$ (see Table B1 and Figure B1). This finding indicates that as effort (i.e., hours studied) increased females believed their grades would improve less than the males. Specifically, males believed they would increase their grade level by .56 of a level for every hour studied ($p < .001$) and females believed they would increase their grade level by .48 of a level for every hour studied ($p < .001$). That is, for every extra hour studied, women believed their grade would be .08 below that predicted by the men. Additionally, the difference between the genders at 0 hours was not significant, $b = 0.72$, $SE = 0.63$, $p = .255$. The interaction between gender and hours studied using the 5-item measure of self-efficacy displayed a similar pattern, but it was only marginally significant, $b_{11} = -0.008$, $SE = 0.005$, $p = .080$ (see Table B2 and Figure B2).

Table B1

Set of Models Fit to the Data for the 8-Item Measure

Hierarchical Linear Modeling Equation for Predicted Grade 8-Item Measure

Predicted Grade = $\pi_0 + \pi_1$HoursStudied + e_i
$\pi_0 = b_{00} + b_{01}$ (Gender) + ζ_{00}
$\pi_1 = b_{10} + b_{11}$ (Gender) + ζ_{10}

	b (*SE*) Main effects	b (*SE*) Full interaction
Simple effects		
Intercept, b_{00}	0.94** (0.30)	0.18 (0.38)
Hours Studied	0.24** (0.01)	0.29** (0.01)
Gender	-0.04 (0.48)	0.72 (0.63)
Interaction effects		
Hours X Gender	--------	-0.04** (0.01)

Note: **p* < .05 ***p* < .01

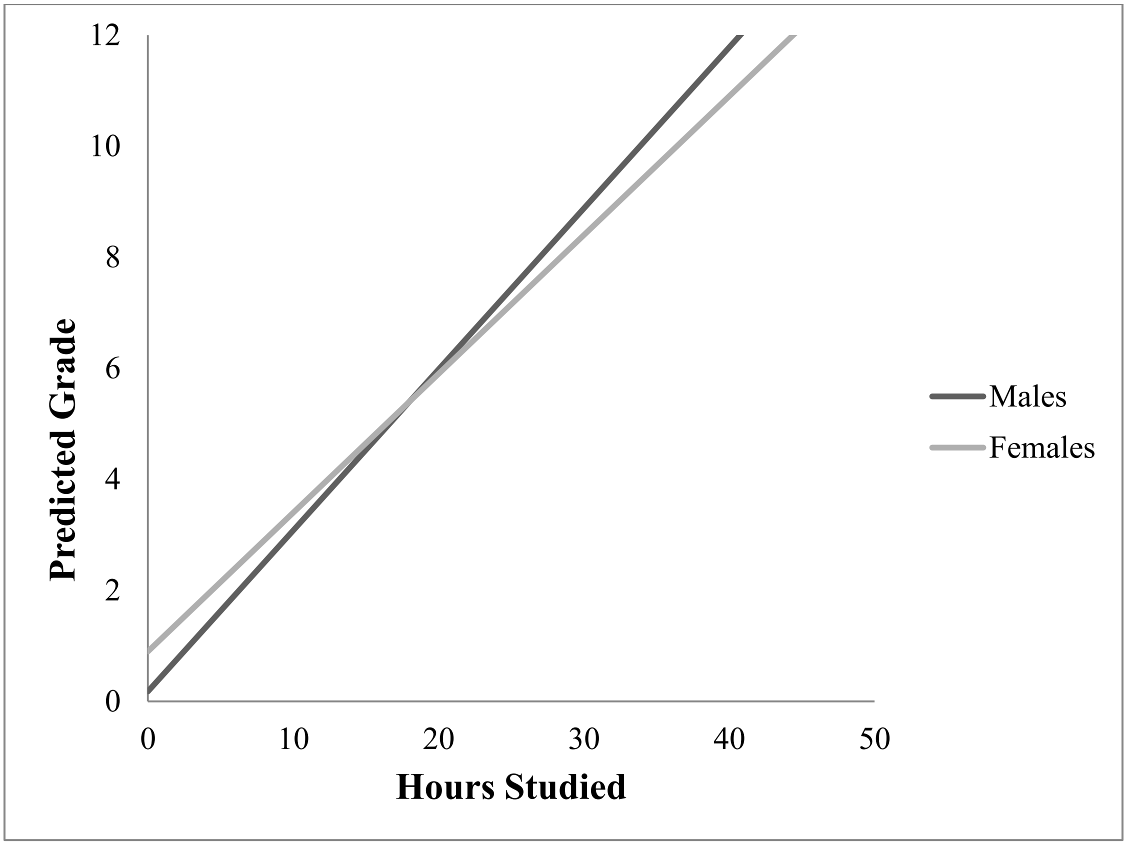

Figure B1. Relationship between hours studied and predicted grade separated by gender using the 8-item measure of self-efficacy. Grade is coded as F = 0 and A = 12.

Table B2

Set of Models Fit to the Data for the 5-Item Measure

Hierarchical Linear Modeling Equation for Predicted Grade 5-Item Measure

Predicted Grade = $\pi_0 + \pi_1$HoursStudied + e_i
$\pi_0 = b_{00} + b_{01}$ (Gender) + ζ_{00}
$\pi_1 = b_{10} + b_{11}$ (Gender) + ζ_{10}

	b (*SE*) Main effects	b (*SE*) Full interaction
Simple effects		
Intercept, b_{00}	1.60** (0.09)	1.40** (0.10)
Hours Studied	0.09** (0.002)	0.1** (0.003)
Gender	-0.17 (0.15)	0.005 (0.17)
Interaction effects		
Hours X Gender	--------	-0.008 (0.005)

Note: *$p < .05$ **$p < .01$

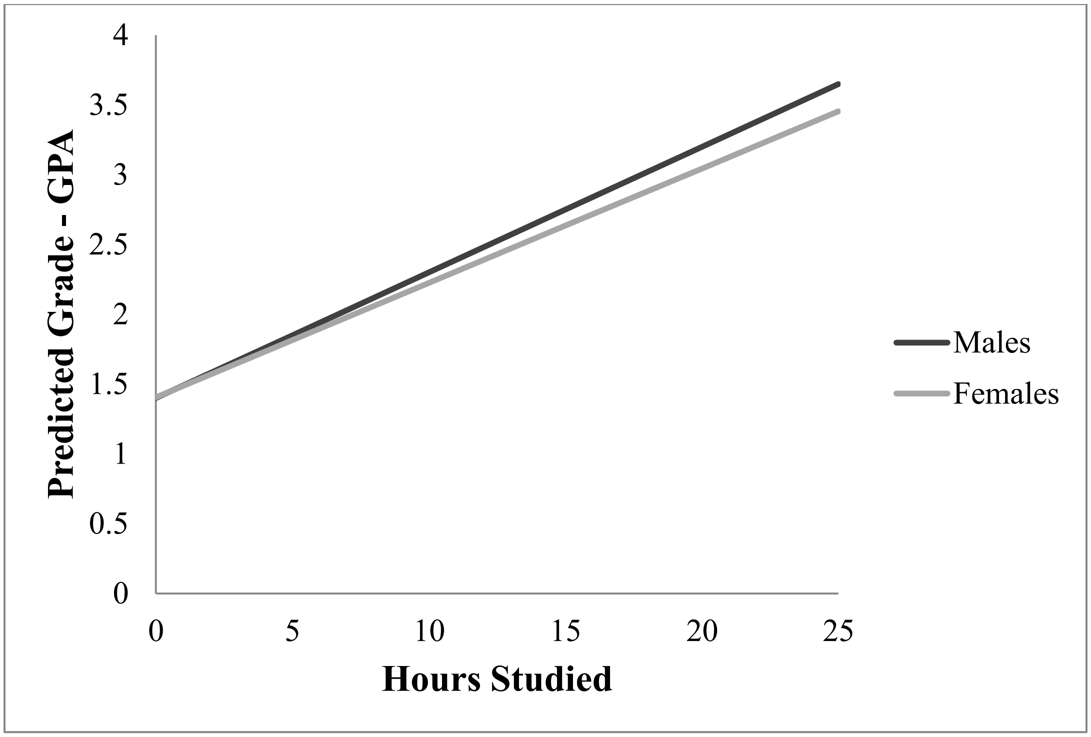

Figure B2. Relationship between hours studied and predicted grade separated by gender using the 5-item measure of self-efficacy. Grade is coded in GPA units.