Boston College
Lynch School of Education and Human Development

Department of Counseling, Developmental, and Education Psychology
Applied Developmental and Educational Psychology

EXPLORING PREDICTORS AND OUTCOMES OF GENDER DIFFERENCES IN MATH CLASSROOM PARTICIPATION

Dissertation
by

MEGHAN L. COUGHLAN

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

May 2022

Abstract

Exploring Predictors and Outcomes of Gender Differences in Math Classroom Participation

Meghan L. Coughlan

Dissertation Chair: Dr. Marina Vasilyeva

Motivated by the underrepresentation of women in math-related majors and occupations, the present dissertation explored the possibility that disparities in frequency of classroom participation—a measure of engagement prone to gender differences—could be partially responsible for gender differences in belonging and identity in math contexts. Via the introduction of a novel psychological construct, *class participation confidence threshold*, and the adoption of a Regulatory Focus Theory framework, the present work aimed to investigate mechanisms underlying gender differences in math classroom participation, as well as how these disparities might contribute to more distal outcomes.

The dissertation consisted of two studies conducted with undergraduate students. Both studies tested a theoretical model that posited potential predictors and outcomes of gender differences in classroom participation. Study 1 ($N = 161$) was a cross-sectional investigation of students' participation frequency, reported based on their general experience in current math and social science classes. Study 2 ($N = 269$) investigated the same associations using a daily diary methodology, including pre- and post-measures of relevant constructs.

Results indicated that, when considering opportunities for participation, women participated less than men—both in math and the comparison domain of social science. In addition to less frequent participation, women generally displayed higher stereotype threat susceptibility and confidence thresholds in both domains. Women also demonstrated higher levels of a prevention focus in math compared to social science, while men's regulatory

orientation was similar across domains. With respect to the proposed outcomes of participation (i.e., belonging, identity, and career interest), women exhibited lower levels than men in math and equal or higher levels in social science.

Path analyses revealed that students' regulatory focus predicted their classroom participation and that this relation was mediated by their confidence threshold. Importantly, these results persisted even when controlling for motivational variables traditionally regarded as predictors of classroom participation according to Expectancy-Value Theory. The results suggest that an increased prevention focus in the math domain may lead women to set higher confidence thresholds and participate less frequently.

Acknowledgements

I owe a number of people a heartfelt thank you for their continued support throughout this dissertation research and my PhD experience in general. First, to my dissertation chair, Dr. Marina Vasilyeva, thank you for your encouragement and guidance during this dissertation work and countless other projects before it. I am so grateful to have the support of a mentor who is always willing to move mountains on my behalf. Dr. David Miele, I owe you a sincere thank you for your mentorship, support, and generous funding of my work that extended far beyond the typical role of a dissertation reader. It was your enthusiasm for and belief in this work from its nascent stages that propelled it to this point, and I am extremely appreciative of your thoughtful contributions along the way. Dr. Eric Dearing, thank you for sharing your vast analytical expertise and optimistic outlook throughout the duration of my dissertation work. Dr. Alex Browman, thank you for always being willing to answer my questions, even in your incredibly busy new faculty role, and for generously assisting me with recruitment at Holy Cross.

I would also like to thank the members of the MML lab for providing feedback on these dissertation studies and completing countless pilot surveys over the past several years. Thank you to Dr. Melissa McTernan and Linxi Lu for providing guidance on some analytical dilemmas. Thank you also to the undergraduate students who shared thoughtful responses for these studies, even in the midst of a seemingly never-ending pandemic. I also owe gratitude to Boston College and the College of the Holy Cross for allowing me to recruit participants from their institutions.

Next, I owe a big thank you to the remarkable friends I made in ADEP who kept me optimistic and sane through every step of the way. In particular, thank you to Sidney May and Lindsey Caola for the daily texts of support, frequent Zoom calls, and nonjudgmental encouragement. Sidney, if a cohort “lottery” existed, I would be the undisputed winner. I am so lucky to have had you as my companion through this adventure. Lindsey, you have been both a wise mentor and a close friend, and I am so thankful for all of your guidance and kindness. I am endlessly grateful to both of you for providing me a space to brainstorm through challenges, ask questions I was too embarrassed to ask elsewhere, and laugh with people who really “get it.” I did not expect to meet lifelong friends when I entered graduate school, but now I wonder how I ever could have done this without you.

I owe a debt of gratitude to my family and friends that I can never hope to repay. Mom and Dad, thank you for your unwavering support in every aspect of my life and for your countless sacrifices that have allowed me to pursue my dreams. I love you both for encouraging me to be whoever I want to be while always reminding me what is most important in life. Matthew, I am grateful to have a “little” brother to look up to as a wiser and more thoughtful version of myself. It is not lost on me how lucky I am to have a sibling who doubles as a best friend. To my dear friend, Beth, thank you for encouraging me to pursue graduate work in the first place and for always being available for a phone call or hug to help me keep things in perspective. To my soon-to-be parents-in-law, Trish and Rick, thank you for your enthusiasm and support since the first day I met you. Your sincere interest in my work means so much to me, and I am grateful to call you family.

Most importantly, thank you to my fiancé, Mac, who has been a steadfast source of encouragement, comfort, and enthusiasm throughout my PhD journey. I do not have adequate words to express how much credit you deserve for this dissertation. Your support on the more challenging days and delight over the smallest of successes along the way are what guided me to this point. Thank you for loving me when I was impossible and for bringing balance into my life. I am eternally grateful for a partner whose pride in me helps me feel proud of myself.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

A number of sources have documented the overwhelming homogeneity of STEM fields in terms of gender. Women are underrepresented in math-related majors and earn fewer STEM degrees than men (National Science Foundation, 2019). Disparities observed within universities become even more apparent when examining who decides to pursue math-related occupations (e.g., Else-Quest et al., 2010). In fact, “horizontal sex segregation” across fields of concentration at the university level is undoubtedly a source of the persistent inequality within careers requiring a STEM university degree (Alon & Gelbgiser, 2011; Smyth & Steinmetz, 2008).

One potential explanation for these disparities that has been posed is gender differences in math ability. Support for this proposition comes from work demonstrating that boys succeed over girls in intense, high-stakes situations, such as standardized tests (Reardon et al., 2018) and math competitions (Ellison & Swanson, 2010). On the other hand, there is evidence that girls actually excel over boys on other measures of math performance, particularly those requiring sustained effort—e.g., as measured by grades in math classes (O’Dea et al., 2018). Since it seems safe to assume that grades reflect some level of ability, the fact that female students in general outperform male students in math grades suggests that many women leave school with the skills they would need to successfully pursue a STEM career. Thus, in exploring the reasons for gender disparities in the STEM workforce, it may be more productive to focus on whether there are social or motivational factors—such as identification with the subject and sense of belonging in math contexts—that better explain women’s reticence to pursue math-related careers.

Sense of belonging encompasses a student’s feelings of fit between themselves and a given context (Walton & Brady, 2017). In extant research, belonging has been found to predict interest in particular careers, including those in STEM fields (Cheryan et al., 2009; Good et al., 2012). Math identity, or a student's perception of themselves as somebody who is a math doer or learner (Anderson, 2007), has also been shown to be a critical influence of STEM career interest (Cribbs et al., 2020; Lock et al., 2013).

One possible reason why female students may experience less belonging in STEM classrooms pertains to how belonging is associated with certain behavioral indicators of student engagement during math learning. In particular, one such behavioral measure of engagement is classroom participation (Fredricks et al., 2004)—e.g., raising one's hand with the intention of volunteering to answer an instructor-initiated question, regardless of being called upon to respond. The sense of belonging and identity students have when they come to class likely shapes their participation behaviors (Zumbrunn et al., 2014). In turn, students' sense of belonging and identity may, over time, be bolstered through a sustained experience of active participation in the classroom (Christenson et al., 2001).

Several studies have established the existence of gender differences in classroom participation behaviors, including in STEM contexts (e.g., Eddy et al., 2014), although the reasons underlying these differences typically have not been investigated. Without an understanding of why female students are less likely to engage in math contexts, we cannot begin to explore potential ways of addressing this disparity.

There are a number of potential mechanisms to explain why some students might actively participate more than others. One that seems particularly compelling draws on an expectancy-value model (Wigfield & Eccles, 2000). In line with this model, researchers have examined why students might perceive it as worthwhile to answer teachers' questions (e.g., Jansen, 2006; Turner & Patrick, 2004), as well as whether they believe they are capable of answering them correctly—i.e., whether they expect to succeed (e.g., Galyon et al., 2012). For instance, Böheim et al. (2020a) examined the role of interest, which could be considered a type of value, and self-concept, which is closely related to students' math competencies (e.g., "I learn math quickly"), via an observational study of high school students' hand-raising behaviors in math and language arts classes. They found that math-specific self-concept positively predicted participation in math classes (as measured by hand-raising), whereas students' language arts-specific situational interest was predictive of their participation in language arts. Thus, depending on the domain, expectancies and value can each be predictive of participation behaviors.

However, students' expectancies and domain-specific value may not fully explain their likelihood of participating. For instance, two students with the same expectancies of success could still require different levels of confidence in their knowledge in order to be willing to participate. That is, some students might not feel comfortable participating unless they have no doubt in the validity of their contribution, whereas others may be more willing to share a response in which they are only somewhat confident. These differences in *class participation confidence thresholds* may be determined by a separate motivational factor that is independent of expectancies.

A good candidate for explaining why some students might adopt different confidence thresholds for participation comes from the social psychology literature. According to regulatory focus theory (Higgins, 1997), people sometimes pursue their goals by vigilantly protecting against potential threats and losses (a prevention focus) and other times by eagerly seeking opportunities for gain (a promotion focus). Thus, students might adopt different confidence thresholds depending on their general regulatory orientation within a domain; a higher threshold might correspond to a prevention focus, whereas a lower threshold might align with a promotion focus.

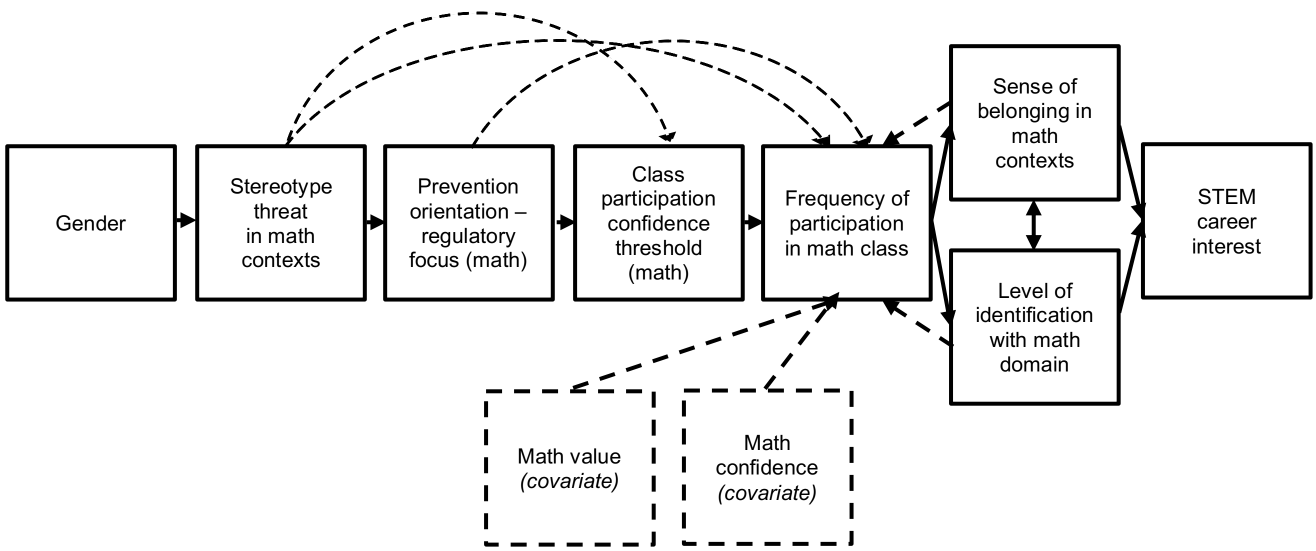
Another reason why a regulatory focus perspective is a potentially useful lens through which to explore potential gender differences in students' confidence thresholds is because of its connection to stereotype threat, which often affects women in STEM settings. Some evidence suggests that the extent to which students are more prevention-focused in math contexts might be due in part to stereotype threat effects, or concerns about confirming negative stereotypes about one's group in contexts in which one's identity is salient (e.g., Ståhl et al., 2012). Given that female students are more susceptible than males to the effects of stereotype threat in math classes (Spencer et al., 1999), it seems possible that women might exhibit lower frequencies of participation in math classrooms because this threat induces a prevention focus, which leads them to be more cautious about ensuring that they do not reply to the instructor with an incorrect answer.

The primary goal of this dissertation is to better understand mechanisms underlying gender differences in active classroom participation in mathematics classrooms, as well as how these disparities might contribute to other, more distal outcomes such as math-related career pursuit. Two studies probed a theoretical model (see Figure 1) accounting for the potential roles of several motivational variables in the link between student gender and STEM career choice. Study 1 was an associational, survey-based exploration of these constructs. Study 2 explored the same associations using a daily diary methodology, including pre- and post-measures of relevant constructs such as belonging and identity, allowing for a deeper exploration of directionality in the model. In addition to examining a number of established psychological constructs, this dissertation also introduces a novel concept: *class participation confidence thresholds*.

Identifying the psychological processes that contribute to gender differences in students' frequency of math class participation is not only theoretically informative. The process also illuminates potential strategies that teachers and other practitioners can use to encourage engagement and persistence in the math domain from a more diverse group of students.

Figure 1

Proposed Theoretical Model



Note. This model attempts to explain variance in students' classroom participation, belonging, identity, and career interest, assuming similar levels of expectancies and value across participants. Solid lines indicate the primary posited associations in the model, whereas dotted lines represent potential direct or reciprocal effects that may not be explicitly investigated.

CHAPTER 2: LITERATURE REVIEW

The existence of gender disparities in STEM majors and careers is well established (National Science Foundation, 2019). In particular, a large body of research has investigated differences in math ability as one way to understand what contributes to these disparities. Accumulating evidence, however, implies that differences in STEM participation cannot be fully accounted for by cognitive and ability-based differences (Hyde et al., 2008; Spelke, 2005), suggesting the potential value of examining motivational factors as well.

Driven by this need for motivational research, the present dissertation tests the model outlined in Figure 1, which proposes mechanisms through which motivational factors may influence students' patterns of behaviors as learners, and eventually their career choices. Specifically, this work focuses on the behavioral construct of classroom participation, exploring motivational factors that may contribute to differences in participation frequency, such as regulatory focus and stereotype threat. Further, it investigates the extent to which classroom participation is related to students' sense of belonging and identification with STEM domains, as well as to their decisions about career choices.

In the literature review that follows, I discuss research related to each of the components of this model and highlight connections between them. I first address the existing research on classroom participation, followed by a review of the motivational factors that are posited as potential predictors—regulatory focus, stereotype threat, and associations between the two constructs. Finally, I discuss extant literature on belonging and identity, constructs that are potentially influenced by students' participation within math classrooms.

Classroom Participation

"Classroom participation" can refer to a number of different behaviors occurring within an academic setting, including responding to an instructor-initiated question, volunteering a comment without answering an instructor's prompt, and asking a question to the instructor (Fritschner, 2000). Hand-raising in particular has been recognized as an easily observable metric of behavioral engagement (Böheim et al., 2020b) and is therefore often used as a proxy for measuring participation within a classroom.

Prior research has established the significance of classroom participation in relation to a variety of important outcomes. For instance, studies have shown that active participation, including hand-raising, positively predicts measures of student achievement, such as course grades and standardized test scores (Narayan et al., 1990; Subotnik & Strauss, 1995; Webb et al., 2014), even when controlling for prior achievement (Böheim et al., 2020b). One potential explanation for this is that active student participation provides instructors with feedback about concepts the class does or does not understand, allowing them to deliver targeted instruction based on what is unclear to students (Stowell & Nelson, 2007). Additionally, student dialogue is instrumental in cultivating conceptual understanding and critical thinking, both of which are vital to math learning (Dixon et al., 2009).

Along with purely achievement-related outcomes, classroom participation also fosters students' sense of agency, accountability, and involvement in their own learning (Dallimore et al., 2004; O'Connor, 2013) and nurtures a sense of belonging (Christenson et al., 2001; Goldstein & Benassi, 1994). As I will explain later in this chapter, this last set of outcomes is particularly important to consider when identifying reasons underlying students' decisions whether or not to commit to a career in STEM.

Gender Disparities in Participation

Despite the important benefits of participation, research has shown that girls and women participate less frequently and less assertively than boys and men in educational settings, including academic seminars (Carter et al., 2018) and university classrooms (Aguillon et al., 2020; Fassinger, 1995; Guinier et al., 1994). In a study of introductory biology classes for science majors, Eddy et al. (2014) found that although women comprised about 60% of the students in the courses, only 37% of responses to questions posed by instructors were from female students. Similarly, in a self-report study of university students conducted by Crombie et al. (2003), male students reported a significantly higher frequency of participation compared to female students in the same classes.

An important body of work that attempts to explain gender disparities in classroom participation emphasizes the role of environmental factors, such as teachers' disproportionate attention to male students (Eliasson et al., 2016; Jones & Dindia, 2004; Sadker & Sadker, 1986) and other kinds of systemic discrimination that disadvantage female students. In an observational study of students in a single-sex math class, Streitmatter (1997) found that seventh- and eighth-grade girls were more willing to answer questions in their female-only math class than in their other, coeducational classes. The authors point to these results as evidence of the influence of environmental factors (i.e., the classroom's gender composition) on females' levels of participation.

In their seminal work, Hall and Sandler (1982) introduced the concept of a "chilly climate," which refers to an environment, dominating many STEM classrooms, of salient gender stereotypes and a generally masculine cultural atmosphere. The authors discuss a variety of ways in which women face discrimination in STEM classrooms, pointing out that these subtle cues from professors and peers might eventually lead women to “believe and act as though their presence in a given class...is at best peripheral, or at worst an unwelcome intrusion [and that] their participation in class discussion is not expected, and their contributions are not important” (p. 3). More recently, Lee and McCabe (2021) investigated whether the “chilly climate” persists in more modern university classrooms. Via both qualitative and quantitative observations, the authors found that men still participate more frequently and assertively than women. Importantly, they cited differential treatment by professors (e.g., referring to female students as “young lady”) as a factor influencing the existence of these “gender status hierarchies” (p. 52). For an understanding of how gender expectations in the environment may become internalized, see the literature on gender role norms (e.g., Diekman et al., 2010; Wood & Eagly, 2012).

Self-Concept and Participation

As detailed above, there are a number of potential factors contributing to these frequently observed gender disparities in the quantity and nature of classroom participation. A logical next step is to consider the particular psychological mechanisms explaining the connection between these environmental factors and resulting student participation behaviors. Markus and Wurf’s (1987) work on identity-based motivation supports the notion that students’ self-concept is dynamic, reacting to their environment in a given moment. Thus, it is reasonable to suppose that external factors, such as the gender composition of a classroom, could be internalized and interpreted by individuals in ways that encourage or hinder their participation via their self-concept.

In fact, students’ self-concept—used interchangeably in the present dissertation with “self-efficacy,” “confidence,” and “perceived competence”—is a reasonable (and perhaps somewhat obvious) psychological explanation for disparities in classroom participation. It seems probable, for example, that students with a higher level of confidence in the class material would participate more frequently. In fact, perceived competence might be particularly important in math contexts. Results from an observational study of high school students by Böheim et al. (2020a) suggest that math-specific self-concept is *especially* predictive of participation in the math domain. Although the researchers also explored the language arts domain, they did not find analogous effects in those classes. They speculated that high self-concept may be particularly critical for prompting participation in math contexts, given the high-stakes nature of answering math questions that typically only have a single correct answer. In contrast, questions in language arts contexts may more often be open to interpretation, allowing for a much wider range of suitable responses.

Despite its importance, a large body of literature documents that relatively low perceived math competence is typical for females of all levels of math ability—even those who exhibit high levels of achievement (Leder, 1988; Thomas & Costello, 1988). As early as first grade, girls report lower perceived math ability compared to boys, despite a lack of corresponding gender differences in actual math achievement (Jacobs et al., 2002; Wigfield et al., 1997) and endorse stereotypes associating "brilliance" with males (Bian et al., 2017). Moreover, although work by Jacobs et al. (2002) has demonstrated an evaporation in gender differences in perceived competence by the end of high school, other studies suggest that this gap only widens as children advance across school years (Ehrtmann & Wolter, 2018). Thus, there is reason to expect that these disparities might continue to be present in a college sample.

These differences in perceived ability may translate to disparities in participation behaviors, in part because students low in perceived math competence may not feel capable of correctly answering the instructor's questions. According to Expectancy-Value Theory, students' expectations of success (which are strongly correlated with their perceived competence) and task values (i.e., their perceptions of whether the task is worth pursuing) influence their behaviors in the form of choices, engagement, and persistence (Wigfield & Eccles, 2000). Thus, differences in perceived competence could reasonably account for some variation in participation behaviors (and ultimately, disparities in STEM career interest).

Class Participation Confidence Threshold

Although perceived competence is undoubtedly predictive of classroom participation, it is unlikely that it is the *sole* psychological factor accounting for differences—gender-related or more generally—in active classroom participation. Galyon et al. (2012) examined the relation between self-efficacy and class participation in undergraduate students and found that there is not necessarily a linear association between the two constructs. The authors identified self-efficacy groups (low, medium, and high) via a cluster analysis of self-efficacy scores—as measured at five time points throughout the semester immediately prior to exams—and selected participation groups based on quartiles of students' cumulative comments throughout the semester. Analyses demonstrated the intuitive finding that students with high self-efficacy were more likely than students with low or medium self-efficacy to exhibit high levels of active participation. Importantly, however, students in the low and medium self-efficacy groups did not differ in their participation.

The finding that only membership in the high self-efficacy group was significantly related to membership in the high participation group points to the potential existence of an internal threshold that specifies how confident a student must be in their answer to be willing to participate in class in a particular domain. This *class participation confidence threshold* is likely independent of students' domain-specific confidence, although the two may interact in influencing behavior. The threshold might also vary between individuals or groups, within individuals across different domains, or within a domain across time.

There is reason to believe that the level of confidence required for a person to volunteer an answer could differ depending on how much that person worries about the consequences of providing an incorrect answer. For instance, one can imagine two students who are both fairly confident in their math ability and about knowing the correct answer to a question posed by their instructor. However, one student may be much more concerned about the small possibility of making a mistake and what the consequences of this mistake might be. Thus, even though this student may be just as confident in their knowledge as the other student, they may be more reluctant to participate.

Such variation in students' confidence thresholds could potentially explain gender differences in math classroom participation. Since gender disparities in the math domain in general appear to be present in certain high-pressure situations, such as while taking standardized tests (Ellison & Swanson, 2010; Reardon et al., 2018), it is possible that gender differences in math classroom participation could reflect a perception among female students (in contrast to male students) that class participation is a rather high-stakes behavior involving a lot to lose.

Regulatory Focus Theory

What might lead students to adopt different confidence thresholds, either within different domains or at different points in time? Regulatory focus theory, which distinguishes between two motivational orientations that guide people's goal pursuit, may offer an explanation for why some students might require more confidence than others in order to participate. In regulatory focus theory, a promotion focus is an eager, gain-oriented state of mind, whereas a prevention focus reflects a vigilant, loss-oriented mindset (Higgins, 1997). A person with a prevention focus is concerned with ensuring that they do not make mistakes or experience losses; they approach tasks in a vigilant manner to guard against potential threats.

A number of empirical studies in the regulatory focus literature demonstrate how regulatory focus can shape people's information processing and behavior. For example, Crowe and Higgins (1997) administered a word recognition memory task to participants who had been randomly assigned to either a prevention or promotion framing condition. Participants in the prevention-focused condition were given instructions including language such as: "As long as you don't do poorly on the word recognition memory task…" On the other hand, the instructions for the promotion framing condition contained language such as: "If you do well on the word recognition memory task…" In a series of three trials, participants first saw a list of 20 nonsense words, and then were asked to rate (in a new set of 40 words) whether they had seen each word before. Results demonstrated that participants in the promotion focus condition were more biased toward taking risks in order to ensure that they identified correct words and avoided errors of omission—i.e., they were more likely to say yes, they had seen the word before. In contrast, participants in the prevention focus condition were more conservatively biased, prioritizing correct rejections and avoiding errors of commission—i.e., they were more likely to answer no, they had not seen the word before. These results suggest that in a state of relative uncertainty, a promotion-focused person might act daringly by sharing their response (in order to take a chance on being correct), whereas a prevention-focused person might act vigilantly by withholding their response (in order to avoid making a mistake).

Importantly, people might adopt distinct regulatory focus orientations in different domains, or even for different tasks or situations within a domain. For example, promotion motivations might be evoked by certain situations with gain-focused incentives, and prevention orientations may be induced by other situations that include loss-focused incentives (Molden et al., 2008; Shah et al., 1998). Furthermore, Browman et al. (2017) found support for the existence of domain- and situation-specific motivations for prevention or promotion that are separate from a person's domain-general regulatory orientations. Thus, although people can have a chronic tendency towards a promotion focus or prevention focus, situation-specific promotion and prevention states can be activated by cues in the environment (e.g., Zou et al., 2020).

Gender Differences in Regulatory Focus

Research suggests that in general, female students might be more susceptible to a prevention focus. In an experiment involving the administration of questions from practice tests for the SAT II United States History and World History exams, Baldiga (2014) found that all test takers, regardless of gender, answered every question when they were told that there was no penalty for incorrect responses. However, in the condition when there was a small penalty for wrong answers, female students responded to significantly fewer questions than male students. This tendency to skip more questions put the female participants at a disadvantage in overall test scores, despite the fact that the experimenters found no actual gender differences in knowledge of the test material. Importantly, the researchers also found *no* gender differences in confidence. The gap appeared to be mostly due to differences in risk tolerance, suggesting that women might experience a higher sense of vigilance in high-pressure situations such as standardized tests.

These effects appear to be particularly strong in STEM contexts. Coffman and Klinowski (2020) examined the influence of a policy change on the Chilean college entry exam, a high-stakes standardized test focusing on STEM topics such as math and chemistry, that recently eliminated penalties for incorrect answers. Prior to the policy change, female test-takers—particularly those in the top quintile of performers—skipped more questions, on average, than male test-takers. However, after the elimination of these penalties, the gender gap in number of skipped questions was reduced by 70% overall (and by 79% among the most talented test-takers), shrinking the gap in achievement by between 8% to 16% and increasing female representation in the group of students with the highest test scores.

Tannenbaum (2012) examined data from the mathematics portion of the 2001 administration of the SAT exam, finding both that female students' scores were significantly lower than those of their male peers and that female students skipped significantly more questions than male students. In fact, the author argues that up to 40% of the gap in SAT math scores could be explained by the gender difference in number of questions skipped. Other studies of risk-taking behaviors and skipped questions on standardized exams in the math domain have found similar results (Atkins et al., 1991; Ramos & Lambating, 1996).

Several studies in the regulatory focus literature have employed manipulations with a similar structure to the point system of the standardized exams described above. For example, in one experimental manipulation, Rosenzweig and Miele (2016) told participants that they had already been entered into a lottery with a $50 prize but would lose entry if their performance did not meet certain criteria (prevention condition), or that they would gain entry into a lottery with a $50 prize if they met the same performance criteria (promotion condition). The loss-framed language of the prevention orientation (i.e., taking away lottery eligibility) is similar to the penalties associated with incorrect responses on the standardized exams described above (i.e., taking away a certain number of points), although participants actually performed better in the prevention condition in these studies. Therefore, it is possible that the female students described in the studies above were more reticent to answer exam questions about which they were unsure because they were particularly sensitive to the penalties associated with incorrect answers. Taken together, these results lend support to the hypothesis that female students are perhaps more likely than male students to adopt a prevention orientation in situations they perceive to be high-pressure or threatening.

Participation and Regulatory Focus

With respect to participation, a person operating with a prevention focus might be hesitant to answer questions posed by the instructor because the possibility of responding incorrectly represents a loss that outweighs whatever is to be gained by answering correctly. It is important to note that this person could have high perceived competence in the subject and even be relatively confident about knowing the correct answer, but their increased state of vigilance—and the high cost associated with answering incorrectly—might discourage their willingness to answer.

Abdullah et al. (2012) conducted a focus group study comparing university students in social sciences courses who they identified, based on a series of observations, as either displaying “active” participation behaviors (e.g., answering questions posed, asking questions, or presenting their opinion on a topic) or “passive” participation behaviors (e.g., listening to the instructor or taking notes without engaging vocally). Findings from the focus groups revealed that the more passive participants cited fear of being incorrect, offending others, or being humiliated by the instructor as reasons for avoiding active participation. Although the study did not examine students’ regulatory focus orientations, the participants’ responses seem to align with concerns that are characteristic of a prevention focus.

Participation in math contexts might be particularly susceptible to the effects of a prevention orientation. Because math is a subject where there is generally a “correct” answer, students face a relatively high risk of being wrong when raising their hand in a math class. Thus, active classroom participation in math is essentially a measure of students' willingness to attempt to answer a question that has an objectively correct answer (and, therefore, numerous incorrect answers as well). In contrast, questions posed in some humanities courses might have more subjective answers and therefore may be less likely to induce a prevention orientation.

Stereotype Threat

What might lead a student—particularly a female student—to adopt a prevention focus in a math setting? One potential source is stereotype threat, a phenomenon of performance decrements occurring when members of negatively stereotyped groups become focused on the possibility of confirming these stereotypes (Steele, 1997). These effects have been demonstrated empirically in academic settings across many stereotyped groups, including racial and ethnic minorities (Aronson & Salinas, 1997; Steele & Aronson, 1995) and individuals of low socioeconomic status (Croizet & Claire, 1998).

With respect to gender stereotypes, extensive research has documented that in the math domain, female students are particularly susceptible to stereotype threat. In a seminal set of three studies, Spencer et al. (1999) explored stereotype threat experienced by women in math contexts by administering math tests that were described either as producing gender differences or not producing such differences. Results showed that when the test was described as susceptible to gender differences (i.e., stereotype threat for female participants was high due to increased concerns about confirming these negative expectancies), women performed significantly worse than men did. However, this gap was eliminated when the test was described as not being prone to gender differences (i.e., when stereotype threat was reduced by making gender-related concerns less salient).

There is also evidence that even when the threat is much more subtle, gender-related stereotype threat effects still influence women's performance in math contexts. For example, Neuville and Croizet (2007) identified stereotype threat effects in seven- and eight-year-old girls whose gender was implicitly made salient prior to a math test. In order to activate gender identity, researchers based a manipulation on a prior study by Ambady et al. (2001), instructing girls to color a picture of a little girl holding a doll and boys to color a picture of a boy holding a ball. Findings showed that when gender identity was activated, girls performed more poorly on difficult problems. However, activation of gender identity did not influence boys' performance, implying the existence of gender-specific stereotype effects.

Inzlicht and Ben-Zeev (2000) found that even something as inconspicuous as the gender makeup of the classroom could influence female students' experiences of threat. The researchers administered math and verbal tests to participants in groups of three people—either in a minority condition (where the two other participants were of the opposite gender of the participant) or a same-sex condition (where all three participants were the same gender). Whereas male participants performed equally well in both conditions, female students in the minority condition performed worse on the math test only.

Taken together, the results of these studies suggest that even subtle cues of gender-related stereotypes can induce threat effects for female participants in math contexts. Various mechanisms have been posited to explain these effects. One potential mechanism is distraction, which posits that threat compromises individuals' performance via executive functioning by diverting working memory and attention to thoughts and concerns that are not relevant to the task (Beilock et al., 2004). Explicit monitoring theory, on the other hand, proposes that threat-inducing situations disrupt individuals' execution of tasks by directing their attention inordinately to skill processes (Beilock & Carr, 2001; Jackson et al., 2006). Beilock et al. (2006) combined these explanations and proposed that stereotype threat undermines performance by simultaneously occupying working memory with concerns *and* boosting perseveration on gaining control, resulting in a “double whammy” (p. 1062). Thus, the performance decrements exhibited in studies exploring stereotype threat are likely a reflection, in part, of participants’ perseveration on avoiding negative outcomes—a description which happens to be aligned with that of a prevention focus.

Notably, research has found that stereotype threat effects may be particularly strong for women who are moderately to highly identified with the math domain (Nguyen & Ryan, 2008; Steele, 1997). In a review of research on stereotype threat, Steele et al. (2002) claim that stereotype threat effects in a given domain are more strongly experienced by (or perhaps even *uniquely* experienced by) individuals for whom the domain is an important component of their identity. In other words, a person must care about a domain in order to be susceptible to stereotype threat effects within it. Thus, not all women will necessarily experience stereotype threat in their math classes.

Associations Between Stereotype Threat and Regulatory Focus

A small set of studies have identified associations between stereotype threat and regulatory focus. Research has found that stereotype threat susceptibility may lead students to experience elevated levels of agitation and anxiety (e.g., Osborne, 2001) and to become concerned with making mistakes and approach the task in a cautious or vigilant manner (Ståhl et al., 2012), all of which are associated with a prevention orientation.

A study by Seibt and Förster (2004) has also lent credence to this relation, suggesting that the experience of stereotype threat induces a prevention focus. More specifically, through a series of five studies, the authors demonstrated that negative stereotypes prompt a vigilant prevention focus (as evidenced by less creative and more analytical performance), while positive stereotypes induce a more eager promotion focus (as suggested by enhanced speed and creativity). A notable implication of this study is that a prevention focus—due to stereotype threat or some other cause—is not always associated with poor performance and other negative outcomes; it depends on the nature of the task. Since a prevention orientation alters one's strategic approach to be more vigilant, the extent to which the orientation might be helpful or harmful may depend on whether it is a good fit for a given task (Scholer et al., 2018). Therefore, in environments where the reward for making an incorrect guess is greater than the penalty of answering incorrectly, withholding responses due to a prevention orientation might be particularly maladaptive.

Research by Oyserman et al. (2007) also explored whether priming membership in a stigmatized group can induce a prevention focus. Undergraduate participants were randomly assigned to stigmatized and non-stigmatized conditions based on a pre-screener. Results demonstrated that making group membership salient can lead to an increased prevention focus in members of negatively stereotyped groups. However, it is critical to note that "regulatory focus" in this study was measured with a scale by Lockwood et al. (2002), which—although widely used—has since been questioned by some experts in the field as an inappropriate measure of regulatory focus, due to concerns about its validity (Molden & Winterheld, 2013). Although the results of this study should be interpreted with a great deal of caution, it potentially provides some additional evidence for an association between unfair treatment (and related stereotype threat expectancy) and prevention orientation.

A Note on the Current State of Stereotype Threat Research

I would like to add an important note with respect to the current state of stereotype threat research in general. Within the last several years, there has been increasing deliberation in the field over both the replicability and measurement of stereotype threat effects. While there is some published literature contributing to this conversation, scholars are primarily engaging on this topic via other mediums, such as blog posts, podcasts, and social media (e.g., Twitter) posts.

Recent meta-analyses have called into question the results of stereotype threat literature, particularly with respect to the experiences of female students in math environments (e.g., Flore & Wicherts, 2015; Stoet & Geary, 2012). In light of these investigations, which highlight the prevalence of publication bias in stereotype threat literature and low rates of replication, some previously enthusiastic champions of stereotype threat began to publicly grapple with their previous work. For example, Dr. Michael Inzlicht, a renowned stereotype threat researcher, admitted his uncertainty via a post on his blog: “Now I am not as certain as I once was about the robustness of the [stereotype threat] effect... There is a lot of evidence supporting it. That said... I would be lying if I said that doubts have not crept in” (Inzlicht, 2016).

In addition to concerns about the robustness of the effects, the conversation in the field has focused on how stereotype threat is operationalized. Historically, stereotype threat has been primarily measured as an effect on performance outcomes, with a separate psychological process driving it. More recently, researchers have noted that there has not been enough focus on the way that Steele (1997) originally wrote about stereotype threat, as a “threat in the air.” Modern researchers argue that the focus should be redirected from the extent to which students experience performance-related outcomes of threat to the extent to which they *feel* stereotype threat. For instance, in a short opinion paper, Lewis and Sekaquaptewa (2016) highlight how the majority of stereotype research has focused on performance outcomes (particularly test scores) and call for a "broader view" of stereotype threat.

It is important to note that despite these concerns, there seems to be a general consensus among social psychologists that stereotype threat most likely does exist; it has just not typically been measured in a way that honors its psychological nature. Although some research has attempted to determine the psychological mediators of the performance deficits caused by stereotype threat (e.g., Schmader et al., 2008), even that work often focuses on the resulting performance deficits. I considered these concerns while developing the measure of stereotype threat susceptibility used for this dissertation work (addressed in more detail in the *Methods* sections below), which attempts to focus on the affective experience of threat rather than performance implications.

Outcomes Influenced by Participation

The present dissertation is not only focused on identifying potential mechanisms contributing to gender differences in participation behaviors such as frequency of raising one's hand in class. It also aims to explore broader outcome measures that might be influenced by these narrow measures of engagement. In particular, evidence suggests that both sense of belonging in math contexts and math identity might be directly influenced by students' participation in class. Furthermore, belonging and identity also tend to be associated with each other (Walton & Brady, 2017) and both contribute to students' interest in and interest in pursuing a STEM career.

It is worth noting that math may be a domain with which it is particularly difficult for women to identify, due to pervasive stereotypes regarding the inferior ability of women in math compared to men. Math is unique in that people use phrases such as "math anxiety" and "math person" to describe their orientation toward the subject. In contrast, it would be surprising to hear somebody claim that they are "not a social science person." Thus, math is often regarded as a subject that is intended for some people but not others (e.g., Ayalon, 1995). Leslie et al. (2015) surveyed faculty members, postdoctoral fellows, and graduate students across 30 academic fields in order to assess their beliefs about ability within their respective domains. Results demonstrated that the higher the number of academics in a given field who believed that field-specific success depended on innate talent, the more male-dominated the discipline was likely to be. Notably, among STEM domains, math was rated the highest with respect to this emphasis on innate brilliance. Results also suggested that the dearth of women in fields with a particular emphasis on innateness could be explained by a belief that women (as opposed to men) are generally less capable of becoming prominent scholars. These peculiarities perhaps support a view of math as a rather unwelcoming domain, encouraging both low identity and belonging, particularly for women.

Sense of Belonging

A critical determinant of whether a student will thrive in STEM fields is their sense of belonging (i.e., membership and acceptance) in math-related contexts (Rainey et al., 2018). Extant research has demonstrated the importance of a sense of belonging within a given environment in increasing intrinsic motivation and bolstering persistence on specific tasks in that context (Walton & Brady, 2017).

In a longitudinal study of college calculus students, Good et al. (2012) determined that stereotypes and messaging regarding women's lack of math ability compared to men's contributed to a deterioration of women's sense of belonging in math, impairing their desire to pursue math in the future. Work by Murphy et al. (2007) demonstrated similar results with respect to how situational cues about gender influence women's sense of belonging. In a factorial design experiment (gender by situational cue), undergraduate STEM majors were shown an advertisement for a fictional STEM leadership conference. Half of the participants saw a video with people at the conference in a ratio of 3 men to 1 woman, whereas the other half saw a ratio of 1 man to 1 woman. Analyses of participants' anticipated sense of belonging at the conference unveiled a significant interaction between gender and situational cue, such that women who watched the gender-unbalanced video reported a significantly lower sense of belonging than women who watched the balanced version of the advertisement. On the other hand, the manipulation had no effect on male participants.

In another series of studies by Cheryan et al. (2009), undergraduate women were asked to imagine that they were deciding between two employment opportunities. Results showed that when the company's office environment was described as aligning with the masculine stereotype of computer science, participants reported a significantly lower sense of anticipated belonging than when the office had a more neutral environment. Further, this lower sense of belonging predicted reduced interest in pursuing a computer science career.

Belonging and Participation

As mentioned briefly in the introduction, classroom participation is one mechanism that is essential to fostering students' sense of belonging (Christenson et al., 2001; Goldstein & Benassi, 1994). Thus, active classroom participation may be one of the ways in which stereotype threat influences students' sense of belonging in the math domain.

Evidence suggests the potential existence of bidirectional relations between students' participation and sense of belonging. At the university level, Freeman et al. (2007) found that the extent to which participation was encouraged within the classroom was strongly associated with participants' sense of belonging. In another study of students in grades 3 through 6, Furrer and Skinner (2003) demonstrated that sense of belonging in the classroom appeared to influence students' levels of behavioral engagement, as measured by participation in the classroom. Students exhibiting high levels of belonging showed greater increases in levels of engagement over time, and the opposite relation also held true. The authors call for additional studies examining multiple time points in order to more thoroughly examine the reciprocal relations between classroom engagement and sense of belonging.

Skinner and Belmont (1993) also found evidence for bidirectional effects by studying associations between student engagement and teacher emotional support, which has been recognized as a metric related to classroom climate and student belonging (Allen et al., 2018). The authors identified reciprocal effects between students' sense of support from teachers—a proxy for student belonging—and behavioral engagement, concluding that students who perceive greater emotional support from their teachers generally also have higher behavioral engagement, and students who show higher behavioral engagement, in turn, experience a greater sense of support in the classroom.

It is possible that this reciprocal relation between belonging and participation is propelled by a "recursive process," as described by Yeager and Walton (2011), who explain how small social-psychology interventions in education often have surprisingly long-lasting effects (e.g., Cohen et al., 2009). The authors attribute these effects—which they say can sometimes seem "magical," sustaining for months or years after the intervention takes place—in part to the fact that the interventions often tap into recursive, self-propelling processes already existing in educational environments (Yeager & Walton, 2011, p. 268). They argue that by tapping into these recursive processes, social psychological interventions might be somewhat self-sustaining.

For instance, they offer the example that as students develop a greater sense of belonging in school and form meaningful relationships with others, these sources of support (e.g., peers and teachers) can further encourage feelings of belonging that persist across time.

With respect to the present studies, perhaps high levels of student belonging have initial effects on frequency of participation (as indicated with the dashed arrows in Figure 1), which, in turn (assuming that the environment is one in which students are supported even after giving incorrect answers), further boosts sense of belonging, in a positive feedback loop with accumulating effects. If the relations between belonging and participation indeed follow a reciprocal pattern, these recursive processes could be capitalized upon in designing interventions to boost female belonging in STEM domains.

Math Identity

Along with positively influencing feelings of belonging, higher frequency of participation could also bolster a student's level of identification with math—i.e., their perceptions of themselves as the type of person who does math (Anderson, 2007). Math identity, like sense of belonging in math contexts, is a vital predictor of whether a student intends to pursue a STEM career (Cribbs et al., 2020; Lock et al., 2013).

Extant literature has established that female students tend to have a more fragile sense of math identity than male students (e.g., Lock et al., 2013). In a study of implicit association by Nosek et al. (2009), results showed that among 34 countries, endorsement of stereotypes associating STEM with males (and not females) was evidenced in 70% of the more than half a million implicit association tests administered. Another study by Nosek et al. (2002) addressed how these stereotypes relate to students' identity. Their study of implicit measures of gender identity and gender-math stereotypes with college students demonstrated that male participants displayed a stronger identification with science and math-related concepts than did female participants, and both men and women endorsed stereotypes of math as a male domain (Nosek et al., 2002). Overall, this work suggests that internalization of cultural stereotypes can influence students' math identity.

Identity and Participation

Research has demonstrated that meaningful participation in classroom discourse—via challenging existing ideas and sharing knowledge and perspectives—can also be an important factor in augmenting students' math identities (Walshaw & Anthony, 2008). It is possible that self-perception effects are responsible for mediating the process of one's participation shaping their identity. Self-perception theory argues that people's attitudes develop as a result of examining and reflecting on their own behaviors (Bem, 1972). For example, observing one's own lack of behavioral engagement in a math class could lead a person to conclude that math is not a central part of their identity and that they should therefore not pursue further opportunities in the domain. In contrast, monitoring their own frequent participation could prompt them to make a conclusion such as: "I am the kind of person who raises their hand to engage in conversations in math contexts; therefore, I am a math person."

Another potential explanation is cognitive dissonance theory (Zanna & Cooper, 1974), which could apply in the case that a student already holds strong beliefs about their math identity, e.g., that they are not a math person. If they raise their hand to participate in math class despite this belief, the discrepancy between their beliefs and actions may lead them to readjust their level of math identity to align with their behavior. While the present work will not explore which mechanism is in fact occurring, both potential mechanisms support the general idea that participation could influence math identity. It is also important to note that like belonging, identity may also have a bidirectional relationship with participation, such that a student's domain-specific identity is also an important predictor of their participation.

Interest in Pursuing a STEM Career

As mentioned previously, a primary source of motivation for this dissertation research is data showing that women are underrepresented in math-related majors, earn fewer STEM degrees than men, and enter STEM occupations at a significantly lower rate than men (National Science Foundation, 2019). Research suggests that both sense of belonging in math contexts and identification with the math domain may contribute to persistence in STEM domains and bolster students' interest in pursuing a STEM career (e.g., Belanger et al., 2020).

For instance, Thoman et al. (2014) recruited female undergraduate STEM majors and surveyed them every two weeks throughout a semester regarding their experiences of belonging and interest in their STEM course and a humanities course. Results suggested that a low sense of belonging can cause women to feel pushed out of STEM contexts, reducing their interest in the domain and also potentially prompting them to feel *pulled into* other domains due to a competing sense of belonging. This study highlighted how a lack of belonging can trigger female students to seek opportunities that are external to STEM fields.

Smith et al. (2013) also explored how sense of belonging influences women's interest in STEM domains. Participants were directed to read a brochure describing a graduate program in a fictional domain named "eco-psychology." In one condition, the brochure contained all male faculty names, and the majority of program photos were of men. In another condition, both the faculty names and the photographs in the brochure were evenly split in terms of gender. After reviewing the brochures, participants rated their interest in pursuing eco-psychology. The interest of female participants in the gender-equal condition did not differ from that of male participants' in either condition; however, women shown the male-dominated brochure expressed significantly lower interest in pursuing eco-psychology compared to participants in all other conditions.

Math identity can be equally important to belonging in influencing students' career interest. Cribbs et al. (2020) utilized data from a national US survey of over 10,000 university calculus students to determine whether math identity predicted students' interest in STEM careers. The results were staggering: a one standard deviation increase in math identity corresponded with a 4.2-times-higher chance of having STEM career intentions. Similarly, in a survey of first-year college students, Lock et al. (2013) found that the gender gap in intention to pursue a physics career was eliminated when controlling for gender differences in math and physics identity.

A series of three studies by Stout et al. (2011) demonstrated that being presented with ingroup role models (e.g., female professors, professionals advanced peers, and other experts) both promoted a greater sense of belonging and bolstered math identity in female undergraduate STEM majors. Further, this increased sense of belonging and identification with the domain, in turn, encouraged heightened commitment to pursuing STEM careers. More specifically, math identity mediated the relation between participants' identification with the female experts and their interest in continuing in STEM after graduation.

Altogether, the results of this research suggest that a lack of belonging and/or low math identity could negatively influence a student's desire and interest in pursuing a STEM career, whereas increased belonging and identity may be predictors of students' persistence in the domain.

The Present Dissertation

Gender disparities in the math domain have been examined from a variety of research perspectives, including through examinations of how differences in belonging and identity might contribute to disparities in career pursuit. Via two studies—one cross-sectional, and a second employing a daily diary methodology—the present dissertation aims to contribute to this literature by focusing on antecedents of these disparities. The investigation is motivated by the possibility that differences in math classroom participation, a measure of engagement that is potentially prone to gender differences, could be partially responsible for these later, more pronounced—and possibly more detrimental—gender inequalities in STEM domains.

With the aim of exploring reasons for gender differences in STEM career interest, the present research employs the theoretical model in Figure 1 in order to examine whether female students in fact demonstrate less frequent participation in their math classes and what mechanisms might be underlying this difference. Specifically, the model accounts for the possibility that women exhibit lower frequencies of participation in math classrooms because gender-related stereotype threat induces a prevention focus, which leads them to be more cautious about potentially giving an incorrect answer, thus increasing the threshold for classroom participation and reducing their participation frequency, which eventually impacts their sense of belonging and math identity.

Academic Domains

One important note is that although the focus of the present investigation is on participation in math classes, the pilot and dissertation studies also include items related to contrasting academic domains as points of comparison. Specifically, social studies and psychology served as domains of comparison for math in the pilot studies, and a specific social sciences course named by each participant was the domain of comparison in both dissertation studies. The purpose of including a contrasting domain was in order to be able to probe the math-specificity of the hypothesized mechanisms. In other words, it allowed me to investigate whether lower levels of participation and higher confidence thresholds are unique to women in math contexts or if gender differences exist in non-stereotyped domains as well.

Although a number of studies examining gender bias in the math domain use English as a point of comparison, I selected different domains for a few reasons. First, there are some existing gender stereotypes in the English domain that potentially reduce its utility as a control, since it is often viewed as a subject more suited to female than male students. Second, questions posed to students in English classes are often not presented with "correct" answers in mind; thus, confidence thresholds may not be particularly meaningful in that domain. Since this investigation hinges on students' perceptions of how "sure" of a correct answer they must be, it was critical to employ contrasting domains—such as social sciences—that, like math, generally require students to share factual information aligned with a "correct" answer (i.e., may involve confidence thresholds) and, unlike math, are not generally affected by gender stereotypes.

Forms of Participation

It is also important to note that the present dissertation examines a specific type of classroom participation: students' attempts to answer instructor-initiated questions (e.g., raising their hand to answer a question) or to contribute in another way that carries a risk of making a mistake (e.g., sharing a response or volunteering to provide an explanation without instructor prompting). These forms of participation might be particularly related to stereotype threat effects, especially in math contexts, where there are typically right and wrong answers and students could be judged for answering a question incorrectly. Because female students are especially prone to stereotype threat in the math domain (Spencer et al., 1999), participation of this nature may be more reflective of gender disparities than other, more subjectively evaluated forms of participation.

Furthermore, participation in these studies is operationalized as raising one's hand to participate, regardless of being selected to respond or not. Based on emerging research on classroom participation (e.g., Böheim et al., 2020b), hand-raising is a convenient proxy for active participation that may reduce measurement error. It also represents an indication of when students' confidence has surpassed their participation confidence threshold, prompting tangible, measurable action.

Research Aims

With the overarching goal of exploring the theoretical model proposed in Figure 1, the present dissertation addresses the following research aims in a set of two studies.

Research Aim 1

The first aim of this dissertation was to investigate whether there were gender differences in the mean levels of all constructs in the proposed model. More specifically, I examined whether gender differences existed in (1) class participation behavior, particularly in the math domain; (2) potential predictors of differences in participation, including stereotype threat, regulatory focus, and confidence thresholds; and (3) potential outcomes of these differences, including sense of belonging, identity, and career interest.

I expected to find significant differences across gender and domain. More specifically, within math but not necessarily social science courses, I expected female students to show higher levels of stereotype threat susceptibility, tendencies toward a prevention focus, and class participation confidence thresholds (such that female undergraduate students would report requiring higher levels of confidence in the validity of their contribution before being willing to raise their hand). On the other hand, I hypothesized lower frequencies of participation, sense of belonging, identity, career interest, confidence, and value for women in math, and no differences in social science. In terms of perceived opportunities for participation, I expect to find differences across domain (such that all participants report a greater number of opportunities to participate in social science compared to math), and no difference across gender.

Research Aim 2

The second aim of this research was to explore relations between constructs in the proposed theoretical model, both by examining bivariate correlations and path analyses replicating the proposed theoretical model. In both domains, I expected to find that variables that are proximal in the model in Figure 1 were significantly associated. I also expected to find significant (but weaker) correlations between indirectly related (more distal) variables.

Specifically, I hypothesized that high stereotype threat would be correlated with a prevention focus, and that a prevention focus would be associated with a higher confidence threshold, which in turn would be associated with less frequent classroom participation. Furthermore, lower participation frequency was expected to be related to a lower sense of belonging, lower domain identity, and lower career interest within a domain. In addition to exploring correlations among constructs, I planned to examine the potential directionality of these relations using multiple regression and path analyses. While there was a relatively strong theoretical argument for the directionality of relations between some constructs in the model, it was difficult to make causal claims with respect to other associations. For example, given that the literature provides support for bidirectional associations between participation and sense of belonging (Skinner & Belmont, 1993), an investigation of the nature of this association was necessary in order to investigate claims about causality.

Finally, I aimed to examine the unique contributions of specific predictors when controlling for other variables in the model. In particular, I expected class participation confidence threshold—a novel construct introduced in this dissertation—to emerge as a unique predictor of variability in participation behaviors, even when controlling for other motivational variables that have traditionally been examined as key motivational predictors of students' behaviors, such as confidence and value in the domain.

Research Aim 3

After examining associations between constructs in the theoretical model within Research Aim 2, the third aim of this dissertation was to explore potential mediating mechanisms underlying these relations. In general, I expected to find support for the indirect relations in the model in Figure 1 (indicated by some of the dashed arrows). In particular, when examining potential predictors of classroom participation, I expected the relation between regulatory focus and classroom participation to be mediated by confidence thresholds.

CHAPTER 3: PILOT STUDIES

Four preliminary studies informed the present work by examining various parts of the theoretical model displayed in Figure 1. Although these pilot studies were designed to probe a range of research questions (some of which are unrelated to the present dissertation), they shared the common goal of exploring associations between gender, stereotype threat, regulatory focus, participation confidence thresholds, and frequency of participation in math classes. Further, these pilot studies were also critically important in the process of measure selection and refinement for the dissertation studies. In this chapter, I focus only on findings that are particularly relevant to the present dissertation. See Chapter 4 for detailed descriptions of the final versions of the measures used in the dissertation studies, which slightly varied from the measures included in each of the pilot studies.

Pilot Study 1: Associational Survey with Middle School Participants

Participants

After exclusions, the final sample of Pilot Study 1 included 150 presumptive seventh- and eighth-grade students from across the United States (50.7% female; 84.0% White; mean age 13.7). Participants were recruited via email invitations to parents through a panel recruitment service.[1]

Methods

Procedure

The study was administered online as a 15-to-20-minute survey about motivation and participation behaviors in math classes. In addition, participants completed analogous

[1] Note that because participants were recruited for money via messages to parents, I was unable to confirm that all participants were, in fact, middle school students.

questionnaires regarding a contrasting domain—social studies—as a point of comparison. The survey included two counterbalanced blocks of items: one with math-related questions and another with social studies-related questions. Within each of the two blocks, scales were presented in a fixed order, although within each scale, items were randomized.

Measures

See Table A1 for a list of measures (including sample items) used in each of the pilot studies. Unless otherwise specified, participants were asked to respond to the items in each measure using Likert-type scales that included between five and six scale points.

Specifically, *stereotype threat susceptibility* was measured with a series of six items adapted from Marx and Goff (2005)—three items regarding the math domain and an analogous series of three items regarding the social studies domain. *Regulatory focus orientation* was operationalized as tendency toward a prevention orientation and was assessed with a 6-item scale adapted from an identity-specific regulatory focus scale by Browman et al. (2017). *Self-reported participation frequency* was assessed with two items intended to gauge how often students participated, as well as their frequency of deciding *not* to participate, both in response to a question posed by an instructor. *Domain-specific self-efficacy* (both math and social studies) was measured with six items from the Academic Self-Description Questionnaire (Marsh, 1990), modified for administration to children (e.g., "Work in math classes is easy for me").

Class participation confidence threshold (CPCT-1) was assessed by presenting students with a vignette description (related to an instructor asking a question in math and social studies, separately) and prompting them to mark their response on a slider. Participants were then presented with a continuous slider from 0 to 100%, with answer options of whole percentages, on which they declared the level of confidence they would require in their response before raising their hand to participate.

Relevant Results

See Tables B1 and B2 for correlations between the study variables of interest.

Associations between Regulatory Focus, Confidence Thresholds, and Participation Behaviors

Across all pilot studies, regulatory focus was operationalized as the mean of the items assessing one's prevention focus minus the mean of the items assessing one's promotion focus, such that positive scores indicate that one is primarily prevention-focused. As expected, in Pilot Study 1, regulatory focus was significantly associated with self-reported participation frequency in both domains, such that the more predominantly prevention-focused participants were, the less often they raised their hands to answer the instructor's questions and the more often they changed their mind after deciding to raise their hand.

Contrary to my expectations, regulatory focus orientation was *not* correlated with the initial measure assessing participants' class participation confidence threshold (CPCT-1) in either domain. In addition, CPCT-1 was not significantly associated with self-reported participation behaviors in either domain (p = .061-.671).

Gender and Stereotype Threat Susceptibility

To examine differences in stereotype threat susceptibility, I conducted a 2 (Gender: male vs. female) x 2 (Domain: math vs. social studies) mixed ANOVA, with Domain as the within-participant factor. Although there was not a significant main effect of gender, $F(1, 148) = .80$, $p = .374$, there was a significant main effect of domain, $F(1, 148) = 11.70$, $p = .001$, which was qualified by a significant interaction between gender and domain, $F(1, 148) = 4.14$, $p = .044$. Pairwise comparisons revealed that when comparing stereotype threat susceptibility between domains within each gender, females reported significantly higher levels of threat susceptibility in math ($M = 2.11$, $SD = .13$) than in social studies ($M = 1.70$, $SD = .11$), $p < .001$. This difference was not present for males, (math: $M = 1.82$, $SD = .13$; social studies: $M = 1.72$, $SD = .11$), $p = .332$. In line with prior work on stereotype threat, these results suggest that women (but not men) are more prone to experiencing stereotype threat in the math domain than in a non-STEM domain such as social studies. Additional comparisons that examined stereotype threat susceptibility between genders within each domain demonstrated no significant gender differences ($p = .116$ for math; $p = .901$ for social studies).

Stereotype Threat

Additionally, in the math domain, stereotype threat susceptibility was positively correlated with regulatory focus, such that being more predominantly prevention-focused was associated with higher levels of stereotype threat. In contrast, there was no association between stereotype threat susceptibility and regulatory focus in the social studies domain. These results are consistent with prior work suggesting that stereotype threat can activate a prevention focus (Oyserman et al., 2007; Seibt & Förster, 2004). In both domains, stereotype threat was also highly correlated with frequency of changing one's mind after having decided to participate.

Pilot Study 2: Associational Survey with Undergraduate Participants

Participants

The second pilot study was administered to a final sample of 171 undergraduate students (87.7% female; 69.6% White) enrolled in introductory developmental psychology courses in exchange for course credit.

Methods

The survey-based study was conducted in a university laboratory setting. As in Pilot Study 1, participants responded to questionnaires about their motivation and hypothetical participation behaviors in math class, as well as in a contrasting domain (psychology). The structure of the survey was similar to that of Pilot Study 1, except for some additional measures. As in Pilot Study 1, the order of the subject matter blocks (i.e., math- versus psychology-related questions) was counterbalanced across participants. See Table A1 for a list of measures included in the study.

Due to the homogeneity of the sample, there was not sufficient power to examine potential gender differences in the present study (though I did conduct some exploratory gender analyses, which are not reported here).

Relevant Results

Tables B3 and B4 present bivariate correlations between the study variables of interest.

Associations between Regulatory Focus, Confidence Thresholds, and Participation Behaviors

Consistent with my expectations, and with the results of Pilot Study 1, regulatory focus was significantly associated with self-reported participation frequency in both domains, such that the more predominantly prevention-focused participants were, the less often they raised their hands to answer the instructor's questions. In the math domain, but not in the psychology domain, predominantly prevention-focused participants were more likely to change their mind after deciding to raise their hand.

Once again, CPCT-1 was not associated with regulatory focus in the math domain; though, in the psychology domain, the correlation between these two variables was significant. Importantly, and in contrast to the results for CPCT-1, regulatory focus did predict participants' confidence threshold in *both* domains when it was assessed with the newly introduced Likert-type measure (CPCT-2).

In contrast to the previous pilot study, both confidence threshold measures were significantly associated with each measure of participation frequency. Specifically, the higher participants' threshold was for voluntarily answering the instructor's questions, the less likely they were to voluntarily answer the instructor's questions and the more likely they were to change their minds about participating. Note that the correlations between confidence thresholds and self-reported participation were much stronger for the CPCT-2 (vs. CPCT-1) measure.

Finally, I calculated a difference score between the participation-specific confidence slider measure and CPCT-1 for each participant. In this pilot study, participation-specific confidence was measured as a percentage in order to align with the format of the CPCT-1 measure (see Table A1 for more information). Thus, the difference score is a measure of the percentage of confidence a person typically has in knowing the correct answer in a domain minus their confidence threshold, or how confident they need to be in order to participate. In other words, a positive difference score would represent the number of percentage points *over* one's confidence threshold a participant typically is. This confidence-threshold difference score was correlated with frequency of participation, both in the math domain, $r = .559$, $p < .001$, and the psychology domain, $r = .512$, $p < .001$.

Mediation Analyses

As a further test of the model in Figure 1, I also examined whether confidence thresholds mediated the relation between regulatory focus and participation frequency. To do so, I used Hayes's (2018) PROCESS macro (v3.5 for SPSS), which is a bias-corrected bootstrapping procedure that randomly selected 5,000 samples with replacement from the data set, estimated regression coefficients for each of the bootstrap samples, and averaged them across all samples.

Results in the math domain demonstrated an indirect effect of regulatory focus on frequency of participation via CPCT-2 (b = -0.23, 95% CI: [-0.35; -0.11]). Similarly, in the psychology domain, results confirmed an indirect effect of regulatory focus on frequency of participation via CPCT-2 (b = -0.20, 95% CI: [-0.33; -0.07]). Thus, in both domains, class participation confidence threshold explained a significant portion of the association between regulatory focus and frequency of participation.

Examining the Unique Effects of Confidence Thresholds on Participation Frequency

To assess whether participants' class participation confidence thresholds predicted the self-reported frequency of their participation when controlling for other motivational variables that are thought to predict classroom participation, I regressed frequency of participation in math contexts onto confidence threshold in math (CPCT-2), math self-confidence, participation confidence in math classes, and three aspects of math value (importance, intrinsic value, and utility value) by entering each variable separately in a simultaneous regression model. The results showed that CPCT-2 remained a significant predictor of participation frequency in the math domain even when controlling for these additional motivational variables, ($\beta = -.58$, $t(169) = -9.74$, $p < .001$).

Thus, focusing solely on confidence and value when attempting to predict participation behavior provides an incomplete picture from a motivational perspective. It appears that confidence thresholds represent a unique, significant predictor of undergraduate students' classroom participation behaviors.

Stereotype Threat

In contrast to the results from Pilot Study 1, I did not find any significant associations between stereotype threat susceptibility and regulatory focus orientation in the math domain or in the psychology domain. In response to this unexpected outcome, I revised the stereotype threat measures for use in Pilot Study 4 (described in a section below), with the intention of including a more subtle measure of stereotype threat that captures participants' affective experience rather than their awareness of gender stereotypes.

It is important to note that the vast majority of participants were majoring in a non-STEM domain (91.8%). Thus, it is possible that, regardless of their gender, most participants did not identify with the math domain. Since stereotype threat is more prominent for individuals who are highly identified with a given domain (Schmader et al., 2008; Steele, 1997), it is possible that the sample's overall lack of identification with math influenced the results of this study.

Pilot Study 3: Experimental Study with Undergraduate Participants

Participants

Pilot Study 3 was administered to undergraduate students enrolled in introductory developmental psychology courses in exchange for course credit. It included a final sample of 124 students (87.1% female; 71.8% White). Due to the nature of the research participation requirement at the university, 56 participants completed both Pilot Studies 2 and 3. However, as explained in the *Limitations* section below, this allowed me to examine the temporal stability of key variables.

Methods

Although Pilot Study 3 included many of the same variables of interest as in Pilot Studies 1 and 2, it was designed to be an experiment that temporarily manipulated students' regulatory focus. The study was originally intended to take place in a university laboratory setting, but it was ultimately administered online due to the COVID-19 pandemic. Accordingly, I included a number of attention check items.

For the manipulation, participants were randomly assigned to either a prevention or promotion condition and asked to write about their academic duties and obligations (prevention) or their academic hopes and aspirations (promotion) (see Molden et al., 2008). They completed both a "general induction" and an "academic-specific induction" formulated based on work by Freitas and Higgins (2002) and Higgins et al. (1994) (see Appendix A).

After the inductions, participants responded to questionnaires similar to the ones administered in Pilot Studies 1 and 2. One primary difference was that questions focused *only* on the math domain. Measures that were theoretically most likely to be influenced by the induction were asked closer to the beginning of the survey, followed by manipulation check items (measuring state regulatory focus, math-specific regulatory focus, and chronic regulatory focus), as well as demographics and attention check items. Importantly, Pilot Study 3 was the first to include the distal outcome measures of the model (i.e., domain-specific identity, sense of belonging, and career interest). See Table A1 for a list of all measures included in the study.

Relevant Results

Manipulation Checks

Neither chronic regulatory focus nor math-specific regulatory focus were significantly influenced by the manipulation (p = .185-.985). In addition, state regulatory focus only differed between conditions for the promotion subscale, $t(122) = 2.15$, $p = .033$, and not the prevention subscale, $t(122) = .82$, $p = .42$.

The manipulation also did not appear to significantly influence any of the dependent variables. Thus, for the purposes of these analyses of this preliminary study, I have collapsed across experimental conditions. See Table B5 for a summary of correlational results.

Associations between Regulatory Focus, Confidence Thresholds, and Participation Behaviors

Perhaps due to my attempts to manipulate participants' regulatory focus, some results of Pilot Study 3 do not align with those of the prior two studies. Unlike in Pilot Studies 1 and 2, regulatory focus was not associated with self-reported participation frequency—the effect was in the expected direction but was only marginally significant. However, as in each of the prior two pilot studies, regulatory focus was significantly correlated with participants' decisions not to participate, such that the more predominantly prevention-focused participants were, the more often they changed their mind after initially deciding to raise their hand.

In alignment with each of the previous pilot studies, CPCT-1 was only marginally associated with math-specific regulatory focus. However, when the Likert-type measure of threshold (CPCT-2) was utilized, a prevention focus was associated with a higher math confidence threshold.

As in Pilot Study 2, both confidence threshold measures were significantly associated with each measure of participation frequency. More specifically, the higher participants' thresholds were, the less likely they were to voluntarily answer the instructor's questions and the more likely they were to change their minds about participating.

Finally, I again calculated a difference score between the participation-specific confidence slider measure and CPCT-1 for each participant. The difference score, which represents how many percentage points of confidence participants typically have *over* their math confidence threshold, was highly correlated with frequency of participation in the math domain, $r = .670$, $p < .001$.

Mediation Analyses

As in the prior pilot studies, I used Hayes's (2018) PROCESS macro (v3.5 for SPSS) in order to examine whether confidence thresholds mediated the relation between regulatory focus and participation frequency. Results suggested an indirect effect of regulatory focus on frequency of participation via CPCT-2 ($b = -0.12$, 95% CI: [-0.22; -0.04]).

Analyses of Newly Included Outcome Measures

Because Pilot Study 3 was the first to incorporate the outcome measures in Figure 1 (i.e., outcomes potentially resulting from participation behaviors), I conducted a series of analyses to probe their relations with each other and other study variables. First, I examined correlations between the two participation behavior measures (frequency and frequency of changing one's mind), confidence threshold, regulatory focus, and all three distal measures. Importantly, I found that sense of belonging was strongly correlated with all other measures of interest, which is consistent with research that has found domain-specific belonging to be a critical predictor of domain-specific outcomes (e.g., Walton & Brady, 2017). Math identity and math-related career interest were also both significantly associated with participation frequency and regulatory focus. Finally, among the outcome measures themselves, they were all rather highly correlated with each other (see Table B5).

It is important to note that as with all of the correlational results reported here, I cannot make strong claims about directionality. While it is possible that participation influenced sense of belonging and math identity in the direction posited by our model, there is also a theoretical argument for the possibility that students who feel a greater sense of belonging and/or have higher math identity are more likely to participate. The daily diaries format of Study 2 of the dissertation was designed to probe this directionality further.

As a preliminary means of testing potential directionality, I conducted some mediation analyses. I first tested for a potential mediation effect of confidence thresholds in the relation between regulatory focus and sense of belonging. Results demonstrated an indirect effect of regulatory focus on sense of belonging via CPCT-2 (b = -0.04, 95% CI: [-0.09; -0.006]), such that class participation confidence threshold explained a significant portion of the association between regulatory focus and sense of belonging. I also tested a sequential mediation by adding an additional mediator to the model, examining potential mediation effects of both confidence threshold and participation frequency in the relation between regulatory focus and sense of belonging. Results showed a small indirect effect of regulatory focus on sense of belonging via two mediators: CPCT-2 and participation frequency (b = -0.06, 95% CI: [-0.11; -0.02]).

In reviewing all results of Pilot Study 3, it is critical to keep in mind is that I attempted to manipulate regulatory focus in this pilot study. Thus, while it is interesting that regulatory focus was associated with all distal outcome measures, further investigation is required in order to draw any stronger conclusions. At the very least, these preliminary results indicate that this is a promising avenue for further exploration.

Pilot Study 4: Associational Survey with Undergraduate Participants

Pilot Study 4 was conducted with two separate samples of undergraduate students at similar northeastern universities. Because the study was administered identically to both samples, for the sake of brevity, here I present the results of the analyses pertaining to the two samples together.

Participants

Similar to Pilot Studies 2 and 3, Pilot Study 4 was administered to undergraduate students in exchange for course credit. After exclusions, the final sample included 219 participants (105 from University A and 114 from University B). The sample was mostly female (78.5%) and White (74.4%). Post-hoc analyses revealed that two participants at University A had also completed Pilot Study 2, and another 8 had participated in Pilot Study 3.

Methods

Pilot Study 4 was administered online in a survey format to both samples. The structure of the survey was very similar to that of Pilot Study 2, although several measures were updated, and some additional measures were added. Table A1 presents a summary of all measures included in the study.

Changes from Prior Pilot Studies

Pilot Study 4 incorporated some important changes based on results of the prior pilot studies. *Stereotype threat susceptibility* was updated to include two new measures: one Likert-type measure based on a scenario-based prompt intended to measure stereotype threat that requires less conscious awareness, and another measure by Deemer et al. (2016), the Stereotype Threat in Math scale. My goal with introducing the scenario-based measure was to include measures that would capture experienced threat and not necessarily perceptions of fairness or awareness of gender stereotypes. The scale by Deemer et al. (2016) was selected because of its seven-item length, compared to the three items offered by Marx and Goff (2005). The scenario-based measure was operationalized as a difference score of self-reported threat in the opposite-gender scenario minus self-reported threat in the same-gender scenario.

Participation behaviors were expanded to include more items than the two present in Pilot Studies 1 through 3. Rather than asking only about a scenario in which an instructor poses a question to the class, additional items were included to ask about a wider variety of participation behaviors. This more "general" set of participation behaviors was also reflected in new measures of *class participation confidence threshold (CPCT)* and *participation-specific confidence* (in addition to the "instructor-prompted participation" versions remaining from prior pilot studies). Finally, participation-specific confidence was changed from a slider to a Likert-type measure to align with the CPCT-2 measure of confidence threshold rather than CPCT-1 (given the stronger associations with CPCT-2 in results of the previous pilot studies).

Relevant Results

Tables B6 and B7 present bivariate correlations between the study variables of interest in Pilot Study 4.

Between-Domain Differences. First, I explored between-domain differences by conducting a series of t-tests. Consistent with my expectations, participants reported a higher sense of stereotype threat (as measured by the scale by Deemer et al., 2016) in the math domain overall compared to the psychology domain, $t(218) = 7.20$, $p < .001$. Participants also demonstrated higher prevention focus, $t(218) = 9.23$, $p < .001$, and confidence thresholds, $t(218) = 3.91$, $p < .001$ (CPCT-1), $t(218) = 2.99$, $p = .003$ (CPCT-2), in the math domain than in the psychology domain.

Associations between Regulatory Focus, Confidence Thresholds, and Participation Behaviors. Consistent with the results of Pilot Studies 1 and 2, regulatory focus in both the math and psychology domains was significantly associated with self-reported participation frequency, such that the more predominantly prevention-focused participants were, the less often they raised their hands to answer the instructor's questions, and the less often they volunteered to provide an explanation or make a comment.

For the first time among all pilot studies, CPCT-1 was associated with regulatory focus in the math domain, such that a prevention focus was correlated with a higher confidence threshold; however, this relation was not present in the psychology domain. Consistent with the results of Pilot Studies 2 and 3, the Likert-type measure of threshold (CPCT-2) was positively associated with a prevention focus in both domains.

As in the two prior pilot studies, both class participation confidence threshold measures were significantly associated with each measure of participation frequency, such that the higher the participants' thresholds, the less likely they were to answer the instructor's questions, make a comment, or volunteer to provide an explanation. CPCT-2 was also associated with both measures of changing one's mind about participation, such that participants with high thresholds were more likely to change their minds about participating in both domains. In the math domain only, CPCT-1 was also associated with the measures related to changing one's mind in the expected direction.

Finally, as with the two prior pilot studies, I calculated a difference score between participation-specific confidence and confidence threshold for each participant. In this study, the measure of participation-specific confidence was designed to align with CPCT-2 instead of CPCT-1 and was therefore updated to be a Likert-type scale rather than a percentage. This confidence-threshold difference score was again significantly associated with frequency of participation, both in the math domain, $r = .608$, $p < .001$, and the psychology domain, $r = .603$, $p < .001$.

Associations with Belonging, Identity, and Career Interest. Pilot Study 4 was the first to include measures of more distal outcomes of the model (sense of belonging in math contexts, math identity, and interest in pursuing a STEM career) for *both* the math and psychology domains. Consistent with the results of Pilot Study 3, I found that sense of belonging was correlated with all other measures of interest. These associations were also present in the psychology domain. Furthermore (and also in line with the results of the prior pilot study), math identity and math-related career interest were also both significantly associated with participation frequency and regulatory focus; in the psychology domain, the association between identification with the psychology domain and frequency of making comments was only marginal, however. Finally, in both domains, all outcome measures were highly correlated with each other. Tables B6 and B7 present a full summary of these correlational results.

Mediation Analyses. Mediation analyses, again using Hayes's (2018) PROCESS macro, explored whether confidence thresholds mediated the relation between regulatory focus and participation frequency. Consistent with the results of the two previous pilot studies, results implied an indirect effect of regulatory focus on frequency of participation via CPCT-2 in the math domain (b = -0.18, 95% CI: [-0.26; -0.10]) and psychology domain (b = -0.19, 95% CI: [-0.30; -0.09]), suggesting that CPCT explained a significant portion of the relation between regulatory focus and participation frequency.

Additional analyses tested for potential mediation effects of both confidence threshold and participation frequency in the relation between regulatory focus and sense of belonging. Consistent with the findings of Pilot Study 3, results demonstrated an indirect effect of regulatory focus on sense of belonging via CPCT-2 in the math domain (b = -0.16, 95% CI: [-0.23; -0.10]), such that class participation confidence threshold explained a significant portion of the association between regulatory focus and sense of belonging, and there was also an indirect effect of regulatory focus on sense of belonging via two mediators: CPCT-2 and participation frequency (*b* = -0.06, 95% CI: [-0.09; -0.03]). These mediation results held in the psychology domain as well, with an indirect effect of regulatory focus on belonging via both CPCT-2 and participation frequency (*b* = -0.02, 95% CI: [-0.04; -0.007]).

Examining Unique Effects, Controlling for Expectancy and Value Measures. Next, I investigated the unique effects of relations between variables, controlling for traditional motivational measures from an expectancy-value theory perspective (Wigfield & Eccles, 2000)—namely, confidence and value. To do so, I regressed frequency of participation in math contexts onto confidence threshold in math (CPCT-2), math self-confidence, participation confidence in math classes, and three components of math value (importance, intrinsic value, and utility value). As in Pilot Study 2, CPCT-2 remained a significant predictor of participation frequency in the math domain, even when controlling for these confidence and value measures, (β = -.44, $t(218)$ = -8.26, $p < .001$).

Noting the significant negative correlation between a prevention focus and math-related career interest, I also decided to examine whether this relation would hold while controlling for confidence and value, which are typically considered important predictors of students' interest in pursuing a STEM career (Wang & Degol, 2013). Interestingly, when controlling for math confidence, participation-specific math confidence, and all three components of math value, regulatory focus remained a significant predictor of STEM career interest (β = -.19, $t(218)$ = -3.58, $p < .001$).

Stereotype Threat Susceptibility. As described above, stereotype threat was assessed using new measures in Pilot Study 4—one scenario-based Likert-type measure and another

Likert-type agreement scale, the Stereotype Threat in Math scale (Deemer et al., 2016). The scenario-based difference-score measure was highly correlated with the measure by Deemer et al. in both domains. Consistent with prior pilot studies, the scenario-based stereotype threat measure was not correlated with any relevant variables. The Deemer et al. measure, on the other hand, was negatively correlated with sense of belonging in both domains and some additional constructs of interest in the psychology domain (see Table B7).

As in the prior pilot studies, the majority of participants in Pilot Study 4 were non-STEM majors (79.9%) and thus might not have been highly identified with the math domain. Although I did not observe strong stereotype threat effects in this study, I hypothesized that the low math identity of the sample influenced this. It was also possible that stereotype threat was not being measured appropriately through retrospective reports. Since neither new measure of stereotype threat appeared to be significantly associated with other measures, it was not clear whether a more explicit measure probing participants beliefs (i.e., the Deemer et al. measure) or one that assesses participants' felt experiences of being in a threatening scenario is preferable. Thus, both measures were carried forward into the dissertation studies.

Assessment of Newly Included Measures

One aim of Pilot Study 4 was to assess the suitability of some newly included measures. As described above, both new measures of stereotype threat were included in the dissertation studies, in order to test them in a more representative sample.

As a reminder, other new measures in Pilot Study 4 included a wider variety of *participation behaviors* in addition to answering instructor-prompted questions (i.e., making a comment or volunteering an explanation, and participating in small groups with peers) and "general" measures of *class participation confidence threshold (CPCT)* and *participation-*

specific confidence (in addition to the "instructor-prompted participation" versions from prior pilot studies).

Bivariate correlations revealed that while a prevention focus was highly correlated with frequency of answering instructor-prompted questions and raising one's hand to make comments or provide explanations in class, participation in small groups specifically was not associated with regulatory focus in the math domain in Pilot Study 4, $r = .037$, $p = .589$. This finding confirmed concerns I originally had with this measure, given the differences in the experiences of participating in a small group versus in front of one's entire class (particularly with respect to the potential for an experience of threat). Thus, the item related to small group participation was removed for the dissertation studies.

I also examined correlations between the "instructor-prompted" and "general" measures of CPCT-1, CPCT-2, and participation-specific confidence in both domains. Tables B8 and B9 display the correlations between these measures. Due to the high associations between each of the two versions of CPCT-1, CPCT-2, and participation-specific confidence, only the more inclusive measures referring to general participation behaviors were included in the dissertation studies.

General Discussion of Preliminary Studies

Taken together, the results of these pilot studies were overall quite consistent, seeming to corroborate the significance of many associations in the proposed theoretical model. First, they suggested that there may be a significant link between regulatory focus and participation behaviors. Specifically, being more prevention-focused (and less promotion-focused) was negatively associated with the self-reported frequency of participating in math class and positively associated with the frequency of changing one's mind about participating across all studies—particularly in the math domain. In addition, being more prevention-focused was, in general, positively associated with participants' confidence thresholds, particularly when this threshold was assessed using Likert-type items (i.e., as with the CPCT-2 measure). Finally, analyses of distal outcomes from Pilot Studies 3 and 4 suggested that they were predicted by key variables in the model, including regulatory focus and participation frequency. Sense of belonging, in particular, was significantly correlated with most other relevant constructs.

Significant findings were scarcer with respect to stereotype threat. Results from Pilot Study 1 provided evidence for its association with regulatory focus, particularly in the math domain. However, the results of Pilot Studies 2-4 were less consistent, perhaps due to the sample's low identification with the math domain.

A Note on Gender-Related Analyses

In Pilot Studies 2-4, it was not possible to explore gender differences due to a lack of sufficient power, given that female participants made up such a large percentage of each sample. It is also possible that regardless of whether there had been adequate power to test these effects, no gender differences would have emerged, given the moderating role of identity in stereotype threat susceptibility (Steele, 1997) and the generally low levels of math identity reported in these samples of primarily education and psychology majors (Pilot Study 3: $M = 3.77$, $SD = .90$; Pilot Study 4: $M = 3.81$, $SD = 0.96$; both on a 6-point scale). With respect to Pilot Study 1, although no significant gender differences in participation emerged, it is critical to keep in mind that this study was conducted with much younger participants than those who are the focus of the present dissertation. Thus, testing gender-related effects within a university sample, and making an effort to recruit as many math-identified women as possible, was a focus of the dissertation.

Limitations

Overall, it is important to note that the pilot studies represent preliminary, exploratory research and, as such, have some important limitations that are worth considering. First, Pilot Study 1 was conducted with an online sample of middle school students, and it is unclear how developmentally appropriate and reliable the measures were for this age group. There was also no way for me to confirm whether the participants were actually middle-schoolers and not their parents, given that the study was conducted entirely online for a financial incentive.

Second, data for Pilot Studies 3 and 4 were collected online during the COVID-19 crisis, and it was difficult to ensure that participants maintained a high level of engagement throughout the study session. In Pilot Study 3, I attempted to address this concern by excluding participants who failed attention checks and other practical metrics (e.g., reasonable survey duration), but I ultimately had to exclude a rather large percentage (26.7%) of the original sample.

Finally, 56 participants completed both Pilot Studies 2 and 3, and 10 participants in Pilot Study 4 had completed one of the two prior studies. Although this could be seen as a limitation, it actually allowed us to assess the temporal stability of our measures. In examining data of the 56 participants who completed both Pilot Studies 2 and 3 (average time between responses = 158.5 days), I found relatively strong correlations across the two time points in our measures of math-specific regulatory focus ($r = .546$, $p < .001$), participation frequency ($r = .645$, $p < .001$), frequency of changing one's mind ($r = .449$, $p = .001$), and both CPCT measures ($r = .543$, $p < .001$ for CPCT-1; $r = .647$, $p < .001$ for CPCT-2). I also conducted a moderation analysis—although this analysis may have been underpowered—and found that none of the significant correlations reported in Pilot Study 3 significantly varied based on prior participation status (i.e., the correlations were similar for participants in Pilot Study 3 who had previously participated in Pilot Study 2 and those who had not).

CHAPTER 4: STUDY 1

Study 1 was an associational, survey-based investigation examining students' participation behaviors reported on the basis of their general experience in a math and social science class in which they were enrolled at the time of data collection. In other words, students' responses reflected reports of their aggregate experiences over multiple class sessions within each domain. The overarching aim of Study 1 was to follow up on the results of the pilot studies with a more representative sample of students who were currently enrolled in math courses. In contrast to the pilot studies, in which some participants may have been reflecting on math classes in which they had been enrolled in the past (in some cases, several years before), the reports of Study 1 participants were intended to be more reflective of their current behavior and motivation.

Method

Participants

Participants were recruited from a selective northeastern university at the beginning of the Spring 2021 semester. Because the primary focus of the study was on students' experiences in a current math course, recruitment was conducted via email invitations to undergraduate math course listservs (including math, computer science, physics, statistics, and mathematically-oriented business and economics courses). A specific list of courses deemed as suitable "math courses" was compiled from course descriptions on the university website. Before sending the recruitment emails, I contacted the instructors of these courses, describing the research, notifying them of the impending recruitment email, and inviting them to alert me if they did not want students in their courses to be contacted. Two instructors (of four math courses, collectively) asked me not to email their classes.

The recruitment email, which was sent to students in 131 math courses, contained information about the purpose of the study, what participation would entail, and financial incentives, along with a link to a brief enrollment form. On the enrollment form, prospective participants consented to participation in the study and provided information such as their gender, major, class year, age, and a list of their current courses (see Appendix C for all items).

Eligibility Requirements

Prospective participants were deemed eligible to participate in this study only if they met several requirements. First, participants had to be enrolled in at least one undergraduate math (or math-related, e.g., physics or computer science) course during the present semester. Second, they were also required to be enrolled in at least one undergraduate social science course during the semester that met the following requirements: was in a domain that (1) typically involves "correct" and "incorrect" answers, and (2) is not generally affected by gender stereotypes (although this assessment was subjective). Some example social science courses were history, psychology, sociology, and religion. The suitability of participants' social science course was determined by me, based on course descriptions on the university website. Third, due to this dissertation's focus on gender-related stereotypes in the math domain, prospective participants were required to identify as male or female. Finally, participants were eligible only if they had not participated in a prior pilot study in this line of work.

Exclusion Process Following Enrollment Form

A total of 594 students responded to the enrollment form. However, 252 of these students did not complete it and were therefore excluded from the study. Most of these students exited the form when asked to enter their course titles and numbers; I suspect that the obstacle of having to look up course numbers (e.g., via a separate browser) may have dissuaded many of these students from continuing to complete the form. Another student was removed for not consenting to participation in the study. Seven additional students were removed from the sample because they indicated they were graduate students and were therefore not enrolled in undergraduate courses. Given that some students were enrolled in multiple math courses and therefore may have received multiple copies of the enrollment email, I reviewed email addresses and student ID numbers to ensure that each response corresponded to a unique participant, leading to one duplicate row being removed from the data. I also compared students' email addresses to those of the participants in related pilot studies; 15 students were identified as having participated in prior versions of this work and were removed from the sample. Finally, I examined the current courses students listed in order to confirm that they met the requirements of currently being enrolled in at least one math and one social science course. I removed five participants who were missing at least one of these courses, yielding a final list of 313 students who would be invited to participate in the study.

Of the 313 students deemed eligible and therefore invited to participate in Study 1, 188 responded to the survey. Nineteen respondents did not finish completing it, and seven respondents incorrectly answered at least one of the attention check items. One additional participant who indicated they identified as nonbinary was removed from the sample. The final sample included 161 participants.

To determine the smallest effect sizes that the study was powered to detect, I conducted a series of sensitivity analyses using G*Power (v3.1.9.6; Faul et al., 2007). Two of the analyses indicated that, for the gender x domain ANOVAs reported below, the study had 80% power to detect interactions and within-participant main effects with an effect size of $f = .111$, with alpha set at .05, and assuming the default correlation among repeated measures of .5. A third analysis indicated that, for the same ANOVAs, the study had 80% power to detect between-participant main effects of $f = .192$. And, a final analysis indicated that, for uncorrected zero-order correlations, the study had 80% power to detect an association of $r = \pm.219$, with alpha set at .05 (two-tailed). I did not conduct power analyses for the other types of analyses reported below.

Description of Participants

Of the 161 participants in the final sample, the majority were female (67.1%), White (63.4%), and STEM majors (52.2%). Forty-three participants (26.7%) were math majors, of which 25 (58.1%) were female. STEM majors included all math majors, plus non-mathematical STEM majors (e.g., biology and environmental science). If students listed more than one academic major, they were counted as math majors or STEM majors as long as one of their majors met this requirement. In terms of university year, the highest percentage of participants were freshmen (36.6%), and the lowest percentage were seniors (15.5%). The majority of students (82.0%) reported having at least one parent who had earned a bachelor's degree, and 47.8% had at least one parent with a master's degree or higher. The demographic characteristics of the final sample of Study 1 are summarized in Table 1.

Table 1

Summary of Demographic Information for the Final Sample of Study 1

	All participants (n = 161)	Female only (n = 108)	Male only (n = 53)
	Mean (SD) or N (%)		
Major			
Math major	43 (26.7%)	25 (23.1%)	18 (34.0%)
STEM major	84 (52.2%)	57 (52.8%)	27 (50.9%)
Race			
White	102 (63.4%)	66 (61.1%)	36 (67.9%)
Black or African American	8 (5.0%)	5 (4.6%)	3 (5.7%)
Asian	45 (28.0%)	33 (30.6%)	12 (22.6%)
Mixed race and other	6 (3.7%)	4 (3.7%)	2 (3.8%)
Hispanic	15 (9.3%)	10 (9.3%)	5 (9.4%)
Age (in years)	20.22 (1.25)	20.25 (1.32)	20.15 (1.10)
Class year			
Freshman	59 (36.6%)	40 (37.0%)	19 (35.8%)
Sophomore	30 (18.6%)	17 (15.7%)	13 (24.5%)
Junior	47 (29.2%)	32 (29.6%)	15 (28.3%)
Senior	25 (15.5%)	19 (17.6%)	6 (11.3%)
Highest parent education level			
Some high school	3 (1.9%)	3 (2.8%)	0 (0.0%)
High school	21 (13.0%)	11 (10.2%)	10 (18.9%)
Associate's degree	5 (3.1%)	4 (3.7%)	1 (1.9%)
Bachelor's degree	55 (34.2%)	41 (38.0%)	14 (26.4%)
Master's degree	45 (28.0%)	29 (26.9%)	16 (30.2%)
Med school, law school, PhD	32 (19.9%)	20 (18.5%)	12 (22.6%)

Note. STEM majors include all math majors, mathematical science majors, and non-mathematical STEM majors such as biology and environmental science.

Procedure

Study 1 consisted of three components: (1) a brief enrollment form distributed at the start of the semester; (2) a 15-20 minute survey sent to eligible participants approximately two months into the semester; and (3) a brief (5 minute) follow-up survey sent to participants in the final week of the semester (although this survey was intended to address research questions that were separate from this dissertation). All components of the study were conducted via Qualtrics.

Enrollment Form

As described above, prospective participants first received a recruitment email containing information about the study and a link to a short enrollment form within the first two weeks of the spring semester. The enrollment survey first presented the study's consent form. If students consented to participation, they were directed to a second page asking for their name, email address, student ID number, gender, major, and class year. Students were also asked to affirm whether they were 18 years or older and whether they had participated in any of the pilot studies related to this research. As long as students confirmed that they were at least 18 years old and had not participated in prior related studies, they were directed to the next page of the form, which asked them to list the titles and course numbers of their current courses.

After the enrollment period ended, I assessed the eligibility of the 594 students who had responded to the survey (as described in the *Participants* section above). Participants were selected to participate only if along with at least one math course, they were also currently enrolled in an eligible social science course for comparison purposes. Thus, I examined each of their listed courses to flag which one(s) could be counted as math courses, which were social science courses, and which met neither requirement (e.g., English, biology, or studio art). Next, a file was created listing up to four math courses and up to four social science courses for each prospective participant. The final lists included 55 unique math courses and 96 unique social science courses. Although I did not collect information about the specific section of each course the students were enrolled in, many of these courses were offered in multiple sections, so the actual number of unique classes participants were in was likely greater than the number of courses. Because students were likely enrolled in so many different class sections, and these sections were unknown, it is unlikely that there would be enough participants per cluster to necessitate the use of multilevel modeling.

Main Survey

Approximately two months after the start of the semester, after participant eligibility and suitability of courses was determined, the main Study 1 survey was dispersed to 313 qualified participants via Qualtrics. I distributed the survey two months into the semester so that students would have accumulated sufficient class experience to provide the basis for their responses to questions about their classroom participation within their particular courses.

The 15-20 minute survey began with an introductory section in which students were asked to select one math course and one social science course on which they would like to focus throughout the following survey. Participants were first presented with a message stating: “Based on the survey you completed several weeks ago, we have determined which of your course qualify as math and/or social science courses.” If I had previously determined that they were enrolled in only one eligible course in each domain, this message was followed by the titles of those two courses and a notice to keep their experiences within those two courses in mind as they completed the remainder of the survey. In the more common case that participants were enrolled in more than one eligible math and/or social science course, I invited them to select which course(s) they would like to reference throughout the survey. Importantly, I asked them to "consider which courses offer the most opportunities for classroom participation via hand-raising" while making their selection(s).

After selecting one course from each domain, participants were presented with two separate blocks of items: one of math-related questions and another of social science-related questions. Each block contained nearly identical measures, with the only difference being the domain or course title referenced in the items. The order in which the two sets of questionnaires were presented was randomized, such that approximately half of the respondents (49.7%) answered math-related questions first, and the other half (50.3%) answered social science-related questions first. Within each of the two blocks, measures were presented in an order designed to align with the psychological pathways proposed in Figure 1, which captured the hypothesized sequence in which participants would experience various components of the model. Within each scale, items were be presented in a random order. In many of the measures, students' particular course titles were included in the question stems. This update was implemented since the pilot studies, with the aim of drawing participants' focus to their particular courses (and genuine experiences within these courses) throughout the survey.

All participants who completed the survey were entered in a raffle for a chance to win one of twenty $50 Amazon eGift Cards. The purpose of these incentives was to encourage more students to participate in the study and to finish the survey.

Follow-Up Survey

A very brief follow-up survey was also distributed during the final week of the semester. This short (5-minute) survey was sent to participants in order to collect data for research questions separate from those addressed in this dissertation and to meet dissertation funding requirements. Of the 161 participants who completed the main survey, 111 (68.9%) also participated in the follow-up survey. As compensation for their participation in this follow-up survey, participants each received a $5 Amazon eGift Card. The measures included in the follow-up survey are included in Appendix C. However, because the data from this survey was not examined for this dissertation, it will not be addressed further.

Measures

The majority of the measures included in Study 1 appeared in at least one of the pilot studies, although some of them were modified based on the pilot study results and other feedback. Below, the order in which the measures are presented reflects the order in which participants answered them within each of the domain-related blocks. Unless otherwise specified, items were presented with six potential response options on a Likert-type scale, from “strongly disagree” to “strongly agree.” See Appendix C for a complete inventory of survey measures, including those on the enrollment form, which are not described below.

Stereotype Threat Susceptibility

Two measures assessed participants’ stereotype threat susceptibility: one novel scenario-based measure created for this dissertation, and a Likert-type scale used in prior work.

Scenario-Based Measure (ST-1). The first was a scenario-based measure with Likert-type items inspired by Marx and Goff (2005), designed to assess participants’ actual experiences of stereotype threat, as opposed to their perceptions of fairness or awareness of gender stereotypes. I developed this measure to address potential shortcomings in the initial stereotype threat measure employed in Pilot Studies 1 and 2 (and as addressed in the *Literature Review*). Participants were instructed to imagine themselves in an exam setting surrounded by either mostly males or mostly females. The order of these two scenarios was counterbalanced, as all participants—both male and female—were asked to imagine themselves in classroom settings with peers and instructors of both the same and opposite gender. Following each scenario, participants were prompted to express their level of agreement with statements intended to gauge stereotype threat effects (e.g., “I would worry about what my professor’s perception of my math [social science] ability would be if I performed poorly on the exam”). The measure was operationalized as a difference score of self-reported threat in the opposite-gender scenario minus self-reported threat in the same-gender scenario, such that positive scores indicated higher levels of stereotype threat in scenarios in which one was surrounded by peers and professors of the opposite gender.

A potential concern about this measure is that it is too explicit in cueing participants to think about their gender rather than focusing on the feeling of stereotype threat. However, I ultimately decided that making gender salient was necessary, as it would otherwise be impossible—particularly in a survey format—to distinguish general exam-related anxiety from potential stereotype threat. In other words, I wanted to ensure that students would tap into their potential experiences of gender-related stereotype threat when completing the measure.

Likert-Type Scale (ST-2). The second measure was a Likert-type scale adapted from the Stereotype Threat in Science Scale by Deemer et al. (2016), a 7-item scale used in prior studies as a self-report measure of stereotype threat. Responses to each item were averaged to create a composite score. This scale was included in the study because it was the best previously used measure of stereotype threat I could find, even though I had some reservations about it being too blunt and focused on awareness of stereotypes rather than the felt experience of fearing evaluation and/or confirming negative stereotypes. The scenario-based measure of stereotype threat (ST-1) was intended to capture this more situated experience of threat, and the second scale (ST-2) was included with the purpose of validating the first measure, which had only been tested in the final pilot study.

Regulatory Focus

Students' domain-specific tendency to pursue their goals with a prevention versus promotion focus was assessed with a 6-item scale adapted from Browman et al. (2017). For each domain, three items assessed tendency toward a prevention focus, and another three assessed tendency toward a promotion focus. Regulatory focus orientation was operationalized as a person's average prevention score minus their average promotion score, such that positive values indicate that participants are predominantly prevention-focused.

Since the pilot studies, I changed one of the items in this scale to align with the structure of the other items. Specifically, in the pilot studies, one of the items was: "I feel like I have made progress toward being successful in math [social science]." This item was adjusted to read: "I frequently think about making progress toward being successful in math [social science]." This format better mirrors the other regulatory focus items, which are framed in terms of what participants *focus on* when pursuing their goals (rather than their perceptions of what they may have already achieved). See Appendix C for a list of specific items.

Class Participation Confidence Threshold

Class participation confidence threshold (CPCT) was assessed using two measures: the slider measure that was used in all four pilot studies (CPCT-1) and the Likert-type measure utilized in Pilot Studies 2 through 4 (CPCT-2).

CPCT-1 (Slider Measure). The first measure of confidence threshold (CPCT-1) presented students with a vignette description related to being in a math or social science course. After reading the vignette, participants were prompted to report the level of confidence required for them to raise their hand to participate. Specifically, they were presented with a continuous slider from 0 to 100% (with answer options of whole percentages), on which they marked their required confidence level. Participants' numeric response on each of the sliders was used as the CPCT-1 measure for each domain. The full vignette is presented in Appendix C.

CPCT-2 (Likert-Type Measure). The second measure of confidence threshold (CPCT-2) was assessed using a Likert-type measure. Participants responded to three items within each domain based on their general willingness to participate under the given circumstances. Responses to each item were averaged to create a composite score.

Although I initially planned to select only one of these piloted threshold measures to include in the dissertation studies, I ultimately decided that they both contributed useful information. The first threshold measure captures an important level of nuance that the three-item scale of the second measure might not. However, the single-item nature of the CPCT-1 scale, combined with the fact that it might be somewhat confusing for participants to navigate, supports the importance of including the CPCT-2 scale. Thus, both measures were included in the study.

Self-Reported Participation Behaviors

Participation behaviors included two related constructs: *self-reported participation frequency* and *frequency of changing one's mind about participating*. Each was measured with two items intended to gauge the regularity of students' participation, as well as frequency of deciding *not* to participate (e.g., "On occasions when you decide that you want to raise your hand to answer a question posed by the instructor in your math class, how often do you change your mind?"). The items were assessed on a Likert-type scale with six response options ranging from

"Never" to "Very often." Responses were averaged to create composite scores for participation frequency and frequency of changing one's mind. See Appendix C for a full list of the items.

Opportunities for Participation

Participants' opportunities for participating in class were assessed with a set of two questions, measured on a Likert-type scale with six response options ranging from "Never" to "Very often." This measure, which was a novel addition after the pilot studies, assessed the frequency with which students had opportunities to participate in each course (i.e., "How often does the instructor of your math course pose questions to the class?" and, "How often do you think students in your math course have an opportunity to participate?"). Responses to each item were averaged to create a composite score.

Self-Confidence in Math and Social Science

Two measures of self-confidence were included in the study: a general measure of confidence within each domain and a more specific measure of confidence with respect to participating within each domain. Specific items included in each measure are in Appendix C. For both measures, responses to each item were averaged to create a composite score.

Domain-Level Confidence. Domain-level self-confidence in math and social science was measured with six items from the Academic Self-Description Questionnaire (Marsh, 1990). The response scale consisted of six options, ranging from "False" to "True."

Participation-Level Confidence. A second self-confidence measure—one that was more specific to participating in class—was also assessed, with the goal of measuring participants' typical levels of certainty that their contribution is valid when presented with an opportunity to participate in class. The intention of including this distinct measure of confidence was to be able to consider participants' typical levels of confidence in the validity of their contributions when examining their confidence thresholds for participation.

Perceived Value of Math and Social Sciences

Perceptions of math's and social science's utility value, intrinsic value, and importance was assessed using six modified items from a perceived task value scale by Eccles and Wigfield (1995). The particular response scale for each item is shown in Appendix C. Individual composite scores for each type of value were calculated as the average of each two-item set.

Costs and Value of Participating in Class

Participants' perceived costs and value of participating in math and social sciences courses was measured with eight Likert-type items. Three items measured students' perceived value of participation (e.g., "I find participating in math class to be enjoyable"), and five items probed participants' perceived costs of participation (e.g., "In math class, I am concerned that if I answer the professor's question incorrectly, other students in my class will think I am dumb.") These measures were exploratory in nature and were not included in the dissertation analyses.

Sense of Belonging in Math and Social Sciences Contexts

Sense of belonging in the math and social sciences domains was measured using an 18-item version of the Math Sense of Belonging Scale (Good et al., 2012). All items in the scale completed the stem, "When I am in a math [social science] setting…" with responses such as "…I feel accepted" and "…I feel like an outsider." Responses to each item were averaged to create a composite score. See Appendix C for a complete list of items.

Math and Social Sciences Identity

Identification with the math and social sciences domains was revised from the 6-item scale used in the pilot studies to be better aligned with an accurate definition of domain-specific identity. I had some concerns that the prior measure was tapping into participants' domain-specific confidence rather than the role of the domain in their identity. For this reason, the identity measure was revised to be comprised of two items compiled from scales by Lesko and Corpus (2006) (i.e., "Being good at math [social science] is *not* an important part of who I am") and Nosek et al. (2002) (i.e., "I consider myself to be a math [social science] person"). A third item (i.e., "Math [social science] is an important part of my identity") was created by me and was included to capture the most basic essence of the construct—i.e., how much of one's identity can be attributed to a particular academic domain. Responses to the three items were averaged to create a composite score.

Career Interest

Interest in pursuing a math-related (STEM) career and a social sciences-related career was measured using the three-item math subscale (adapted for social sciences for the corresponding block) from the Educational Career Interest Scale in Science, Technology, and Mathematics (Oh et al., 2013). An example item was, "I am interested in working in a career that allows me to use math-related skills or knowledge." Responses to each item were averaged to create a composite score. See Appendix C for the full list of items.

Demographics

The items at the end of the survey measured participants' gender (in order to ensure that it had not changed since the enrollment form); date of birth; race and ethnicity; and parents' highest level of education. Additional items asked about participants' academic major, typical grades in math and social science courses, and typical grades in general.

Attention and Suspicion Check

Two attention-check items were included in a randomly-determined position in both the math and social science sense of belonging scales. Both attention checks prompted: "...please leave this item blank for validation purposes (i.e., please do not select a response)." At the end of the survey, participants were also prompted to indicate (1) how distracted they were while completing the questionnaire, (2) how seriously they filled out the questionnaire, and (3) whether they found anything in the survey to be strange or difficult to answer. I reviewed participants' responses to these end-of-survey items against a predetermined threshold for exclusion, and ultimately, no participants met the requirements for exclusion based on these scales.

The attention check items embedded in the sense of belonging scales were used to exclude participants who failed to answer them correctly. The items at the end of the survey were reviewed to determine whether any participants reported a particularly high level of distraction, a low level of care, or a concerning comment that indicated they would have had a qualitatively different experience on the survey than other participants (e.g., a mid-survey computer crash).

Results

Descriptive Analyses

SPSS 27 and Stata 17 were used to conduct the analyses for Study 1. After creating composite variables, I examined descriptive statistics for each construct, including measures of central tendency, skewness, and kurtosis. Table 2 displays the means and standard deviations of all primary variables of interest, arranged according to domain and participant gender.

Analyses of skewness and kurtosis indicated that all constructs were in an acceptable range (between -2 and +2) based on conventional guidelines (George & Mallery, 2010). I also explored normal probability plots to ensure that all constructs were normally distributed. In addition, I examined each variable to check for outliers greater than three standard deviations from the mean. The only variable to display values outside this range was the slider measure of class participation confidence threshold (CPCT-1). However, I decided to retain outliers in this case, as the scores all represented meaningful variability in participants' reported thresholds.

Finally, I examined the reliability of each scale, ensuring that all measures of Cronbach's alpha exceeded 0.70. For the difference score measures (e.g., regulatory focus and ST-1), I assessed the reliability of each subscale separately. With respect to domain value, the distinction between subtypes of value was not critical for this dissertation, and internal consistency was high among the six items ($\alpha = 0.89$ for both math and social science). Therefore, I combined the three original components of value to create a single composite measure.

Table 2

Descriptive Statistics for Key Variables in Study 1

	Math						**Social Science**					
	All participants		Female only		Male only		All participants		Female only		Male only	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Primary Constructs in Model												
ST-1 (diff. score)	0.34	0.74	0.48	0.82	0.06	0.43	0.24	0.74	0.36	0.80	0.01	0.53
ST-2 (Likert)	3.32	1.52	4.05	1.23	1.85	0.83	2.82	1.29	3.17	1.28	2.09	0.96
Reg. focus (diff. score)	0.95	1.00	1.04	0.93	0.78	1.12	0.46	1.05	0.33	0.97	0.72	1.16
CPCT-1 (slider)	79.49	22.32	83.07	19.90	72.19	25.22	72.59	22.34	74.81	20.82	68.08	24.75
CPCT-2 (Likert)	4.53	1.18	4.68	1.11	4.21	1.27	4.16	1.16	4.29	1.11	3.91	1.24
Participation frequency	2.83	1.14	2.71	1.16	3.06	1.05	3.32	1.10	3.31	1.12	3.36	1.07
Freq. of changing mind	3.59	1.30	3.88	1.27	2.98	1.17	3.67	1.03	3.81	1.01	3.38	1.01
Sense of belonging	3.93	0.94	3.76	0.86	4.29	1.00	4.30	0.84	4.33	0.82	4.24	0.91
Domain identity	3.55	1.33	3.34	1.36	3.99	1.17	3.63	1.21	3.81	1.13	3.26	1.29
Career interest	4.11	1.40	3.84	1.43	4.65	1.17	4.49	1.33	4.77	1.11	3.92	1.55
Covariates												
Participation opportunities	4.40	1.15	4.41	1.16	4.40	1.16	4.88	1.22	4.93	1.15	4.79	1.36
Math confidence	4.45	1.08	4.30	1.10	4.76	0.96	4.71	0.93	4.82	0.93	4.50	0.92
Participation confidence	3.59	1.19	3.47	1.16	3.84	1.21	3.82	0.97	3.80	0.95	3.86	1.02
Confidence diff. score	-0.93	2.00	-1.21	1.94	-0.36	2.01	-0.35	1.69	-0.49	1.68	-0.05	1.70
Math value	3.55	0.81	3.46	0.08	3.73	0.11	3.68	0.77	3.82	0.07	3.40	0.10

Research Aim 1: Gender and Domain Differences in Constructs

In order to address Research Aim 1 (i.e., *to explore whether there are gender and domain differences in the mean levels of all constructs in the model*), I conducted a series of 2 (Gender: male vs. female) x 2 (Domain: math vs. social science) mixed ANOVAs, with domain as the within-participant factor, for each construct. I first examined the main effects and then, if they were qualified by significant interactions, I explored the simple effects based on the estimated marginal means. In conducting the ANOVAs, I was initially concerned about the unequal sizes of the gender groups, given that women represented about two-thirds of the sample. However, unequal groups are only a concern for ANOVAs when there is heterogeneity of variance (Wickens & Keppel, 2004). Levene's test of homogeneity of variance was nonsignificant for all of the mixed ANOVAs, indicating that the unequal sample sizes were not ultimately a concern.

Table 3 summarizes the ANOVA results for each construct. Because the results are displayed in the table, below I will focus mostly on exploring the differences that emerged as statistically significant.

Table 3

Summary of Gender x Domain ANOVA Results for Study 1 Constructs

	Gender			Domain			Gender x Domain		
Construct	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
ST-1 (difference score)	13.89	<.001	.080	1.67	.198	.010	0.27	.606	.002
ST-2 (Likert)	95.89	<.001	.376	10.78	.001	.064	34.22	<.001	.177
Regulatory focus (difference score)	0.28	.597	.002	9.90	.002	.059	6.92	.009	.042
CPCT-1 (slider)	8.20	.005	.049	9.29	.003	.055	1.05	.308	.007
CPCT-2 (Likert)	7.36	.007	.044	9.46	.002	.056	0.16	.690	.001
Participation frequency	1.75	.188	.011	16.00	<.001	.091	1.69	.196	.011
Frequency of changing mind	16.95	<.001	.096	2.70	.102	.017	5.76	.018	.035
Sense of belonging	3.43	.066	.021	9.80	.002	.058	13.47	<.001	.078
Domain identity	0.13	.720	.001	0.67	.413	.004	14.26	<.001	.082
Career interest	0.03	.875	<.001	0.39	.534	.002	24.35	<.001	.133
Participation opportunities	0.30	.582	.002	9.18	.003	.055	0.16	.686	.001
Domain confidence	0.33	.568	.002	1.40	.239	.009	12.29	<.001	.072
Participation confidence	2.12	.148	.013	2.66	.105	.016	2.29	.132	.014
Confidence difference score	6.75	.010	.041	8.25	.005	.049	1.28	.260	.008
Domain value	0.68	.412	.004	0.02	.900	<.001	12.36	<.001	.072

The results of the analysis of both measures of stereotype threat indicated a main effect of gender ($F(1, 157) = 13.89$, $p < .001$, $\eta_p^2 = .08$ for the scenario-based measure, ST-1; $F(1, 157) = 95.89$, $p < .001$, $\eta_p^2 = .38$ for the Likert-type measure, ST-2). Specifically, on average, women displayed higher levels of stereotype threat ($M = 0.42$, $SE = .06$; $M = 3.61$, $SE = .10$, for the ST-1 and ST-2 measures, respectively) than men ($M = 0.04$, $SE = .08$; $M = 1.97$, $SE = .14$, respectively). Analysis of the Likert-type measure of stereotype threat only (ST-2) revealed a main effect of domain, $F(1, 157) = 10.78$, $p < .001$, $\eta_p^2 = .06$, such that participants on average exhibited higher levels of stereotype threat susceptibility in the math domain ($M = 2.95$, $SE = .09$) compared to the social science domain ($M = 2.63$, $SE = .09$). The main effects for ST-2 were qualified by a significant interaction, $F(1, 157) = 34.22$, $p < .001$, $\eta_p^2 = .18$. Pairwise comparisons revealed that when comparing stereotype threat susceptibility between domains within each gender, females reported significantly higher levels of threat in math ($M = 4.05$, $SE = .11$) than in social science ($M = 3.17$, $SE = .11$), $p < .001$. This difference was not present for males, (math: $M = 1.85$, $SE = .15$; social science: $M = 2.09$, $SE = .16$), $p = .119$. Further pairwise comparisons revealed that within each domain, female students reported higher levels of stereotype threat susceptibility than male students ($p < .001$ in both domains). Because of the theoretical importance of identification with the math domain in predicting stereotype threat susceptibility (Steele et al., 2002), I conducted additional ANOVAs with math major status and its interactions with each other predictor included. Results revealed no main effect of major, $F(1, 157) = 1.09$, $p = .299$, $\eta_p^2 = .007$, and no significant interactions involving major.

Regulatory focus, operationalized as a tendency toward a prevention focus over a promotion focus, did not show a main effect of gender, $F(1, 157) = 0.28$, $p = .597$, $\eta_p^2 = .002$ (males: $M = 0.75$, $SE = .10$; females: $M = 0.69$, $SE = .07$). However, a main effect of domain was present, $F(1, 157) = 9.90$, $p = .002$, $\eta_p^2 = .06$, such that on average, participants indicated a tendency toward a prevention focus in math ($M = 0.91$, $SE = .08$) more than in social science ($M = 0.53$, $SE = .09$). This main effect was also qualified by a significant interaction, $F(1, 157) = 6.92$, $p = .009$, $\eta_p^2 = .04$. Pairwise comparisons by gender showed that women indicated higher levels of a prevention focus in math ($M = 1.04$, $SE = .10$) than in social science ($M = 0.33$, $SE = .10$), $p < .001$, whereas men did not exhibit any difference (math: $M = 0.78$, $SE = .14$; social science: $M = 0.72$, $SE = .14$), $p = .753$. This interaction seemed to be driven mostly by gender differences in the social science domain, $p = .029$, as opposed to the math domain, $p = .126$.

Both measures of class participation confidence threshold, CPCT-1 and CPCT-2, revealed main effects of gender and domain. The slider measure of confidence threshold, CPCT-1, was significantly higher for women ($M = 78.94$, $SE = 1.77$) than for men ($M = 70.13$, $SE = 2.52$), $F(1, 157) = 8.20$, $p = .005$, $\eta_p^2 = .05$. Participants on average reported higher thresholds in math ($M = 77.63$, $SE = 1.83$) compared to social science ($M = 71.44$, $SE = 1.86$), $F(1, 157) = 9.29$, $p = .003$, $\eta_p^2 = .06$. Similarly, analysis of the Likert-type measure of confidence threshold, CPCT-2, revealed main effects of gender, $F(1, 157) = 7.36$, $p = .007$, $\eta_p^2 = .04$, and domain, $F(1, 157) = 9.46$, $p = .002$, $\eta_p^2 = .06$. As with the first measure of threshold, women exhibited higher thresholds than men (males: $M = 4.06$, $SE = .13$; females: $M = 4.49$, $SE = .09$), and all participants reported significantly higher thresholds in math than in social science (math: $M = 4.45$, $SE = .10$; social science: $M = 4.10$, $SE = .10$). The main effects were not qualified by significant interactions for either measure of confidence threshold (see Table 3).

Analysis of self-reported participation frequency showed a significant main effect of domain, $F(1, 157) = 16.00$, $p < .001$, $\eta_p^2 = .09$, such that students reported participating more frequently in social science classes ($M = 3.33$, $SE = .09$) than in math classes ($M = 2.89$, $SE =$

.09). Results also revealed a main effect of domain with respect to perceived opportunities for participation, such that students perceived significantly more opportunities to participate in their social science courses (*M* = 4.86, *SE* = .10) than their math courses (*M* = 4.40, *SE* = .10), $F(1, 157) = 9.18$, $p = .003$, $\eta_p^2 = .06$. Contrary to expectations, there was no main effect of gender for participation frequency, $F(1, 157) = 1.75$, $p = .118$, $\eta_p^2 = .01$. However, noting the significant difference in opportunities for participation between domains, I conducted a supplemental exploratory analysis, in which participation was operationalized as a ratio of frequency of participation to opportunities for participation. Although this measure, computed as a ratio of numeric values obtained from Likert-type scales, cannot be interpreted as a precise indicator of students' participation relative to the number of opportunities, it can be used in an exploratory analysis as an approximate indicator of how students with similar levels of opportunity for classroom participation took advantage of this opportunity. Analysis of this new variable did reveal a significant gender difference, $F(1, 157) = 4.72$, $p = .031$, $\eta_p^2 = .03$, with men (*M* = 0.81, *SE* = .05) participating more frequently than women (*M* = 0.69, *SE* = .03).

With respect to frequency of changing one's mind about participating, results revealed a main effect of gender, $F(1, 157) = 16.95$, $p < .001$, $\eta_p^2 = .10$, with female participants (*M* = 3.85, *SE* = .09) reporting a higher frequency of changing their mind about participating than male participants (*M* = 3.18, *SE* = .13). Although there was no main effect of domain, analyses revealed a significant interaction, $F(1, 157) = 5.76$, $p = .018$, $\eta_p^2 = .04$, such that male students changed their minds about participating more frequently in social science classes (*M* = 3.38, *SE* = .14) than in math classes (*M* = 2.98, *SE* = .17), $p = .015$, while female students reported similarly high levels in both domains (math: *M* = 3.88, *SE* = .12; social science: *M* = 3.81, *SE* =

.10), $p = .511$. Further, women changed their minds about participating more frequently than men in both math ($p < .001$) and social science ($p = .012$).

Exploration of gender and domain differences in the outcome variables in the model (i.e., sense of belonging, identification with the domain, and career interest) revealed significant interactions for all three constructs. Notably, the only significant main effect was a main effect of domain for sense of belonging, $F(1, 157) = 9.80$, $p = .002$, $\eta_p^2 = .06$, such that participants on average reported higher sense of belonging in social science courses ($M = 4.28$, $SE = .07$) than math courses ($M = 4.02$, $SE = .08$). For sense of belonging, a significant interaction, $F(1, 157) = 13.47$, $p < .001$, $\eta_p^2 = .08$, revealed that women experienced a lower sense of belonging in the math domain ($M = 3.76$, $SE = .09$) than in the social science domain ($M = 4.33$, $SE = .08$), $p < .001$, whereas men experienced similar levels of belonging in both domains (math: $M = 4.29$, $SE = .12$; social science: $M = 4.24$, $SE = .12$), $p = .742$. Further, this interaction was driven by gender differences in the math domain ($p < .001$) as opposed to social science ($p = .562$).

Pairwise comparisons intended to explore the significant domain-by-gender interaction for domain-specific identity, $F(1, 157) = 14.26$, $p < .001$, $\eta_p^2 = .08$, highlighted that men reported higher levels of math identity ($M = 3.99$, $SE = .19$) than social science identity ($M = 3.26$, $SE = .16$), $p = .006$. For women, on the other hand, the opposite was true: they reported higher levels of social science identity ($M = 3.81$, $SE = .11$) than math identity ($M = 3.34$, $SE = .13$), $p = .011$. When comparing mean levels of identity by domain, significant gender differences were present in both math ($p = .003$) and social science ($p = .006$). Similarly, the significant interaction for career interest ($F(1, 157) = 24.35$, $p < .001$, $\eta_p^2 = .13$) revealed that male participants, on average, expressed higher interest in math-related ($M = 4.65$, $SE = .19$), as opposed to social science-related ($M = 3.92$, $SE = .18$), careers, $p = .009$, whereas women indicated the opposite preference (math: $M = 3.84$, $SE = .13$; social science: $M = 4.77$, $SE = .12$), $p < 0.001$). Within each domain, significant gender differences were present ($p < .001$ in both domains).

Finally, I explored domain and gender differences in the variables included in the study as covariates—namely, confidence and value. With respect to general confidence within each domain, no significant main effects were present. However, a significant interaction ($F(1, 157) = 12.29$, $p < .001$, $\eta_p^2 = .07$) revealed that female participants reported significantly lower levels of math confidence ($M = 4.30$, $SE = .10$) than social science confidence ($M = 4.82$, $SE = .09$), $p < .001$, whereas male participants expressed no significant difference (math: $M = 4.76$, $SE = .15$; social science: $M = 4.50$, $SE = .13$), $p = .158$. There were also significant gender differences in confidence in both domains, with women displaying higher social science confidence than men ($p = .042$) and men exhibiting higher math confidence than women ($p = .010$). No mean differences were present for the measure of participation-specific confidence. However, when utilizing the difference score measure of confidence (i.e., participation-specific confidence over and above one's typical confidence threshold, measured as CPCT-2), main effects of gender ($F(1, 157) = 6.75$, $p = .010$, $\eta_p^2 = .04$) and domain ($F(1, 157) = 8.25$, $p = .005$, $\eta_p^2 = .05$) were revealed. Although both men and women reported, on average, negative confidence difference scores (such that their typical level of confidence was below their typical confidence threshold), women reported a significantly greater difference ($M = -0.85$, $SE = .14$) than men ($M = -0.21$, $SE = .20$). With respect to domain differences, all participants on average reported a wider gap between confidence and threshold in their math courses ($M = -0.79$, $SE = .16$) than in their social science courses ($M = -0.27$, $SE = .14$). Finally, analyses of domain-specific value revealed a significant interaction, $F(1, 157) = 12.36$, $p < .001$, $\eta_p^2 = .07$, such that men reported significantly higher math value ($M = 3.73$, $SE = .11$) than social science value ($M = 3.40$, $SE =$

.10), $p = .040$, whereas women expressed the opposite (math: $M = 3.46$, $SE = .08$; social science: $M = 3.82$, $SE = .07$), $p = .002$. Further, comparisons within domain revealed that men reported significantly higher math value than women ($p = .042$), while women also expressed significantly higher social science value than men ($p = .001$).

Research Aim 2: Relations Between Constructs in the Theoretical Model

To address Research Aim 2 (i.e., *to explore relations between constructs in the theoretical model, including the directionality of these relations*), I first examined bivariate correlations between constructs within each domain. Then, I employed path analysis to examine the directionality of the relations, as well as how they were affected by the inclusion of other variables in the models.

Bivariate Correlations

Zero-order correlations between the primary variables of interest in the theoretical model, as well as potential covariates, are summarized in Table 4 (math) and Table 5 (social science). Here, I will focus on exploring correlations between variables with posited relations in the theoretical model.

Because some of the key variables in this study were assessed with more than one measure, I first examined the bivariate correlations between the separate versions of these measures. In both the math and social science domains, both the scenario-based and Likert-type measures of stereotype threat susceptibility (ST-1 and ST-2, respectively) were significantly correlated with each other (math: $r = .453$, $p < .001$; social science: $r = .293$, $p < .001$). As discussed previously, ST-2 was deemed a measure of awareness of gender-related stereotypes, rather than the feeling “in the air” (i.e., situated feeling) that stereotype threat has come to be conceptualized as. Thus, given the significant correlation between both stereotype threat measures, I decided to proceed with using the scenario-based measure (ST-1) for the remaining analyses.

I also examined the correlations between the two measures of confidence threshold. As discussed previously, the slider measure (CPCT-1) captures a level of nuance that the three-item Likert scale (CPCT-2) does not. However, because CPCT-1 was measured with only a single item, I decided to use CPCT-2 as the measure of threshold for the remaining analyses. Notably, the two threshold measures were strongly correlated in both the math ($r = .650$, $p < .001$) and social science ($r = .755$, $p < .001$) domains. Finally, I noted that the two participation measures (frequency of participation and frequency of changing one's mind about participating) were, as expected, negatively correlated in both domains (math: $r = -.189$, $p = .016$; social science: $r = -.297$, $p < .001$).

Table 4

Zero-Order Correlations Between Key Variables of Interest in Study 1 (Math)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Primary Constructs															
1. ST-1 (scenario)	1														
2. ST-2 (Likert)	.453***	1													
3. Regulatory focus	.025	.029	1												
4. CPCT-1 (slider)	.101	.257**	.139^{+}	1											
5. CPCT-2 (Likert)	.108	.299***	.239**	.650***	1										
6. Participation frequency	-.079	-.251**	-.266**	-.547***	-.653***	1									
7. Freq. of changing mind	.008	.377***	.100	.184*	.307***	-.189*	1								
8. Sense of belonging	-.159*	-.273***	-.419***	-.326***	-.422***	.423***	-.388***	1							
9. Domain identity	-.128	-.075	-.514***	-.068	-.139^{+}	.165*	-.165*	.597***	1						
10. Career interest	-.142^{+}	-.076	-.590***	-.044	-.061	.121	-.118	.512***	.767***	1					
Covariates															
11. Participation opportunities	.094	-.029	-.188*	-.056	-.159*	.254**	.015	.112	-.006	.144^{+}	1				
12. Domain confidence	-.008	-.149^{+}	-.366***	-.237**	-.263**	.256**	-.192*	.677***	.703***	.565***	-.077	1			
13. Participation confidence	-.083	-.263**	-.262**	-.271**	-.417***	.438***	-.392***	.532***	.183*	.190*	.182*	.387***	1		
14. Conf. diff. score	-.113	-.333***	-.297***	-.547***	-.842***	.648***	-.415***	.566***	.191*	.149^{+}	.203**	.386***	.842***	1	
15. Domain value	-.099	.021	-.585***	-.133^{+}	-.153^{+}	.207**	-.058	.565***	.770***	.824***	.145^{+}	.600***	.264**	.248**	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5

Zero-Order Correlations Between Key Variables of Interest in Study 1 (Social Science)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Primary Constructs															
1. ST-1 (scenario)	1														
2. ST-2 (Likert)	.293***	1													
3. Regulatory focus	-.131	-.125	1												
4. CPCT-1 (slider)	.064	.184*	.242**	1											
5. CPCT-2 (Likert)	.096	.219**	.200*	.755***	1										
6. Participation frequency	-.034	-.066	-.311***	-.461***	-.558***	1									
7. Freq. of changing mind	-.049	.343***	.104	.328***	.454***	-.297***	1								
8. Sense of belonging	.037	-.174*	-.390***	-.312***	-.328***	.590***	-.362***	1							
9. Domain identity	.197*	.270**	-.566***	-.175*	-.115	.387***	-.051	.498***	1						
10. Career interest	.127	.269**	-.624***	-.094	.011	.296***	.087	.500***	.736***	1					
Covariates															
11. Participation opportunities	.023	.021	-.142^{+}	-.085	-.134^{+}	.289***	-.078	.169*	.145^{+}	.198*	1				
12. Domain confidence	.140^{+}	.081	-.425***	-.246**	-.150^{+}	.410***	-.114	.619***	.452***	.485***	.058	1			
13. Participation confidence	-.040	-.080	-.065	-.233**	-.253**	.393***	-.248**	.547***	.243**	.210**	.168*	.419***	1		
14. Conf. diff. score	-.089	-.197*	-.175*	-.652***	-.832***	.609***	-.454***	.540***	.219**	.113	.189*	.344***	.748***	1	
15. Domain value	.151^{+}	.275***	-.590***	-.192*	-.121	.382***	-.018	.516***	.757***	.861***	.201*	.535***	.256**	.230**	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

In addition to examining correlations between measures intended to capture the same underlying construct, I also reviewed zero-order correlations between variables that were hypothesized to be proximally related in the theoretical model. Although the scenario-based measure of stereotype threat susceptibility (ST-1) was not significantly correlated with any relevant variables, ST-2 was positively associated with both measures of confidence threshold in both domains. In the math domain only, ST-2 was negatively associated with participation frequency and, in both domains, positively associated with frequency of changing one's mind. Regulatory focus (i.e., tendency toward a prevention focus) exhibited positive correlations with CPCT-1 and CPCT-2 in both domains (although the association was only marginally significant with CPCT-1 in the math domain). A predominant prevention focus was also negatively correlated with participation frequency and strongly negatively correlated with all outcome measures in the model (sense of belonging, identity, and career interest) in both domains. Both threshold measures displayed strong negative correlations with participation frequency and positive correlations with frequency of changing one's mind in both domains. Frequency of participation was significantly positively correlated with both sense of belonging and domain identity in both domains, which in turn were both strongly correlated with career interest. In sum, the majority of the posited relations between variables were significant, other than the correlations involving stereotype threat susceptibility.

Path Analyses

After exploring bivariate correlations, I conducted path analyses in Stata 17 to test the theoretical model introduced in Figure 1.

Approach to Path Analysis. The general approach I adopted was to utilize path analysis as "a technique for testing models, not for building them" (Streiner, 2005, p. 121). Although I did examine modification indices and largest standardized residuals, I made adjustments with caution and an overall aim of balancing theory, parsimony, and fit within the constraints of my data. Furthermore, I generally prioritized the ability to compare models across domains and studies over attainment of best model fit. Therefore, although in some cases it is possible that I could have removed certain pathways to achieve a better model fit, I opted to leave in many nonsignificant pathways for the purposes of comparing related models. Given constraints such as sample size, I ran multiple versions of each model in order to determine the best approach to address this research aim; these additional model versions are available upon request. Below, I will focus only on the tests of the initial theoretical model and the "final" models for each domain.

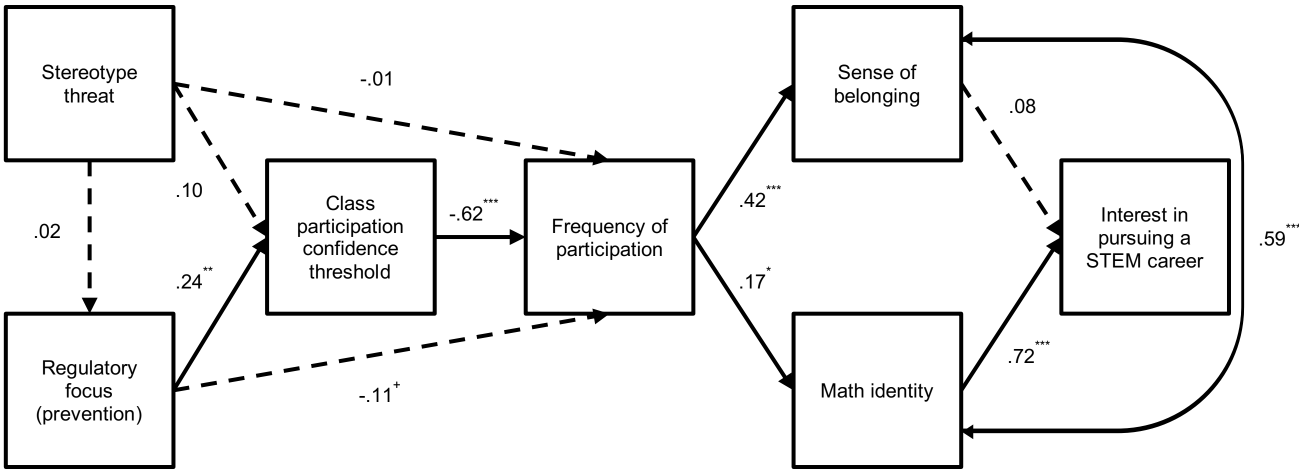
For all path analyses, model fit was evaluated in alignment with guidelines for acceptable ranges of conventional fit indices: χ^2/df in the range of 1 to 3, comparative fit index (CFI $\geq$ 0.95), the root mean square error of approximation (RMSEA $\leq$ 0.08), and the standardized root mean square residual (SRMR $\leq$ 0.08) (Hu & Bentler, 1999; Weston, et al., 2008).

Math Domain. In the first step of the analysis, the hypothesized path model (as depicted in Figure 1) was fit to the data for each domain. Figure 2 displays the results of this initial path analysis for the math domain.

The fit indices indicated a poor fit to the data, with CFI = .817, RMSEA = .219 CI [.178, .263], and SRMR = .157. The χ^2 test was also significant, $\chi^2(10) = 87.051$, $p < .001$. Although it is generally inappropriate to evaluate parameters with such a poor level of model fit, it is noteworthy that all pathways from stereotype threat susceptibility to other constructs were nonsignificant.

Figure 2

Initial Path Model (Study 1 – Math)



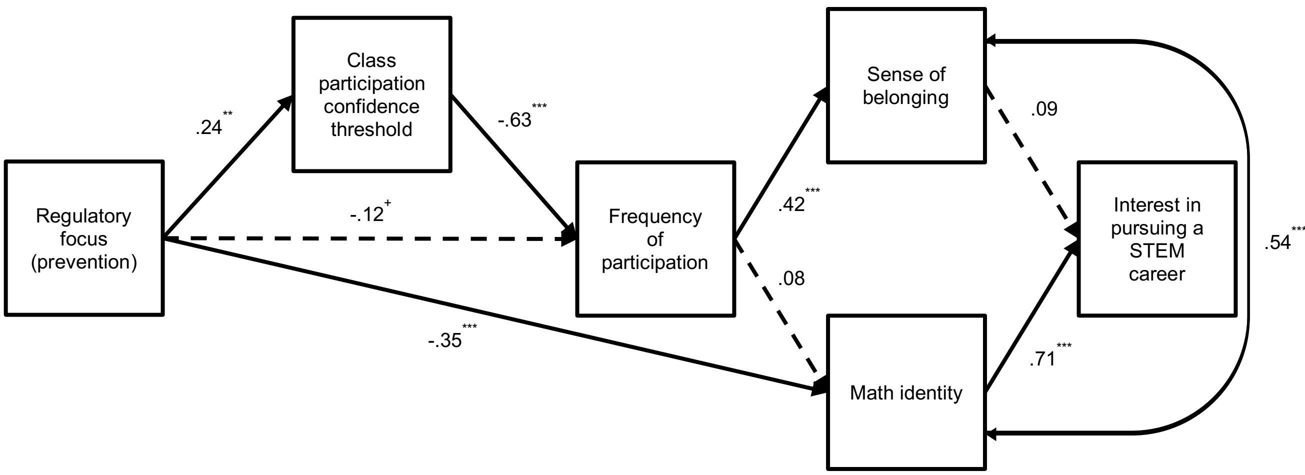
Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .817, RMSEA = .219 CI [.178, .263], and SRMR = .157. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

In the next step of the analysis, I removed stereotype threat from the model altogether. Taking into account the concerns I had with how stereotype threat was being assessed, the lack of significant associations between stereotype threat and other variables in the examination of bivariate correlations, as well as the nonsignificant pathways in the first iteration of the path model, I decided that it did not make sense to keep stereotype threat susceptibility in the model. Finally, based on modification indices, and in view of theoretical considerations, I added a pathway from prevention orientation to math identity. Consistent with self-perception theory (Bem, 1972), it makes sense that students who chronically adopt a prevention focus within a domain might experience unpleasant levels of anxiety and concern and therefore may begin to deidentify with the domain.

The final model for the math domain, taking these two modifications into account, is displayed in Figure 3. The fit indices still indicated a relatively poor fit to the data, although some of the fit indices improved since the initial model (CFI = .880, RMSEA = .229 CI [.177, .285], and SRMR = .117). The χ^2 test was still significant, $\chi^2(6) = 56.361$, $p < .001$.

Figure 3

Final Path Model (Study 1 – Math)



Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .880, RMSEA = .229 CI [.177, .285], and SRMR = .117. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

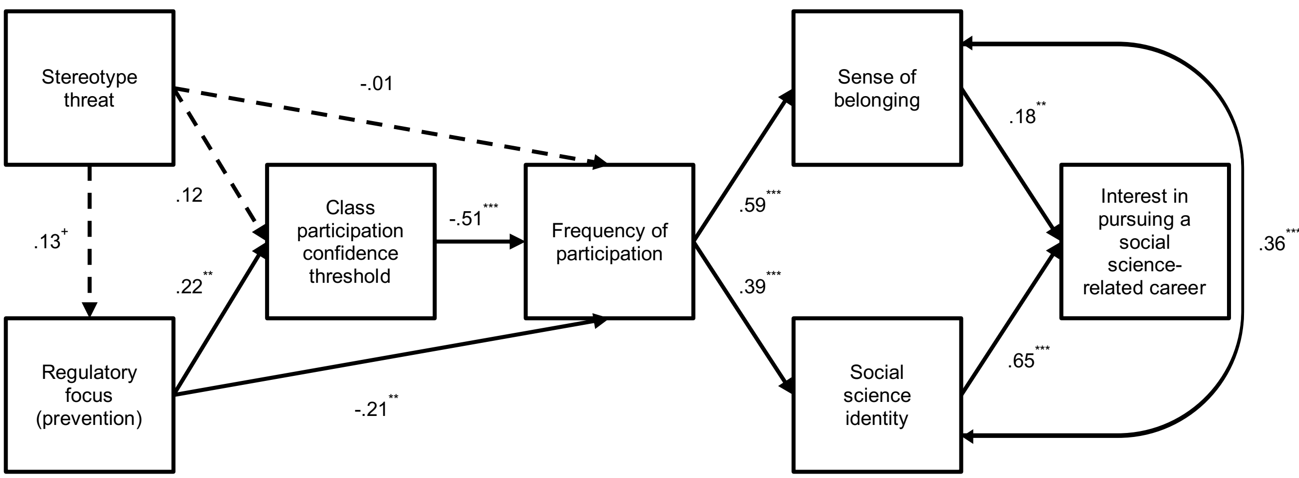
As will be addressed in the *Limitations* section, interpretations of the path coefficients in this model—given the poor model fit—must be done with a considerable degree of caution. Despite these valid concerns, I will discuss what has emerged within the model as it stands. A stronger prevention focus in the math domain predicted a higher class participation confidence threshold ($\beta = 0.24$, $p = .001$) and reduced math identity ($\beta = -.35$, $p < .001$). It also had a marginally significant direct effect on participation frequency ($\beta = -0.12$, $p = .054$). Higher class participation confidence thresholds, as measured by CPCT-2, were associated with lower participation frequency ($\beta = -.63$, $p < .001$). No significant relation emerged between participation frequency and math identity. However, as expected, higher frequency of participation predicted a higher sense of belonging in the math domain, $\beta = 0.42$, $p < .001$.

Finally, although sense of belonging did not significantly predict interest in pursuing a STEM career, β = 0.09, p = .181, math identity was shown to be a significant predictor of career interest, β = 0.71, p < .001.

Social Science Domain. Next, the theoretical path model was tested in the social science domain (see Figure 4). The fit indices indicated a poor fit to the data, with CFI = .789, RMSEA = .232 CI [.191, .275], SRMR = .150, and a significant χ^2 test, $\chi^2(10) = 96.117$, p < .001.

Figure 4

Initial Path Model (Study 1 – Social Science)



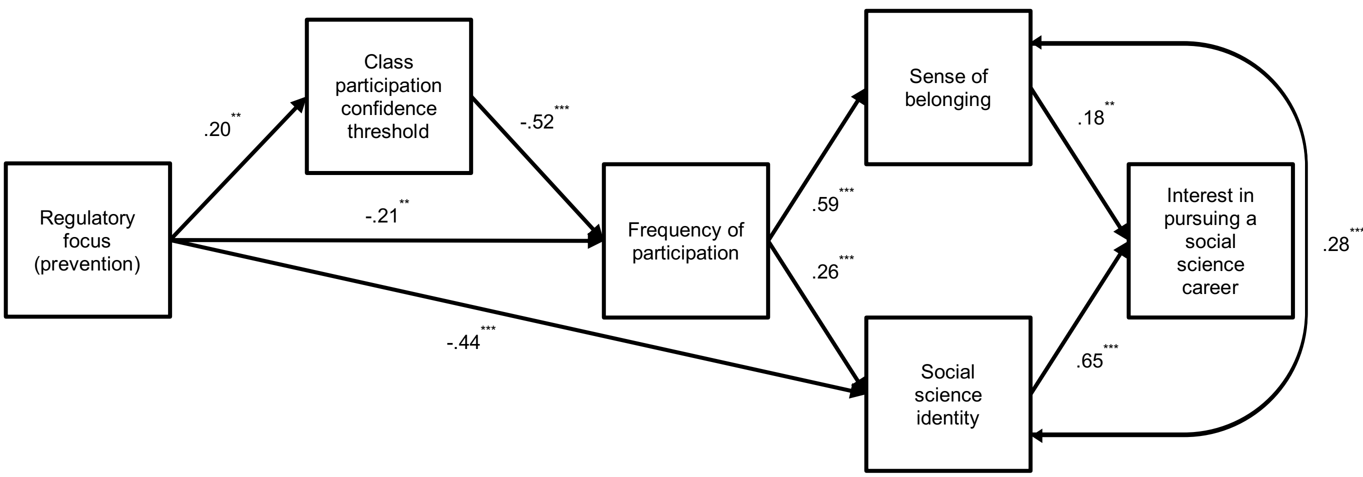
Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .789, RMSEA = .232 CI [.191, .275], and SRMR = .150. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

As in the math domain, stereotype threat susceptibility failed to predict any other variables. Thus, in the next steps of model revisions, I removed it from the model. This decision was also justified by the fact that, like in the math domain, the zero-order correlations between stereotype threat and other variables were also nonsignificant in the social science domain. The modification indices for this model also suggested the same addition as the indices for the math model: a pathway from regulatory focus to social science identity.

The final model for the social science domain is presented in Figure 5. It is important to note that the fit indices still indicated a relatively poor fit to the data (CFI = .883, RMSEA = .222 CI [.169, .278], and SRMR = .096), although the indices were somewhat improved from the prior iteration of the model. The χ^2 test was significant, $\chi^2(6) = 53.219$, $p < .001$.

Figure 5

Final Path Model (Study 1 – Social Science)



Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .883, RMSEA = .222 CI [.169, .278], and SRMR = .096. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

As with the final model in the math domain, the relatively poor fit statistics are an important caveat for the interpretation of the model parameters. Again, I will discuss what has emerged within the model as it stands, noting the stipulation that these estimates may be incorrect due to inadequate model fit. In alignment with results in the math domain, prevention focus positively predicted class participation confidence threshold ($\beta = 0.20$, $p = .008$) and negatively predicted identification with the domain ($\beta = -.44$, $p < .001$). In contrast to the math domain, in which the effect was only marginal, there was also a significant direct effect of regulatory focus on participation frequency (β = -0.21, $p = .001$). Further, higher class participation confidence threshold emerged as a predictor of lower participation frequency (β = -.52, $p < .001$), also mirroring results in the math domain. Frequency of participation predicted increased sense of belonging, β = 0.59, $p < .001$, as well as increased identification with social science, β = 0.26, $p < .001$. Finally, both sense of belonging (β = 0.18, $p = .003$) and social science identity (β = 0.65, $p < .001$) emerged as significant predictors of interest in a social science-related career.

Research Aim 3: Mediation Effects

In addition to exploring the direct effects in the model, I was also interested in examining potential indirect effects. In order to address Research Aim 3 (i.e., *to explore potential mediating mechanisms underlying the relations identified by the investigation of Research Aim 2*), I conducted more targeted investigations of the primary portion of the model (i.e., regulatory focus → confidence threshold → frequency of participation → the outcome variables of belonging and identity). Although a direct pathway from regulatory focus to sense of belonging was not posited in the final models for Study 1, I included it in these more targeted investigations as a means of exploring indirect effects.

Because the sample size of Study 1 was too small to support the addition of more parameters in the path models, I was not able to investigate the impact of relevant control variables in the models—namely, confidence and value, as suggested by Expectancy-Value Theory (Wigfield & Eccles, 2000). Therefore, I included these covariates—general (domain) confidence, specific (participation-related) confidence, and domain value—in more targeted investigations, among fewer parameters. I also included a fourth covariate—perceived opportunities for participation—that seemed important to investigate with respect to participation frequency. Below, I present four mediation models with covariates—two in each domain, with sense of belonging and domain identity as the final outcome measures.

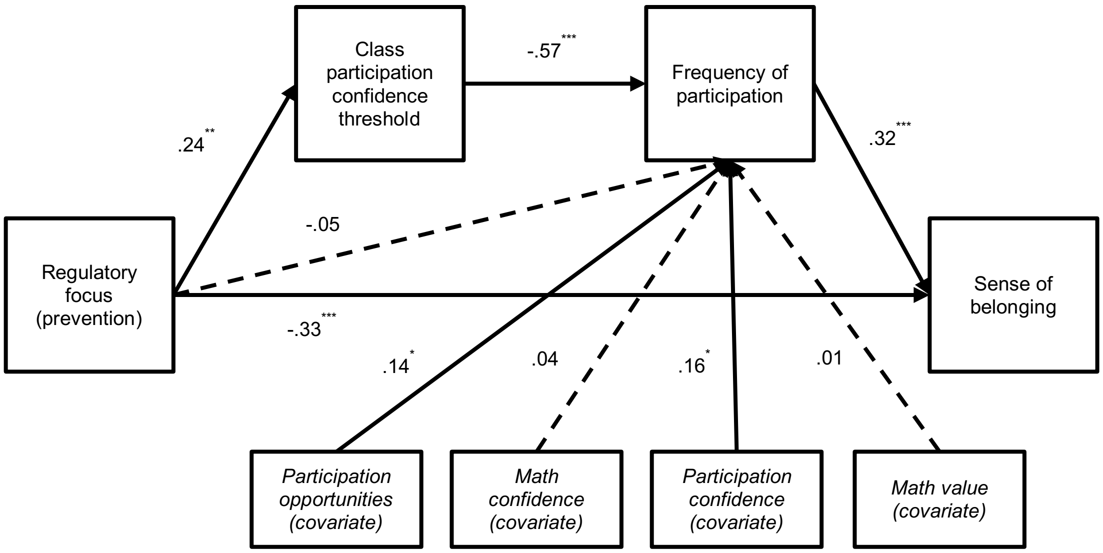
Regulatory Focus to Sense of Belonging

I first explored mediation models for regulatory focus to sense of belonging, with CPCT-2 and participation frequency as potential mediators. Figure 6 displays the results of the model in the math domain.

Before interpreting the coefficients and the results of the mediation, I will call attention to the poor model fit (see the note on Figure 6). In the context of this poor model fit, however, I will also note the significant direct effects from regulatory focus to sense of belonging and class participation confidence threshold, confidence threshold to frequency of participation, and participation frequency to sense of belonging. It is notable that these effects still hold while controlling for the effects of expectancy and value covariates. Despite its significance in the final model in the math domain (Figure 3), the direct pathway from regulatory focus to participation frequency was no longer marginally significant with the addition of the covariates.

Figure 6

Mediation Model of Regulatory Focus to Sense of Belonging (Study 1 – Math)



Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .571, RMSEA = .288 CI [.245, .333], and SRMR = .143. All covariates were allowed to covary. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

With respect to indirect effects—summarized in Table 6 for all four mediation models—results revealed a significant indirect effect of regulatory focus on belonging via CPCT-2 and participation frequency (β = -.06, *SE* = .03, *p* = .047). Furthermore, there was also a significant indirect effect of regulatory focus on frequency of participation via CPCT-2 (β = -.15, *SE* = .05, *p* = .003).

Table 6

Indirect Effects in Targeted Mediation Models of Study 1

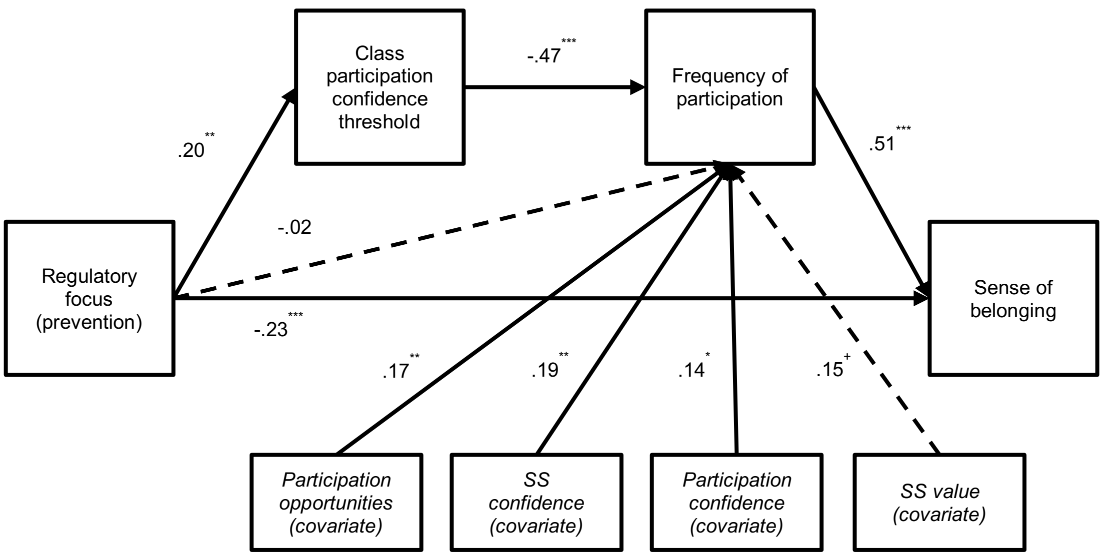
	Math						**Social Science**					
	Direct effect		Indirect effect		Total effect		Direct effect		Indirect effect		Total effect	
	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE
Model: Reg. focus → CPCT-2 → participation frequency → belonging												
Reg. focus → participation frequency	-.05	.08	$-.15^{**}$	.05	$-.19^{*}$	.09	-.02	.08	$-.09^{*}$	.04	-.12	.08
Reg. focus → belonging	$-.33^{***}$	.06	$-.06^{*}$	.03	$-.39^{***}$	.07	$-.23^{***}$	.05	-.05	.03	$-.29^{***}$	.06
Model: Reg. focus → CPCT-2 → participation frequency → identity												
Reg. focus → participation frequency	-.05	.08	$-.14^{**}$	.05	$-.19^{*}$	.09	-.02	.08	$-.09^{*}$	.04	-.12	.08
Reg. focus → identity	$-.51^{***}$	.09	-.01	.02	$-.51^{***}$	.09	$-.49^{***}$	.08	-.03	.02	$-.52^{***}$	.08

Note. All models include the following covariates: opportunities for participation, domain confidence, participation-specific confidence, and domain value.
$^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

The same mediation model is depicted for the social science domain in Figure 7. I will again note the relatively poor fit statistics (detailed in the note below the figure) as a caveat for interpretation of the following results.

Figure 7

Mediation Model of Regulatory Focus to Sense of Belonging (Study 1 – Social Science)



Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .732, RMSEA = .218 CI [.175, .264], and SRMR = .103. All covariates were allowed to covary. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

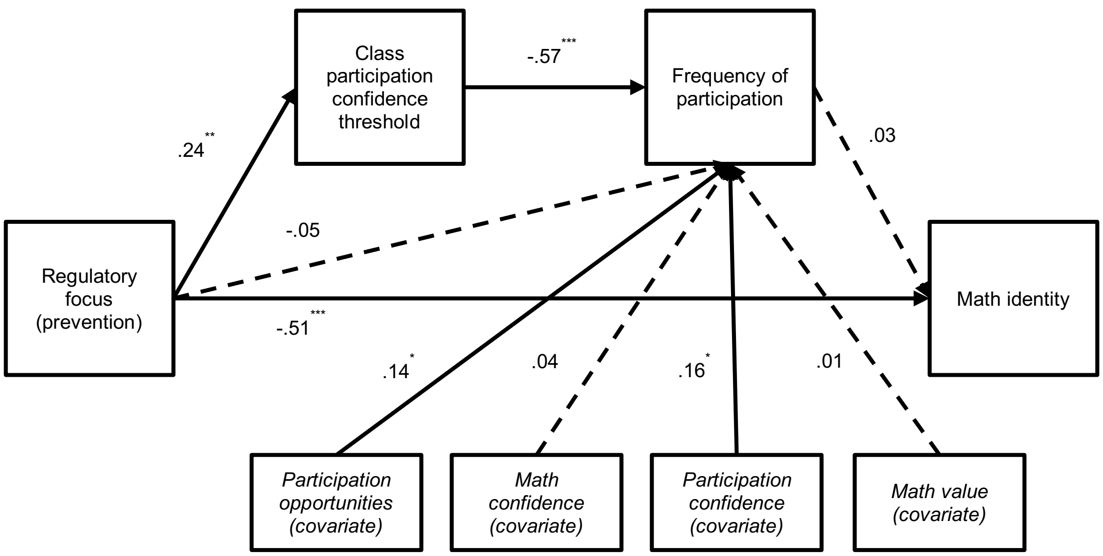
As in the math domain, there were significant direct effects from regulatory focus to sense of belonging and class participation confidence threshold, confidence threshold to frequency of participation, and participation frequency to sense of belonging, but not regulatory focus to participation frequency. Results also exhibited a significant indirect effect of regulatory focus on frequency of participation via CPCT-2 ($\beta = -.09$, $SE = .04$, $p = .014$). In contrast to the results in the math domain, there was no significant indirect effect of regulatory focus on sense of belonging in social science (see Table 6).

Regulatory Focus to Domain Identity

In addition to exploring mediation models from regulatory focus to sense of belonging, I also explored models with identity as the final outcome variable (with the same potential mediators of CPCT-2 and participation frequency). Figure 8 depicts this mediation model for the math domain.

Figure 8

Mediation Model of Regulatory Focus to Domain Identity (Study 1 – Math)



Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .481, RMSEA = .342 CI [.299, .387], and SRMR = .142. All covariates were allowed to covary. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

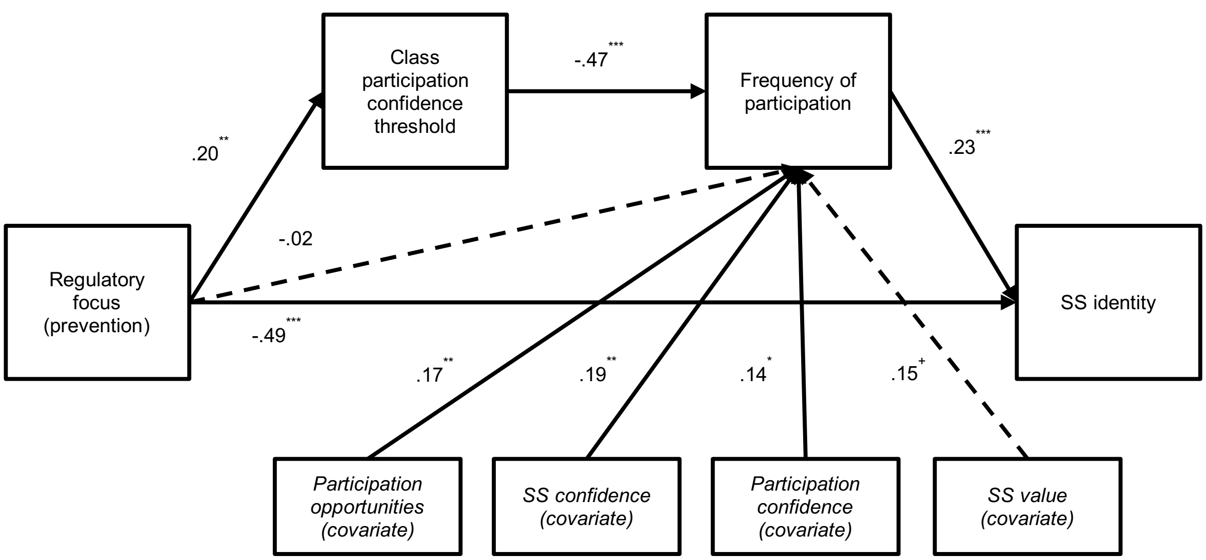
I will again caveat the following results, noting the poor fit statistics of the model. Direct effects of regulatory focus on math identity and class participation confidence threshold, and confidence threshold on frequency of participation were significant in this model. The addition of the covariates rendered the effect of regulatory focus on frequency of participation nonsignificant. With respect to indirect effects, results revealed a significant indirect effect of regulatory focus on frequency of participation via CPCT-2 ($\beta = -.14$, $SE = .05$, $p = .003$).

However, there was no significant indirect effect of regulatory focus on math identity (see Table 6).

The final mediation model I explored was regulatory focus to domain identity in social science (depicted in Figure 9), which had similarly poor fit statistics as the other mediation models. Nevertheless, as in the prior models, direct effects of regulatory focus on social science identity and class participation confidence threshold were significant, as well as the direct effect of confidence threshold on frequency of participation, and participation frequency on identity. Explorations of indirect effects (see Table 6) highlighted a significant indirect effect of regulatory focus on frequency of participation via CPCT-2 (β = -.09, *SE* = .04, *p* = .014). As results confirmed in the math domain as well, there was no significant indirect effect of regulatory focus on identity.

Figure 9

Mediation Model of Regulatory Focus to Domain Identity (Study 1 – Social Science)



Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .692, RMSEA = .236 CI [.193, .282], and SRMR = .090. All covariates were allowed to covary. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Discussion

The primary goal of Study 1 was to explore the questions addressed in the pilot studies, but with a sample of students who could reflect on current (rather than past) experiences in their math classes. Another aim of this study was to be able to test gender differences, which could not be explored in the pilot studies due to a lack of statistical power. Below, I present a summary of the results of Study 1, followed by a discussion of the study's limitations and how they were addressed in Study 2. The interpretation of the key findings across both studies will be offered in the *General Discussion and Conclusions* chapter.

Research Aim 1: Gender and Domain Differences in Constructs

My hypotheses regarding gender differences in the mean levels of the constructs in the model were mostly supported. As a reminder, I expected to find higher levels of stereotype threat, prevention focus, and confidence thresholds for women in the math domain, but not necessarily in social science. On the other hand, I hypothesized that results would show lower frequencies of participation, sense of belonging, identity, and career interest for women in math, and perhaps higher levels of these variables for women in social science.

Predictors of Participation

As expected, women's levels of stereotype threat were higher in math than in social science, while men displayed equal levels of stereotype threat in both domains. My hypotheses were also supported regarding mean level differences in regulatory focus and confidence thresholds. Overall, participants exhibited more of a tendency toward a prevention focus in math as opposed to social science, which aligns with the general perception of math as a subject that heightens vigilance due its objective nature. As hypothesized, domain differences in prevention focus appeared to affect women more than men; while men did not exhibit any difference in prevention focus across domains, women reported higher levels of prevention in math than in social science. With respect to confidence thresholds, as expected, thresholds in general were higher in the math domain, and women exhibited higher thresholds than men in both domains.

Participation

In terms of participation frequency, participants reported a higher frequency of participation in social science compared to math classes. Consistent with this finding, participants also perceived more opportunities for participation in their social science courses than their math courses. Thus, the overall higher levels of participation in social science could be a reflection of the greater number of opportunities for participation in that domain, compared to math. Although there were no significant gender differences in participation frequency, an additional exploratory analysis, operationalizing participation frequency as a ratio of frequency to opportunities, did reveal a gender difference, with men participating more frequently in both domains. Further, women reported changing their minds about participating more frequently than men in both domains.

Outcomes of Participation

With respect to the outcome variables in the model (i.e., sense of belonging, domain identity, and career interest), I hypothesized that women would display lower mean values than men in math but equal or higher mean values in social science. Analyses of sense of belonging supported my hypotheses, with women exhibiting lower levels of belonging than men in math and equal levels of belonging in social science. For domain identity and career interest, also in line with my hypotheses, women demonstrated lower levels of both variables in the math domain and higher levels in the social science domain. Similar patterns were exhibited for domain confidence and value—women displayed lower math and higher social science confidence and value than men.

In sum, my hypotheses regarding gender differences in the mean levels of the constructs in the theoretical models were largely supported, with women in the math domain exhibiting higher levels of the predictor variables (stereotype threat susceptibility, prevention focus, and confidence thresholds) and lower levels of the outcome variables (participation frequency out of total opportunities, sense of belonging, math identity, and STEM career interest) than men.

Research Aim 2: Relations Between Constructs in the Theoretical Model

The second aim of this study was to test how well the relations between the measured variables aligned with the hypothesized theoretical model, as presented in Figure 1, in both math and social science. I did not expect domain to moderate the pathways in the model, just the intercept values of the variables, so I did not expect to find differences in the models across domains.

In both domains, the bivariate correlations were mostly aligned with the hypothesized model. More specifically, a tendency toward a prevention focus was positively correlated with both measures of confidence threshold and negatively correlated with participation frequency. Interestingly, a prevention focus was also strongly and negatively associated with sense of belonging, identity, and career interest in both domains. As predicted, in both domains, confidence thresholds were negatively correlated with participation frequency and positively associated with frequency of changing one's mind. Relations in the second half of the model also aligned with hypotheses in both domains: frequency of participation was positively correlated with both sense of belonging and domain identity, and both belonging and identity exhibited strong associations with career interest.

Despite the support for the majority of the posited relations between variables, correlations involving stereotype threat susceptibility were less aligned with the hypotheses. More specifically, the scenario-based measure of stereotype threat susceptibility (ST-1), which is the primary measure of this construct, was not significantly correlated with any proximal variables in either domain. The secondary measure (ST-2), however, did exhibit some significant correlations. Although it was not associated with regulatory focus in either domain, it was positively correlated with both measures of confidence threshold and with frequency of changing one's mind about participating. In the math domain only, it was also negatively associated with participation frequency. It should be noted that ST-2 was a measure perhaps capturing one's awareness of gender stereotypes more than experience of them. I will discuss potential reasons for these observed correlations in the general discussion of both studies in Chapter 6.

The correlation analysis was followed by path analyses, which allowed me to examine multiple relations at once while controlling for other variables. Although the final model for social science (Figure 5) appeared to corroborate the theoretical model, some of the pathways were not significant in the math model (Figure 3). One such pathway was the direct relation between regulatory focus and participation frequency, which was only marginally significant in the math domain. Notably, however, the total effect of this relation in the targeted mediation models was actually greater in math than in social science (as displayed in Table 6). Another relation that was not significant in the final model for math was the path from frequency of participation to math identity. A potential explanation for this result is that in this final model, regulatory focus is a significant direct predictor of math identity and therefore may account for part of the association between those two variables.

The final relation that did not emerge as significant in the math domain—but was significant in social science—was the path from sense of belonging to career interest. It is worth noting that this path coefficient was also nonsignificant in the *initial* path model in the math domain. There are a few potential reasons for this result. First, it is possible that this effect is less reliable than others in the model, and with a larger sample, it would emerge as significant. It is also possible that there is a mediation effect that was unaccounted for in the present model, such that math identity mediates the relation between sense of belonging and STEM career interest. In other words, the extent to which a student feels like they belong in a domain may influence how much they identify with it, which in turn may impact their career aspirations. Because the final model did not measure any directional paths between belonging and identity, I was not able to test this possibility. The final potential explanation is that this finding is evidence of a lack of an effect of belonging on career interest in the math domain, in which case, belonging can still be considered an important outcome of participation in and of itself. Even in the case that it is not associated with career interest, sense of belonging also predicts other critical outcomes in the math domain that were not measured in this study, such as student learning, even when controlling for other motivational variables (e.g., Morrow & Ackermann, 2012).

Research Aim 3: Mediation Effects

Finally, the targeted mediation analyses (see Figures 6-9) revealed some interesting findings, particularly with respect to indirect effects and the influence of the addition of motivational covariates such as confidence and value. All four models displayed similar significant direct effects—specifically, the effects of a prevention focus on confidence thresholds, sense of belonging, and identity; confidence thresholds on participation frequency; and participation frequency on sense of belonging. Notably, the direct effect of regulatory focus on participation frequency emerged as nonsignificant in all of them. However, in all of the models, confidence thresholds significantly mediated this relation, even taking into account the influence of motivational covariates. This is an important result because it shows that not only was the confidence threshold a novel motivational factor accounting for some variance as a predictor of participation; it actually accounted for more variance than any other motivational variable in all models. The indirect effect of regulatory focus on more distal outcomes (i.e., belonging and identity) was less robust. In the math domain only, the indirect effect of regulatory focus on belonging was significant, although the standardized coefficient was a similar size in the social science domain.

Limitations and Implications for Study 2

As with all research, there are some important limitations to note with respect to interpreting the findings of Study 1. Below, I present some initial considerations; additional limitations are described in the final chapter of the dissertation.

One limitation of Study 1 was its relatively small sample size of 161 participants. More specifically, the sample was too small to provide adequate power for the path analyses testing the theoretical model. In general, path analysis requires more participants per estimated variable than other multivariable techniques such as multiple regression, given that there are often multiple parameters estimated per variable. Specifically, methodologists recommend having 20 participants for each estimated parameter (Kline, 2015). Thus, the relatively low sample size is likely one explanation for the poor model fit statistics in this study.

Relatedly, perhaps the most substantial limitation of Study 1 was the poor fit of the path models—i.e., the data did not fit the theoretical models well. In general, it is inappropriate to interpret parameter estimates prior to establishing an adequate model fit (Bauer & Curran, 2012).

Thus, all of the findings from the path models for this study should be considered preliminary. One explanation for the generally poor model fit statistics is that the proposed theoretical model is incorrect and does not accurately explain the predictors and outcomes of participation frequency. Alternatively, it is possible that the models were merely too ambitious (i.e., attempting to estimate too many parameters) in light of the relatively small sample size. Study 2 attempted to address this limitation via the recruitment of a larger sample.

A final caveat of Study 1 was that the data were collected in one survey and therefore were cross-sectional. This was a limitation because path models assume a component of temporal ordering (Land, 1969). Relatedly, it was not possible to confirm the directionality of certain relations that could be theoretically justified to operate in more than one way—e.g., the association between participation and belonging. Further, since motivational constructs in this study were measured at a single point in time, they could have captured either stable individual differences or only a snapshot of students' changing beliefs and concerns at a particular moment in time, limiting the robustness of the findings. It is also possible that participants' reports of their participation frequency could have been subject to memory decay, since the survey requested general estimates of participation frequency and did not ask participants to reflect on particular class days. The second study was designed to address this limitation by introducing multiple timepoints for data collection and asking participants to report their participation on the same day they attended a given class.

CHAPTER 5: STUDY 2

Study 2 employed a daily diary methodology (e.g., Gunthert & Wenze, 2012) to explore the same theoretical model investigated in Study 1 in a more naturalistic manner. In contrast to the first study, the goal of this methodology was to assess students' participation behaviors reported on the basis of their specific experiences on particular days in their math and/or social science class. Thus, students' responses reflected their participation-related behavior and thoughts on the days when they completed a diary entry, therefore presumably offering a more accurate representation of motivation and participation than was provided by Study 1.

The general structure of Study 2 was four phases: an enrollment form, followed by an initial survey (Phase 1), a period of daily diaries surveys over two weeks within a period of several weeks (Phase 2), and a final survey at the end of the semester (Phase 3).

Method

Participants

Undergraduate students were recruited from two selective northeastern universities at the beginning of the Fall 2021 semester. As in Study 1, potential participants received an email directly from me inviting them to participate in a survey-based study over the course of the semester. However, in contrast to Study 1, rather than emailing math course listservs, I received IRB approval to contact the students directly. At both universities, I provided a list of suitable "math courses" (including heavily math-focused courses such as computer science and calculus-based physics) to the appropriate administrative office and received lists containing the names and email addresses of enrolled students in return.

Eligibility Requirements

In order to be eligible for participation in Study 2, prospective participants had to meet the following requirements based on their responses to the enrollment form. As in Study 1, participants must have: (1) been enrolled in both a qualified undergraduate math and social science course during the current semester; (2) identified as male or female; and (3) not participated in any prior versions of the study, including the pilot studies and Study 1. Additionally, participants were asked to indicate which weeks they expected to be off campus, unavailable to participate in the study, or not attending classes during the upcoming semester. Students who did not indicate being available for at least two weeks during the semester were excluded from the study.

Exclusion Process Following Enrollment Form

The enrollment form received a total of 821 responses – 78.9% from University A and 21.1% from University B. However, 121 did not complete the form, 11 did not consent to participation, one was not an undergraduate student, and five were under 18 years old. Ten additional participants were removed because their gender was not male or female. Next, while checking for duplicate responses, I found that 20 participants had completed the form twice, so I removed these erroneous responses. An additional seven participants were excluded due to indicating a lack of availability during the upcoming semester (i.e., fewer than two weeks of availability). While comparing to participant lists from pilot studies and Study 1, I found that 67 respondents had participated in a prior version of the study. Finally, I examined the courses students listed and excluded 98 students who were missing either a math or social science course. I assume that those missing a math course had dropped their math course between the time that the course enrollments were downloaded and the enrollment form was sent to students. Thus, out of the initial responses, 481 were from participants eligible to be invited to Study 2.

Of the 481 students invited to complete the Phase 1 survey, 377 participated in it. Twenty-six participants did not complete the full survey and therefore were excluded. An additional 24 participants were unavailable and/or had exams during at least four of the five remaining weeks in the semester designated as potential weeks for the daily diaries (Phase 2), and were therefore removed from the study. Thus, 327 participants were invited to Phase 2.

Of these 327 participants invited to Phase 2, 312 participated in at least one of the eight possible daily surveys. Through a review of the data, it was determined that two of these participants were enrolled in a course that had become asynchronous since the Phase 1 survey, and three had dropped one of their courses after the Phase 1 survey. I also reviewed participants' diary responses to ensure that they had provided data at least once for each domain (math and social science). Twenty-seven additional participants were excluded for missing data from at least one of the domains. Finally, 11 participants were removed due to data issues in the Phase 1 survey that were not identified until the final data review process. More specifically, six failed attention checks, four demonstrated evidence of straight-lined responses, and one participant failed to meet a requirement of spending at least 10 seconds on the first CPCT page they were presented (a criterion set due to the significant amount of text on the page and applied to all pilot and dissertation studies).

Description of Participants

The final sample of Study 2 is considered to be the 269 participants remaining after these exclusions. All of these participants have at least some data for Phases 1 and 2, and 246 (91.4%) of these participants also participated in Phase 3. The majority of these 269 final participants were from University A (77.0%), White (69.1%), and reported having at least one parent with a master's degree or higher (56.1%). Participants were also mostly female (61.0%). Math majors comprised 23.4% of the sample, with 55.6% of them male. As in Study 1, approximately half of the participants were STEM majors (46.5%). If students listed more than one academic major, they were labeled as math or STEM majors as long as one of their majors met this requirement. STEM majors again included math majors, as well as non-mathematical STEM majors such as biology and environmental science. Table 7 summarizes the demographic characteristics of the final sample of Study 2.

To determine the smallest effect sizes that the study was powered to detect, I conducted a series of sensitivity analyses using G*Power (v3.1.9.6; Faul et al., 2007). Two of the analyses indicated that, for the gender x domain ANOVAs reported below, the study had 80% power to detect interactions and within-participant main effects with an effect size $f = .086$, with alpha set at .05, and assuming the default correlation among repeated measures of .5. A third analysis indicated that, for the same ANOVAs, the study had 80% power to detect between-participant main effects of $f = .148$. And, a final analysis indicated that, for uncorrected zero-order correlations, the study had 80% power to detect an association of $r = \pm.170$, with alpha set at .05 (two-tailed). I did not conduct power analyses for the other types of analyses reported below.

Table 7

Summary of Demographic Information for the Final Sample of Study 2

	All participants (n = 269)	Female only (n = 164)	Male only (n = 105)
	Mean (SD) or N (%)		
University			
University A	207 (77.0%)	122 (74.4%)	85 (81.0%)
University B	62 (23.0%)	42 (25.6%)	20 (19.0%)
Major			
Math major	63 (23.4%)	28 (17.1%)	35 (33.3%)
STEM major	125 (46.5%)	75 (45.7%)	50 (47.6%)
Race			
White	186 (69.1%)	116 (70.7%)	70 (66.7%)
Black or African American	12 (4.5%)	8 (4.9%)	4 (3.8%)
Asian	50 (18.6%)	26 (15.9%)	24 (22.9%)
Mixed race and other	21 (7.8%)	14 (8.5%)	7 (6.7%)
Hispanic	23 (8.6%)	15 (9.1%)	8 (7.6%)
Age (in years)	19.62 (1.09)	19.60 (1.12)	19.65 (1.05)
Class year			
Freshman	103 (38.3%)	60 (36.6%)	43 (41.0%)
Sophomore	80 (29.7%)	53 (32.3%)	27 (25.7%)
Junior	54 (20.1%)	30 (18.3%)	24 (22.9%)
Senior	32 (11.9%)	21 (12.8%)	11 (10.5%)
Highest parent education level			
Some high school	4 (1.5%)	4 (2.4%)	0 (0.0%)
High school	39 (14.5%)	26 (15.9%)	13 (12.4%)
Associate's degree	9 (3.3%)	7 (4.3%)	2 (1.9%)
Bachelor's degree	66 (24.5%)	31 (18.9%)	35 (33.3%)
Master's degree	94 (34.9%)	66 (40.2%)	28 (26.7%)
Med school, law school, PhD	57 (21.2%)	30 (18.3%)	27 (25.7%)

Note. STEM majors include all math majors, mathematical science majors, and non-mathematical STEM majors such as biology and environmental science.

Procedure

As referenced above, the general structure of Study 2 included three phases following the brief enrollment form: an initial survey (Phase 1), eight short daily surveys (Phase 2), and a final survey (Phase 3). The surveys in each phase, which were all conducted via Qualtrics, are described in detail below.

Given the longitudinal nature of this study, in order to ensure continued participation throughout all phases, participants were offered incentives based on the extent of their commitment to participating in each of the study components. All payments were in the form of Amazon eGift Cards distributed at the end of the study. Participants were compensated $3 for completing the Phase 1 survey, $1 for each of the daily Phase 2 surveys (a maximum of $8), and $4 for completing the Phase 3 survey. In addition, participants who completed all ten surveys were entered into a raffle to win one of ten $50 Amazon eGift cards.

Phase 0: Enrollment Form

As detailed in the *Participants* section above, students in math classes were first invited to complete a short enrollment form at the beginning of the fall semester. The first page of the enrollment form contained the study's consent form, which included information about the purpose and structure of the study. Once students consented to participating in the study, they were directed to the next page of the enrollment form, which asked for their name, email address, student ID number, gender, academic major(s), class year, and whether they were at least 18 years old. Participants were also asked to list their current course titles as precisely as possible. In contrast to Study 1, I did not require participants to enter course numbers in addition to course titles in this study, in an effort to avoid the attrition evidenced at that point of the enrollment form in the first study. Finally, participants were asked to indicate which weeks of the semester they anticipated being "off campus, unavailable to participate in the study, or not attending classes" during the upcoming semester. The intended purpose of gathering this information was to ensure that participants had enough anticipated availability throughout the semester to be able to commit to participating in the study.

After the enrollment form was closed, I reviewed each entry to determine which participants were eligible to be invited to Phase 1. Prospective participant eligibility was determined based on the criteria outlined in the *Participants* section above. After reviewing each student's courses, I created a file listing their eligible math and social science courses. The final lists contained 73 unique math and 176 unique social science courses. However, as in Study 1, participants were likely enrolled in more than this number of actual course sections, even though I did not collect section-level data and therefore was not able to verify the exact section each student was in. Because students were enrolled in multiple different courses with many different instructors, multilevel modeling was not necessary for the analyses.

Phase 1: Initial Survey

Two weeks after the enrollment form was first distributed, in early October 2021, eligible participants were sent the Phase 1 survey, which was a 15-20 minute questionnaire very closely resembling the main survey in Study 1. As in the Study 1 survey, participants were first prompted to select which of their math courses and which social science course would be their focus throughout the duration of the study. Again, they were encouraged to consider which courses would offer the most opportunities for hand-raising, based on their experience in the course so far. This survey was also structured identically to the Study 1 survey, with two blocks of questions (math and social science) presented in a random order, such that 50.9% of the final sample completed the math block first. Also like the Study 1 survey, within each of the two blocks, measures were presented in an order designed to align with the psychological pathways proposed in Figure 1, with items randomized within each questionnaire.

Phase 2: Daily Diaries

Phase 2 was administered over five weeks during the semester (within a six-week period from October to December, as one week was a holiday week and was not included in the study). Participants received eight two- to five-minute daily diaries surveys distributed on Monday through Thursday over two separate weeks during the semester. The Monday through Thursday schedule was selected in order to account for the most common undergraduate course schedules (Monday-Wednesday-Friday and Tuesday-Thursday) while also avoiding Friday evenings, when very low response rates would be expected. At 5pm each evening during their participation weeks, participants received an email asking them to respond to the brief survey. Reminder emails were sent at 9pm and 1am (the following calendar day) to participants who had not yet completed the daily survey, and the questionnaire was locked at 5am the following calendar day.

On the first page of every daily survey, participants were asked to indicate whether or not they attended each of their particular math and social science courses that day (e.g., "Did you attend Multivariable Calculus today?"). Due to the ongoing COVID-19 pandemic, students were presented with four possible response options: (1) *Yes, I attended in person*; (2) *Yes, I attended online synchronously (i.e., live, at the same time as the instructor and other students)*; (3) *Yes, I attended online asynchronously (i.e., not live, in a self-paced manner)*; and (4) *No, I did not attend this class today.* Then, for each of the classes they attended either (1) in person or (2) online synchronously, they were directed to answer questions (described in the *Measures* section below) relating to those courses. If participants indicated that they (3) attended online asynchronously or (4) did not attend a particular course, they would not be presented with questions relating to that course on that particular day.

I used information from the enrollment form and Phase 1 survey in order to determine on which two weeks each participant would receive the daily surveys. In the Phase 1 survey, one question asked participants to indicate the dates of midterm exams, final exams, and in-class presentations on their syllabi throughout the semester, in order to gauge weeks when participants might not have any opportunities to participate in class. After the Phase 1 survey was closed, I reviewed the weeks participants listed as exam/presentation weeks and, combining those weeks with the ones indicated in the enrollment form as weeks off-campus or out of class, I created lists of which weeks each participant *would* be available to receive the daily diaries surveys. Some participants were unavailable for at least four of the five weeks of Phase 2 and were therefore excluded from the remainder of the study. Next, I assigned each of the remaining 327 participants to two of the five diary weeks, taking care to make selections with at least one full week between them whenever possible. Four participants were only available on two consecutive weeks and therefore were set to receive the Phase 2 surveys on those weeks. Most commonly, in 43.9% of cases, participants were sent the two Phase 2 surveys two weeks apart.

Phase 3: Final Survey

In the last week of the semester, participants were sent the Phase 3 survey, which was a five to 10-minute survey distributed to all participants who had completed at least one of the Phase 2 surveys. This survey was closed on the last day of classes, before final exams began, to ensure that all participants completed it during the same point in the semester (i.e., rather than having some participants complete it while still enrolled in the course and others complete it while taking final exams or while home on the break between semesters). As in the other surveys, participants were presented with blocks of questions presented in a random order (math vs. social science).

Measures

A complete inventory of measures implemented in each phase of Study 2 is presented in Appendix D, including those on the enrollment form, which is not described below. Each of the phases also included some version of the cognitive strategies measures that were first presented in Study 1. As in Study 1, because these measures were intended to address research questions unrelated to this dissertation, they will not be discussed further.

Phase 1 Measures

Phase 1 contained measures nearly identical to those utilized in the main Study 1 survey. More specifically, the survey assessed participants' *stereotype threat susceptibility* (again with two different measures), *domain-specific regulatory focus*, *participation confidence thresholds* (again with two different measures), *participation frequency*, *confidence* in and *perceived value* of math and social science, *sense of belonging* in math and social sciences domains, *math and social sciences identity*, and *interest in pursuing a math- and/or social sciences-related career*. Like the Study 1 survey, the Phase 1 survey also contained demographic questions and attention and suspicion checks. See Appendix D for the specific items comprising each measure.

Modifications from Study 1. Because of the overwhelming similarities to the Study 1 survey, I will not describe the measures in detail here. However, there were some notable changes from the Study 1 survey that warrant being addressed. First, the second measure of stereotype threat susceptibility (adapted from Deemer et al., 2016) was revised for clarity. Instead of reading "my gender group" or "members of the opposite gender," I used Qualtrics logic to pipe in the appropriate gender relevant to each participant. For example, the item "I am afraid of being negatively evaluated by members of the opposite gender in this class" was changed to "I am afraid of being negatively evaluated by men in this class" for female participants.

Another modification from Study 1 was to the sense of belonging scale adapted from Good et al., 2012. The 18 items included in Study 1 were reduced to 14 items in Study 2, due to concerns that some of the items were too closely related to participation behaviors. The items that were removed were: "I wish I could fade into the background and not be noticed," "I try to say as little as possible," "I enjoy being an active participant," and, "I wish I were invisible."

Finally, as described in the *Procedure* section above, the Phase 1 survey also contained a question regarding the dates of in-class academic tasks participants expected to engage in throughout the semester, with the goal of anticipating when they might not have as many opportunities for participating as usual.

Phase 2 Measures

Due to its daily format, the second phase of data collection consisted of very abbreviated versions of scales used in the prior surveys, with the aim of keeping the survey very short and minimizing participant attrition.

Stereotype Threat. An abbreviated version of the Likert-type scale adapted from the Stereotype Threat in Science Scale – Gender by Deemer et al. (2016) was created from the original 7-item scale used in Phase 1. I selected the four items based on which ones seemed to best capture the felt experience of stereotype threat. Although the scenario-based measure is a more suitable measure of this affective experience, I included this shorter scale (without the vignette text required for the scenario-based measure) for brevity. The stem of the items was also adapted to align with the daily reflective nature of the diaries: "Immediately before or during [course name] today...," and the items themselves were adjusted to begin with "I felt..." Responses to each item were averaged to create a composite score.

Regulatory Focus. Regulatory focus orientation was assessed with a scale similar to the one adapted from Browman et al. (2017) and used in the prior surveys, modified to ask participants for recent, retrospective reports of their regulatory focus in class that day. Specifically, participants were asked to respond to prompts about what they were focused on immediately before or during each of their classes that day (e.g., "not making mistakes" or "achieving my hopes and aspirations for this class"). As in the Study 1 and Phase 1 surveys, regulatory focus orientation was operationalized as a participant's average prevention score minus their average promotion score, such that positive values indicated a predominant prevention focus.

Confidence Thresholds. Daily participation confidence thresholds were assessed with a modified version of the CPCT-2 Likert-type scale used in Phase 1. Participants were prompted to reflect on their participation behaviors in each of their math and social science courses on a given day, rating their level of agreement with three items asking about their experiences in class that day (e.g., "I was only willing to participate in class today when I was absolutely certain that my potential contribution was valid.")

Classroom Participation Behaviors. Rather than reporting an aggregate experience of class participation as in prior surveys, participants in Phase 2 were prompted to share actual frequency with which they participated in their classes on a given day in an open response (i.e., whole number) format. Specifically, participation frequency was measured with four questions (based on those used in Study 1) probing students' experiences that day. Two of the items in each domain prompted students to report the number of times they raised their hand to (1) answer a question posed by the instructor; or to (2) make a comment, share their experience, or provide an explanation. The other two items in each domain prompted students to report the number of times they changed their mind about engaging in each behavior described in (1) and (2).

Opportunities for Participation. The number of opportunities for participation on a given day was measured with a single open response question prompting participants to enter the number indicating approximately how many times their instructor posed questions to the class that day. Responses were averaged to create a composite score.

Sense of Belonging. An abbreviated version of the Math Sense of Belonging Scale (Good et al., 2012) was also included in the Phase 2 surveys. Four items from the 14-item version were selected based on how well they represented the construct. The measure began: "During [course name] today...," with items such as "...I felt like an outsider," and "I felt accepted." Responses were averaged to create a composite score.

Phase 3 Measures

In the third phase of data collection, the primary focus was on capturing a final measure of each of the primary outcome variables of the theoretical model (i.e., sense of belonging, identity, and career interest). In addition, two sets of questions asked about the courses students had been referencing in the surveys throughout the semester. See Appendix D for the full list of items.

Course Schedule. Participants were first asked to report on which day(s) of the week they had each of their math and social science courses. This question had not been included in an earlier survey, with the aim of reducing the number of questions participants were asked. However, it was added at this phase in order to assist with data analysis.

Gender Composition of Courses. Two questions asked participants to report the general gender composition of their math and social science courses ("mostly male," "mostly female," or "about evenly split"). Both math and social science courses were reported to be "about evenly split" by 49.2% and 51.2% of participants, respectively. An additional 35.0% of respondents shared that their math classes were mostly male, while 35.8% said their social science courses were mostly female.

Sense of Belonging. Participants' sense of belonging in math and social science settings was assessed with the 14-item version of Good et al.'s (2012) Math Sense of Belonging Scale that was used in the Phase 1 survey. Responses were averaged to create a composite score.

Math and Social Science Identity. Identification with the math and social sciences domains was measured using the 3-item scale adapted from Lesko and Corpus (2006) and Nosek et al. (2002) and used in the Study 1 and Phase 1 surveys. Responses were averaged to create a composite score.

Career Interest. Finally, interest in pursuing a math-related career and a social sciences-related career was measured using the same three-item subscale from the Educational Career Interest Scale in Science, Technology, and Mathematics (Oh et al., 2013) that was used in both the Study 1 and Phase 1 surveys. Responses were averaged to create a composite score.

Results

As in Study 1, both SPSS 27 and Stata 17 were used to conduct the analyses for Study 2.

Initial Exploration and Operationalization of Phase 2 Data

The unique format of the data collected during Phase 2 warranted some additional exploration before conducting analyses more explicitly related to the study research questions.

Completion Rates of Daily Diaries

Out of the 269 final participants, 156 (58.0%) completed all eight daily surveys. On average, participants completed 6.87 (85.9%) of the daily surveys they received. However, it is critical to note that a completed daily survey did not necessarily equate to meaningful data. Because each daily survey began with questions regarding whether participants attended each class on a given day, in many cases, participants would only answer these attendance questions (e.g., "No, I did not attend this class today") and then be directed to the final page of the survey. Thus, although instances such as this one constituted a "complete" daily survey, they did not contain any useful survey information.

For this reason, I also investigated the number of "diary entries" per participant—i.e., the number of times they indicated that they did, in fact, attend their math and/or social science class that day and then answered the ensuing questions. In theory, the maximum possible entries per participant was 16 (i.e., 8 daily surveys x 2 domains), although it is very unlikely—if not impossible—that a student would have both of their designated courses scheduled on all four days, Monday through Thursday. The number of complete entries for the final sample ranged from 2 to 12, with 8 entries being the most frequent (for 64 participants, or 23.8%). The average number of entries across all participants was 6.15 (3.20 math entries and 2.95 social science entries). I also examined the number of entries across the two weeks participants received the surveys. Participants completed an average of 3.39 entries on the first week they received the surveys, with a mode of 4 complete entries (for 110, or 40.9%, of participants). On the second week they received the surveys, participants completed an average of 2.76 entries, with the most frequent count again being 4 complete entries (for 88, or 32.7%, of participants). As was addressed in the *Participants* section above, participants were excluded from the study if they did not complete at least one entry per domain. In other words, as long as participants responded to at least one of the eight daily surveys and addressed both their math and social science courses, they were retained in the study.

I also compared participants' actual entries to when I would expect them to have a complete entry based on their course schedule reported in the Phase 3 survey. The most common math course schedules were Monday-Wednesday-Friday or Monday-Wednesday (63.4%) and Tuesday-Thursday (32.5%). Social science courses were also most frequently held on Monday-Wednesday-Friday or Monday-Wednesday (47.6%), followed by Tuesday-Thursday (40.2%). Other, less frequent, course schedules included combinations such as Monday-Tuesday-Thursday and Wednesday-Friday. Based on participants' reported class schedules, 79 participants reported missing at least one class they should have attended based on their course schedules (with an average of 0.74 class sessions missed across all 269 participants and 2.51 sessions missed among those who missed at least one). Perplexingly, there were also 58 instances of participants reporting that they *did* attend a given course on a date that did not align with their reported course schedule. Based on emails from participants and other anecdotal evidence, I hypothesize that student and instructor absences, schedule revisions, and courses changing from in-person to online-asynchronous during the study were largely due to the COVID-19 pandemic and the ensuing turmoil from quarantining guidelines and unexpected illness. Unfortunately, I did not ask any further questions of participants when they reported not attending class on a given day, so this potential explanation is only a conjecture.

Phase 2 Data Operationalization

The puzzling nature of students' actual course schedules during the COVID-19 pandemic influenced how the Phase 2 data was ultimately operationalized. In other words, the pandemic had introduced extensive additional variables that would influence students' expected experience, and it was beyond the scope of the present dissertation to determine exactly how to specify missingness on the level of class day for multiple imputation. Thus, I decided to create composite measures of the diaries data, averaging and totaling variables across days to find amalgamated measures by week and domain.

For each week and domain (e.g., Week 1 math, or Week 2 social science), as well as each domain overall (i.e., math across both weeks and social science across both weeks), I calculated average scores for each participant, including total counts for participation frequency and opportunities for participating. I also computed a percentage of participation frequency: the total times a participant reported raising their hand across the diaries period, divided by the total reported number of questions asked by the instructor. For the majority of Study 2 analyses, this "percent participation" variable is used as the measure of students' frequency of participation. In rare cases where participants reported a greater count of participation frequency than count of opportunities for participation, I considered this erroneous data as missing data to be imputed.

Addressing Missing Data

Description of Missing Data

As briefly discussed above, due to the longitudinal nature of Study 2, there was some missing data. Across all cells in the data (i.e., from all three phases), 7.07% of Study 2 data was missing (entirely from Phases 2 and 3). Although I decided to use composite measures of the Phase 2 data, with variables averaged across days, some of the final participants were still missing data for one of the week-domain combinations (e.g., Week 1 social science). In total, 38 participants (14.1%) were missing data for one week-domain set, and 37 participants (13.8%) had no data for two week-domain sets (e.g., both Week 1 social science and Week 2 math data). Because final participants were required to have data for each domain in at least one week, the maximum number of missing week-domain combinations for a given participant was two. In addition to data missing at the week-domain level, 13 participants (4.8%) had some missing item-level data, typically due to closing out of the daily diaries survey before completing the final items. While missing data in Phase 2 was mostly due to missed diaries entries, Phase 3 data was missing mostly as a result of participant attrition—23, or 8.55%, of the final participants of Study 2 did not participate in the Phase 3 survey. Based on initial analyses, the missingness in both phases appeared to be fairly random (i.e., not related to particular characteristics of the participants).

Handling Missing Data

Two techniques—multiple imputation and full information maximum likelihood—were used to address missing data in Study 2. Missing data theorists have suggested that these techniques will generate similar results when the same auxiliary variables are taken into account and a sufficiently large number of imputations is conducted (Collins et al., 2001). Due to its superiority in yielding unbiased parameter estimates for structural equation models (Enders & Bandalos, 2001), full information maximum likelihood (FIML) was employed during the estimation of path models in Stata. Thus, when it was available (i.e., during path model estimation), FIML was utilized. For all other Study 2 analyses, I used multiple imputation to impute missing data for missing average weekly data (following Thomas et al., 2016). SPSS 27 was used to create 20 imputed data sets, following guidance by Graham et al. (2007).

Descriptive Analyses

The same initial analyses performed for Study 1 were conducted for Study 2. More specifically, after creating composite variables and conducting multiple imputation to address missingness, I examined descriptive statistics for each construct, including measures of central tendency, skewness, kurtosis, and reliability of scales. Table 8 displays the means and standard deviations of all primary variables of interest in all three phases of Study 2, arranged by domain and participant gender.

With respect to normality, all variables in Phases 1 and 3 appeared to be fairly normal. However, some Phase 2 variables were collected as count data and therefore did not reflect a normal distribution. More specifically, participation frequency and opportunities for participation were both measured as whole numbers in Phase 2, with nearly half of respondents (45.4%) reporting that they participated zero times across all diaries in their math classes, and 30.9% reporting that they participated zero times in their social science courses. Logarithmic and inverse transformations were considered for this zero-inflated data but did not yield improved distributions. Because participation frequency was operationalized as a percentage for the majority of Study 2 analyses, and path analyses tend to be quite robust to non-normality of a single variable (Knief & Forstmeier, 2021), the zero inflation of this variable was not ultimately addressed for the present dissertation. The implications of this decision are addressed in the *Limitations* section of the final chapter.

Table 8

Descriptive Statistics for Key Variables in Study 2

	Math						Social Science					
	All participants		Female only		Male only		All participants		Female only		Male only	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Phase 1												
ST-1 (diff. score)	0.18	0.69	0.34	0.75	-0.06	0.48	0.25	0.79	0.33	0.85	0.13	0.67
ST-2 (Likert)	3.12	1.51	3.93	1.33	1.85	0.71	2.69	1.15	3.04	1.18	2.13	0.86
Reg. focus (diff. score)	0.84	1.01	0.90	0.98	0.75	1.07	0.61	1.10	0.44	1.12	0.88	1.02
CPCT-1 (slider)	75.20	23.87	78.88	22.55	69.45	24.93	68.10	24.27	70.47	24.00	64.40	24.46
CPCT-2 (Likert)	4.43	1.21	4.65	1.12	4.08	1.27	4.01	1.26	4.15	1.27	3.80	1.22
Participation frequency	2.84	1.16	2.71	1.19	3.04	1.09	3.50	1.31	3.41	1.39	3.64	1.16
Freq. of changing mind	3.54	1.20	3.77	1.18	3.17	1.15	3.50	1.17	3.70	1.18	3.19	1.08
Sense of belonging	4.03	0.83	3.87	0.83	4.27	0.77	4.32	0.78	4.37	0.81	4.23	0.73
Domain identity	3.42	1.32	3.33	1.36	3.56	1.27	3.47	1.27	3.64	1.33	3.20	1.13
Career interest	4.40	1.31	4.27	1.34	4.61	1.24	4.40	1.28	4.59	1.26	4.10	1.26
Participation opportunities	4.32	1.16	4.44	1.16	4.12	1.13	5.33	0.82	5.36	0.82	5.28	0.83
Domain confidence	4.43	1.06	4.36	1.12	4.56	0.96	4.66	0.90	4.75	0.92	4.53	0.88
Participation confidence	3.79	1.12	3.69	1.19	3.96	0.99	3.91	1.10	3.83	1.16	4.04	1.00
Confidence diff. score	-0.63	1.84	-0.96	1.87	-0.12	1.67	-0.10	1.90	-0.32	2.02	0.24	1.66
Domain value	3.65	0.70	3.59	0.73	3.73	0.64	3.64	0.71	3.75	0.71	3.46	0.67

	Math						Social Science					
	All participants		Female only		Male only		All participants		Female only		Male only	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Phase 2												
ST-2 (Likert)	2.41	1.17	2.83	1.19	1.74	0.63	2.05	0.84	2.21	0.84	1.79	0.73
Reg. focus (diff. score)	-0.08	0.57	-0.07	0.54	-0.11	0.55	-0.20	0.69	-0.25	0.68	-0.12	0.63
CPCT-2 (Likert)	4.09	1.04	4.23	1.05	3.86	0.99	3.77	1.05	3.89	1.06	3.59	1.02
Participation frequency	2.07	3.13	1.68	2.66	2.70	3.70	4.52	7.20	3.46	4.37	6.17	9.97
Participation opportunities	16.17	15.81	15.75	15.13	16.84	16.95	25.58	38.67	21.90	19.39	31.33	56.76
Percent freq. of participation	0.11	0.15	0.10	0.14	0.14	0.16	0.16	0.17	0.15	0.16	0.19	0.19
Sense of belonging	4.24	0.82	4.12	0.88	4.43	0.62	4.52	0.76	4.52	0.73	4.51	0.78
Phase 3												
Sense of belonging	4.18	0.86	4.04	0.91	4.40	0.72	4.52	0.77	4.55	0.78	4.47	0.72
Domain identity	3.60	1.23	3.58	1.29	3.64	1.17	3.51	1.13	3.64	1.14	3.29	1.09
Career interest	4.15	1.31	4.04	1.36	4.32	1.27	4.36	1.21	4.57	1.11	4.05	1.33

Research Aim 1: Gender and Domain Differences in Constructs

For Study 2, Research Aim 1 (i.e., *to explore whether there are gender and domain differences in the mean levels of all constructs in the model*) was addressed using a combination of mixed ANOVAS (for Phase 1) and random effects regression models (for Phases 2 and 3). For Phase 1 constructs, I conducted mixed ANOVAs similar to those employed in Study 1, with a 2 (Gender: male vs. female) x 2 (Domain: math vs. social science) design. Phases 2 and 3 necessitated an alternative analysis approach due to the multiply imputed nature of the data. Rubin's Rules (Rubin, 2004) for pooling parameter estimates following analysis of imputed data rely on the assumption that the pooled result follows a normal distribution. Thus, since analysis of variance uses an F-distribution, pooling F-tests is not possible with multiply imputed data. As an alternate approach, I utilized random effects regression models, predicting each construct of interest as a function of gender, domain, and their interaction, which accounted for the nesting of domain within participant and used multiple imputation for the missing data.

Phase 1

Table 9 summarizes the results of the mixed ANOVA for each construct in Phase 1, including significant differences across gender, domain, and their interaction. Despite the unequal sizes of the gender groups, Levene's test of homogeneity of variance was again generally nonsignificant, indicating that the unequal sample sizes did not violate the assumptions of the ANOVAs (Wickens & Keppel, 2004). Because the results are fully detailed in Table 9, below I focus primarily on differences that emerged as significant.

Table 9

Summary of Gender x Domain ANOVA Results for Study 2 – Phase 1 Constructs

	Gender			Domain			Gender x Domain		
Construct	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Phase 1									
ST-1 (diff. score)	15.77	<.001	.056	2.93	.088	.011	3.49	.063	.013
ST-2 (Likert)	156.33	<.001	.369	20.56	<.001	.072	74.49	<.001	.218
Regulatory focus (difference score)	2.36	.125	.009	3.49	.063	.013	10.39	.001	.037
CPCT-1 (slider)	8.99	.003	.033	20.59	<.001	.072	1.28	.258	.005
CPCT-2 (Likert)	13.58	<.001	.048	19.75	<.001	.069	1.51	.221	.006
Participation frequency	4.65	.032	.017	64.79	<.001	.195	0.33	.567	.001
Frequency of changing mind	20.14	<.001	.070	0.11	.744	<.001	0.34	.558	.001
Sense of belonging	2.58	.110	.010	13.93	<.001	.050	19.21	<.001	.067
Domain identity	1.16	.282	.004	0.03	.873	<.001	6.55	.011	.024
Career interest	0.47	.494	.002	0.64	.424	.002	11.40	<.001	.041
Participation opportunities	4.76	.030	.018	147.83	<.001	.356	2.10	.149	.008
Domain confidence	0.01	.921	<.001	4.35	.038	.016	6.13	.014	.022
Participation confidence	4.59	.033	.017	1.93	.166	.007	0.18	.668	.001
Confidence difference score	13.81	<.001	.049	14.30	<.001	.051	1.15	.285	.004
Domain value	1.76	.186	.007	.752	.387	.003	9.94	.002	.036

As results indicated in Study 1, the analysis of both measures of stereotype threat revealed a main effect of gender ($F(1, 265) = 15.77$, $p < .001$, $\eta_p^2 = .06$ for the scenario-based measure, ST-1, and $F(1, 265) = 156.33$, $p < .001$, $\eta_p^2 = .37$ for the Likert-type measure, ST-2), such that on average, women reported higher levels of stereotype threat ($M = 0.34$, $SE = .05$; $M = 3.49$, $SE = .08$, for the ST-1 and ST-2 measures, respectively) than men ($M = 0.04$, $SE = .06$; $M = 1.99$, $SE = .09$, respectively). Also similarly to the results of Study 1, analysis of the Likert-type measure of stereotype threat susceptibility (ST-2) revealed a main effect of domain, $F(1, 265) = 20.56$, $p < .001$, $\eta_p^2 = .07$, with participants on average reporting higher levels of stereotype threat in the math domain ($M = 2.89$, $SE = .07$) compared to social science ($M = 2.58$, $SE = .07$). The main effects for ST-2, but not ST-1, were qualified by a significant interaction, $F(1, 265) = 74.49$, $p < .001$, $\eta_p^2 = .22$. Pairwise comparisons revealed that female students reported significantly higher levels of threat in math ($M = 3.93$, $SE = .10$) than in social science ($M = 3.04$, $SD = .09$), $p < .001$. The opposite relationship was present for male students, who expressed higher levels of stereotype threat susceptibility in social science classes ($M = 2.13$, $SE = .08$) than math classes ($M = 1.85$, $SE = .07$), $p < .001$. These significant differences were also present when I examined gender differences within each domain ($p < .001$). I also conducted additional ANOVAs with math major status and its interactions with each other predictor included. As in Study 1, results revealed no main effect of major, $F(1, 157) = 0.48$, $p = .490$, $\eta_p^2 = .002$, and no significant interactions involving major.

Regulatory focus (i.e., tendency toward a prevention focus) did not exhibit a main effect of gender, $F(1, 265) = 2.36$, $p = .125$, $\eta_p^2 = .01$, or domain, $F(1, 265) = 3.49$, $p = .063$, $\eta_p^2 = .01$. However, a significant interaction was present, $F(1, 265) = 10.39$, $p = .001$, $\eta_p^2 = .04$. Pairwise comparisons revealed higher levels of a prevention focus for women in math ($M = 0.90$, $SE =$

.08) than in social science (*M* = 0.44, *SE* = .09), $p < .001$, whereas men did not display a significant difference in tendency toward a prevention orientation across domains (math: *M* = 0.75, *SE* = .10; social science: *M* = 0.88, *SE* = .11), $p = .386$. As in Study 1, this interaction was mostly driven by gender differences in social science ($p = .001$), as opposed to in math ($p = .237$).

As in the results of Study 1, both measures of class participation confidence threshold, CPCT-1 and CPCT-2, exhibited main effects of gender and domain. CPCT-1 was revealed to be significantly higher for women (*M* = 74.67, *SE* = 1.62) than for men (*M* = 66.92, *SE* = 2.02), $F(1, 265) = 8.99$, $p = .003$, $\eta_p^2 = .03$. In addition, on average, thresholds for all participants were higher in math (*M* = 74.16, *SE* = 1.47) compared to social science (*M* = 67.44, *SE* = 1.51), $F(1, 265) = 20.59$, $p < .001$, $\eta_p^2 = .07$. Analysis of CPCT-2 also revealed main effects of gender, $F(1, 265) = 13.58$, $p < .001$, $\eta_p^2 = .05$, and domain, $F(1, 265) = 19.75$, $p < .001$, $\eta_p^2 = .07$, such that women exhibited higher thresholds than men (females: *M* = 4.40, *SE* = .08; males: *M* = 3.94, *SE* = .10), and participants on average reported significantly higher thresholds in math (*M* = 4.37, *SE* = .07) than in social science (*M* = 3.98, *SE* = .08). As in the results of Study 1, these main effects were not qualified by significant interactions for either measure of confidence threshold (see Table 9).

Unlike in Study 1, analysis of self-reported participation frequency did show a main effect of gender, $F(1, 265) = 4.65$, $p = .032$, $\eta_p^2 = .02$, with men exhibiting a higher frequency of participation (*M* = 3.34, *SE* = .10) than women (*M* = 3.06, *SE* = .08). There was also a significant main effect of domain, $F(1, 265) = 64.79$, $p < .001$, $\eta_p^2 = .20$, such that students reported participating more frequently in social science classes (*M* = 3.52, *SE* = .08) than in math classes (*M* = 2.88, *SE* = .07). I also conducted the same exploratory analysis I performed in Study 1, operationalizing participation frequency as a ratio of frequency of participation to opportunities for participation. Analysis of this exploratory variable again revealed a significant gender difference, $F(1, 265) = 5.71$, $p = .018$, $\eta_p^2 = .02$, with men ($M = 0.74$, $SE = .03$) participating more frequently than women ($M = 0.65$, $SE = .02$) when participation opportunities were taken into account.

Analysis of frequency of changing one's mind about participating revealed a main effect of gender, $F(1, 265) = 20.14$, $p < .001$, $\eta_p^2 = .07$, with female participants ($M = 3.74$, $SE = .08$) reporting a higher frequency of changing their mind about participating than male participants ($M = 3.18$, $SE = .09$). With respect to perceived opportunities for participation, results indicated main effects of both domain, $F(1, 265) = 147.83$, $p < .001$, $\eta_p^2 = .36$, and gender, $F(1, 265) = 4.76$, $p = .030$, $\eta_p^2 = .02$. More specifically, participants reported significantly more opportunities to participate in their social science courses ($M = 5.32$, $SE = .05$) than their math courses ($M = 4.28$, $SE = .07$), and female students perceived more opportunities to participate overall ($M = 4.90$, $SE = .06$) than male students ($M = 4.70$, $SE = .07$).

Patterns of mean differences in the outcome variables of the theoretical model (i.e., sense of belonging, identification with the domain, and career interest) mirrored the results of Study 1. More specifically, the only significant main effect was a main effect of domain for sense of belonging, $F(1, 265) = 13.93$, $p < .001$, $\eta_p^2 = .05$, such that participants reported a higher sense of belonging in social science courses ($M = 4.30$, $SE = .05$) than math courses ($M = 4.07$, $SE = .05$). Also in line with the results of Study 1, significant interactions were revealed for all three outcome measures. A significant interaction for sense of belonging, $F(1, 265) = 19.21$, $p < .001$, $\eta_p^2 = .07$, revealed that women experienced a lower sense of belonging in the math domain ($M = 3.87$, $SE = .06$) than in the social science domain ($M = 4.37$, $SE = .06$), $p < .001$, whereas men experienced similar levels of belonging in both domains (math: $M = 4.27$, $SE = .08$; social science: $M = 4.23$, $SE = .08$), $p = .677$. This interaction was driven mostly by gender differences in math ($p < .001$) rather than social science ($p = .138$).

Pairwise comparisons intended to explore the significant domain-by-gender interaction for domain-specific identity ($F(1, 265) = 6.55$, $p = .011$, $\eta_p^2 = .02$) revealed that women reported marginally significantly lower levels of math identity ($M = 3.33$, $SE = .10$) than social science identity ($M = 3.64$, $SE = .10$), $p = .056$, and another marginally significant difference in identity was present for men (math: $M = 3.56$, $SE = .13$; social science: $M = 3.20$, $SE = .12$), $p = .083$. The primary difference was between men and women in the social science domain ($p = .006$), not the math domain ($p = .168$). With respect to career interest, a significant interaction ($F(1, 265) = 11.40$, $p < .001$, $\eta_p^2 = .04$) showed that male participants, on average, reported higher interest in math-related ($M = 4.61$, $SE = .13$), as opposed to social science-related ($M = 4.10$, $SE = .12$), careers, $p = .008$, whereas women expressed the opposite preference (social science: $M = 4.59$, $SE = .10$; math: $M = 4.27$, $SE = .10$), $p = .040$. Within-domain explorations also revealed significant gender differences in both math ($p = .036$) and social science ($p = .002$) career interest.

The final set of ANOVAs examined domain and gender differences in mean levels of confidence and value. With respect to general domain confidence, there was a significant main effect of domain, $F(1, 265) = 4.35$, $p = .038$, $\eta_p^2 = .02$, such that participants overall reported higher confidence levels in social science ($M = 4.64$, $SE = .06$) than in math ($M = 4.46$, $SE = .07$), $p = .038$. This main effect was qualified by a significant interaction ($F(1, 265) = 6.13$, $p = .014$, $\eta_p^2 = .02$), which revealed that women reported significantly lower levels of math confidence ($M = 4.36$, $SE = .08$) than social science confidence ($M = 4.75$, $SE = .07$), $p < .001$, whereas male participants expressed similar levels of confidence in both domains (math: M = 4.56, SE = .10; social science: M = 4.53, SE = .09), p = .803. Within social science, women exhibited marginally higher confidence than men (p = .051), while no significant gender difference was present in math (p = .127). Unlike in Study 1, where there were no mean differences in participation-specific confidence, a main effect of gender was revealed in Study 2, $F(1, 265) = 4.59$, $p = .033$, $\eta_p^2 = .02$, with men (M = 4.00, SE = .09) reporting significantly higher levels of participation confidence than women (M = 3.76, SE = .07). As in Study 1, analyses of the difference score measure of confidence (i.e., participation-specific confidence over and above one's typical confidence threshold) revealed main effects of gender ($F(1, 265) = 13.81$, $p < .001$, $\eta_p^2 = .05$) and domain ($F(1, 265) = 14.30$, $p < .001$, $\eta_p^2 = .05$). On average, men reported positive confidence difference scores, such that their typical level of confidence was above their typical confidence threshold (M = 0.059, SE = .15), whereas women reported negative scores (M = -0.64, SE = .12). With respect to domain differences, participants on average reported a wider, negative gap between their confidence and threshold in their math courses (M = -0.54, SE = .11) than in their social science courses (M = -0.04, SE = .12). Finally, analyses of domain-specific value revealed a significant interaction of domain and gender, $F(1, 265) = 9.94$, $p = .002$, $\eta_p^2 = .04$, with men reporting significantly higher math value (M = 3.73, SE = .07) than social science value (M = 3.46, SE = .07), p = .011, and women reporting marginally significantly higher value for social science (M = 3.75, SE = .05) than for math (M = 3.59, SE = .05), p = .068. This interaction was driven primarily by differences in social science (p = .001) rather than math (p = .114).

Phases 2 and 3

Table 10 displays the coefficients, standard errors, and significance values of the random effects regression models conducted in order to explore mean differences in Phases 2 and 3 while taking into account the multiple imputation.

Table 10

Summary of Random Effects Regression Results for Study 2 – Phases 2 & 3 Constructs

	Gender			Domain			Gender x Domain		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Phase 2									
ST-2 (Likert)	1.09	.13	<.001	0.05	.13	.688	-0.67	.16	<.001
Reg. focus (diff. score)	0.04	.09	.665	-0.02	.08	.858	-0.17	.11	.119
CPCT-2 (Likert)	0.38	.14	.009	-0.27	.13	.047	-0.08	.17	.647
Participation frequency	-1.02	.69	.137	3.48	.64	<.001	-1.70	.82	.038
Participation opportunities	-1.09	3.68	.768	14.50	3.73	<.001	-8.34	4.77	.081
Percent frequency of participation	-0.05	.02	.027	0.05	.02	.017	0.007	.02	.760
Sense of belonging	-0.31	.11	.004	0.08	.11	.501	0.33	.15	.024
Phase 3									
Sense of belonging	-0.35	.11	.001	0.07	.10	.508	0.44	.14	.002
Domain identity	-0.05	.15	.728	-0.35	.17	.044	0.41	.22	.061
Career interest	-0.27	.16	.090	-0.27	.18	.123	0.80	.23	<.001

Results of the random effects regression analyses for Phase 2 showed a significant effect of gender for the Likert-type measure of stereotype threat susceptibility, ST-2 (b = 1.09, $p <$ .001), with men exhibiting lower levels of stereotype threat overall than women. Although there was no main effect of domain (unlike in Study 1 and Phase 1 of Study 2), the interaction between gender and domain also emerged as a significant predictor of ST-2, b = -0.67, p < .001.

Surprisingly, neither gender, domain, nor their interaction emerged as significant predictors of regulatory focus. Mirroring the results of Study 1 and Phase 1 of the present study for the Likert-type measure of confidence threshold (CPCT-2), both gender ($b = 0.38$, $p = .009$) and domain ($b = -0.27$, $p = .047$) were revealed as significant predictors, with women exhibiting higher thresholds than men, and all participants reporting higher thresholds in the math domain than in the social science domain.

The random effects regression analyses for the participation variables (participation frequency, number of participation opportunities, and percent frequency of participation) also revealed interesting results. As in Study 1 and Phase 1, domain emerged as a significant predictor of perceived participation opportunities, $b = 14.50$, $p < .001$, with participants overall reporting a higher perceived number of opportunities for participating in their social science classes compared to their math classes. For participation frequency, domain ($b = 3.48$, $p < .001$) and the gender-by-domain interaction ($b = -1.70$, $p = .038$) emerged as significant predictors, such that participants overall reported participating more frequently in their social science classes compared to their math classes. Lastly, both gender ($b = -0.05$, $p = .027$) and domain ($b = 0.05$, $p = 0.017$) were revealed as significant predictors of percent frequency of participation (out of total opportunities to participate). Specifically, participants reported higher percentages of participation in their social science classes compared to their math classes, and men had higher percentages of participating in general compared to women.

Finally, I explored outcome measures in the model—namely, sense of belonging, domain identity, and career interest. Sense of belonging was measured in both Phases 2 and 3, while domain identity and career interest were measured only in Phase 3. In both phases, gender (Phase 2: $b = -0.31$, $p = .004$; Phase 3: $b = -0.35$, $p = .001$) and the domain-by-gender interaction (Phase 2: $b = 0.33$, $p = .024$; Phase 3: $b = 0.44$, $p = .002$) were both significant predictors of students' belonging. In both cases, female students reported a lower overall sense of belonging than their male peers. Domain was a significant predictor of domain-specific identity ($b = -0.35$, $p = .044$), with participants reporting lower levels of identification with the social science domain compared to the math domain. The interaction between gender and domain was also a marginally significant predictor of identity, $b = 0.41$, $p = .061$. Finally, as in Study 1, the domain-by-gender interaction was a significant predictor of career interest, $b = 0.80$, $p < .001$, with women reporting higher levels of interest in social science-related careers and men expressing more interest in math-related careers.

Research Aim 2: Relations Between Constructs in the Theoretical Model

I addressed Research Aim 2 (i.e., *to explore relations between constructs in the theoretical model, including the directionality of these relations*) in Study 2 with similar analyses as were utilized in Study 1. More specifically, I reviewed bivariate correlations between constructs within each domain and then used path analysis to further examine the directionality of the relations within greater models.

In order to take advantage of the longitudinal nature of Study 2, I utilized particular variables from each of the three phases in the analyses—namely, predictors and covariates from Phase 1 (at the start of the semester), the participation measure from Phase 2 (throughout the semester), and outcome variables from Phase 3 (at the end of the semester).

Bivariate Correlations

I first examined zero-order correlations between the key variables of interest across all three phases in the theoretical model, as well as posited covariates. Correlations between relevant constructs are summarized in Table 11 for the math domain and Table 12 for the social science domain. Below, I focus on relations between variables across all phases that are posited to be proximally related in the theoretical model. For a full set of correlations within each phase (which will not be addressed here), see Tables E1-E6 in Appendix E.

As in Study 1, the first correlations I explored were those between sets of two variables that were intended to measure the same underlying construct—specifically, the measures of stereotype threat susceptibility (i.e., ST-1 and ST-2) and confidence threshold (i.e., CPCT-1 and CPCT-2). Because of the concerns with ST-2 and CPCT-1 that I acknowledged earlier, I again decided to use ST-1 as the measure of stereotype threat susceptibility and CPCT-2 as the measure of confidence threshold for further analyses. In both domains, the two measures of stereotype threat susceptibility were significantly correlated with each other (math: $r = .374$, $p < .001$; social science: $r = .166$, $p = .006$). The measures of threshold were also strongly correlated in both math ($r = .704$, $p < .001$) and social science ($r = .634$, $p < .001$).

Table 11

Zero-Order Correlations Between Key Variables of Interest Across All Phases of Study 2 (Math)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
Primary Constructs														
1. ST-1 *(P1)*	1													
2. ST-2 *(P1)*	.374***	1												
3. Regulatory focus *(P1)*	.093	.010	1											
4. CPCT-1 *(P1)*	.129*	.231***	.184**	1										
5. CPCT-2 *(P1)*	.197**	.339***	.220***	.704***	1									
6. Percent freq. of participation *(P2)*	-.089	-.105^{+}	-.123*	-.179**	-.307***	1								
7. Sense of belonging *(P3)*	-.144*	-.308***	-.252***	-.225***	-.310***	.164*	1							
8. Domain identity *(P3)*	.048	.063	-.418***	-.100	-.121^{+}	.057	.390***	1						
9. Career interest *(P3)*	-.046	.028	-.441***	-.003	-.059	.047	.377***	.689***	1					
Covariates														
10. Domain confidence *(P1)*	.098	-.028	-.225***	-.143*	-.132*	.094	.456***	.544***	.479***	1				
11. Participation confidence *(P1)*	-.111^{+}	-.243***	-.093	-.146*	-.246***	.124*	.379***	.124*	.082	.260***	1			
12. Domain value *(P1)*	.018	.039	-.492***	-.057	-.100	.094	.294***	.586***	.613***	.526***	.135*	1		
13. Sense of belonging *(P1)*	-.101^{+}	-.329***	-.374***	-.184**	-.309***	.203***	.657***	.412***	.365***	.562***	.433***	.416***	1	
14. Domain identity *(P1)*	.055	.046	-.502***	-.081	-.125*	.108^{+}	.339***	.751***	.618***	.604***	.088	.717***	.480***	1

Note. *P1, P2,* and *P3* indicate Phases 1, 2 and 3, respectively. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 12

Zero-Order Correlations Between Key Variables of Interest Across All Phases of Study 2 (Social Science)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
Primary Constructs														
1. ST-1 *(P1)*	1													
2. ST-2 *(P1)*	.166**	1												
3. Regulatory focus *(P1)*	.055	-.063	1											
4. CPCT-1 *(P1)*	-.050	.175**	.044	1										
5. CPCT-2 *(P1)*	.042	.239***	.209***	.634***	1									
6. Percent freq. of participation *(P2)*	-.080	-.039	-.105^{+}	-.204***	-.320***	1								
7. Sense of belonging *(P3)*	-.077	-.121^{+}	-.286***	-.054	-.193**	.183**	1							
8. Domain identity *(P3)*	-.009	.146*	-.482***	-.023	-.173**	.192**	.461***	1						
9. Career interest *(P3)*	-.009	.169**	-.436***	-.004	-.143*	.187**	.423***	.685***	1					
Covariates														
10. Domain confidence *(P1)*	-.015	-.066	-.291***	-.022	-.198**	.167**	.382***	.408***	.343***	1				
11. Participation confidence *(P1)*	-.067	-.173**	-.064	-.133*	-.292***	.252***	.352***	.199**	.158*	.411***	1			
12. Domain value *(P1)*	-.009	.203***	-.425***	-.073	-.164**	.247***	.410***	.574***	.649***	.386***	.287***	1		
13. Sense of belonging *(P1)*	-.099	-.168**	-.327***	-.057	-.262***	.213***	.596***	.473***	.412***	.563***	.502***	.561***	1	
14. Domain identity *(P1)*	-.103^{+}	.159**	-.509***	.008	-.141*	.202***	.421***	.759***	.632***	.469***	.240***	.690***	.577***	1

Note. *P1, P2,* and *P3* indicate Phases 1, 2 and 3, respectively. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Next, I explored the zero-order correlations between variables I expected to be proximally related in the theoretical model. Neither stereotype threat susceptibility measure was associated with regulatory focus in either domain. However, both measures were positively associated with each measure of confidence threshold in the math domain (whereas in social science, only ST-2 was associated with the threshold measures). A prevention focus was positively correlated with CPCT-1 in math and CPCT-2 in both domains. A predominant prevention focus was also negatively correlated with participation frequency (although only marginally in the social science domain). Mirroring the results of the first study, regulatory focus was significantly negatively correlated with all outcome measures in the model (sense of belonging, identity, and career interest) in both domains. In both domains, each of the threshold measures had negative correlations with participation frequency. Frequency of participation was positively correlated with sense of belonging, but not identity, in the math domain, whereas it was significantly correlated with both in social science. Finally, in both domains, career interest exhibited strong correlations with sense of belonging and identity. Again, the majority of the posited associations between variables were significant and in the expected direction. In contrast to Study 1, there were some more significant correlations with stereotype threat susceptibility, although most were with ST-2, which had already been deemed an unsuitable measure.

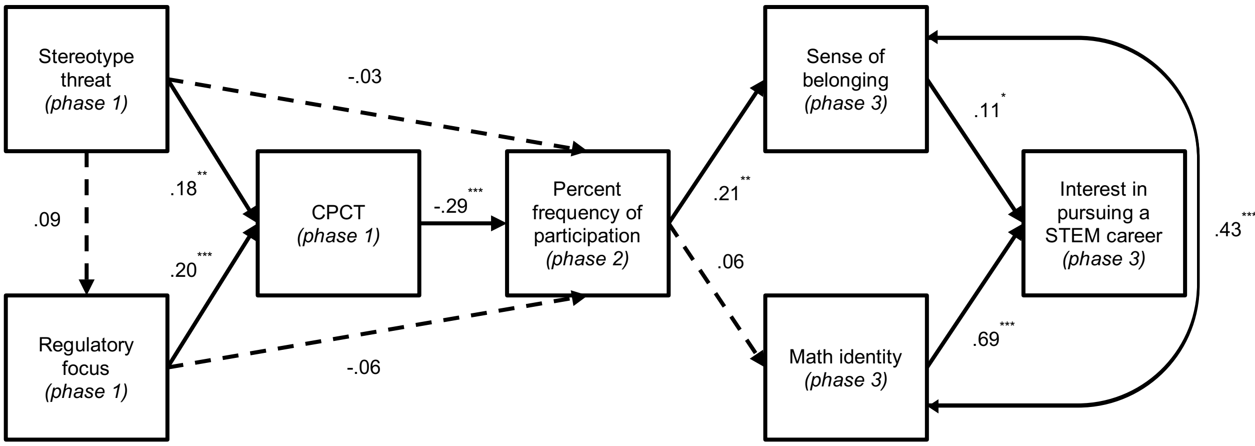
Path Analyses

As was done for the first study, path analyses were conducted in Stata 17 to test the theoretical model introduced in Figure 1. In contrast to Study 1, in which variables were measured at a single time point, Study 2 offered an opportunity to explore causality based on the timing of the data collection in each phase. All path models for the second study included predictors and covariates from Phase 1, the participation measure from Phase 2, and outcome variables from Phase 3. As previously discussed, in order to account for missing data in Phases 2 and 3—which was primarily a result of missing diary entries and attrition before Phase 3—path models were estimated using full information maximum likelihood (FIML). My overall aim in testing path models in Study 2 matched my approach in the first study: to balance theory, parsimony, and fit within the constraints of my data, while also prioritizing the comparison of models across domains and studies. Therefore, the "final" models in math and social science do not necessarily reflect the best possible models; they are merely a reflection of how the data fits the initially hypothesized model, with some minor adjustments. As with Study 1, in the process of determining the best approach while managing sample size constraints, I ran additional versions of each model that are not addressed in this dissertation but are available upon request. Model fit was again evaluated based on established guidelines for conventional fit indices (Hu & Bentler, 1999; Weston et al., 2008).

Math Domain. The first model I tested in the math domain was the theoretical model (as depicted in Figure 1). Figure 10 displays the results of this initial path analysis for the math domain. The fit indices for this initial model indicated a relatively poor fit to the data, with CFI = .754, RMSEA = .192 CI [.160, .225]. The χ^2 test was also significant, $\chi^2(10) = 108.663$, $p < .001$.

Figure 10

Initial Path Model (Study 2 – Math)

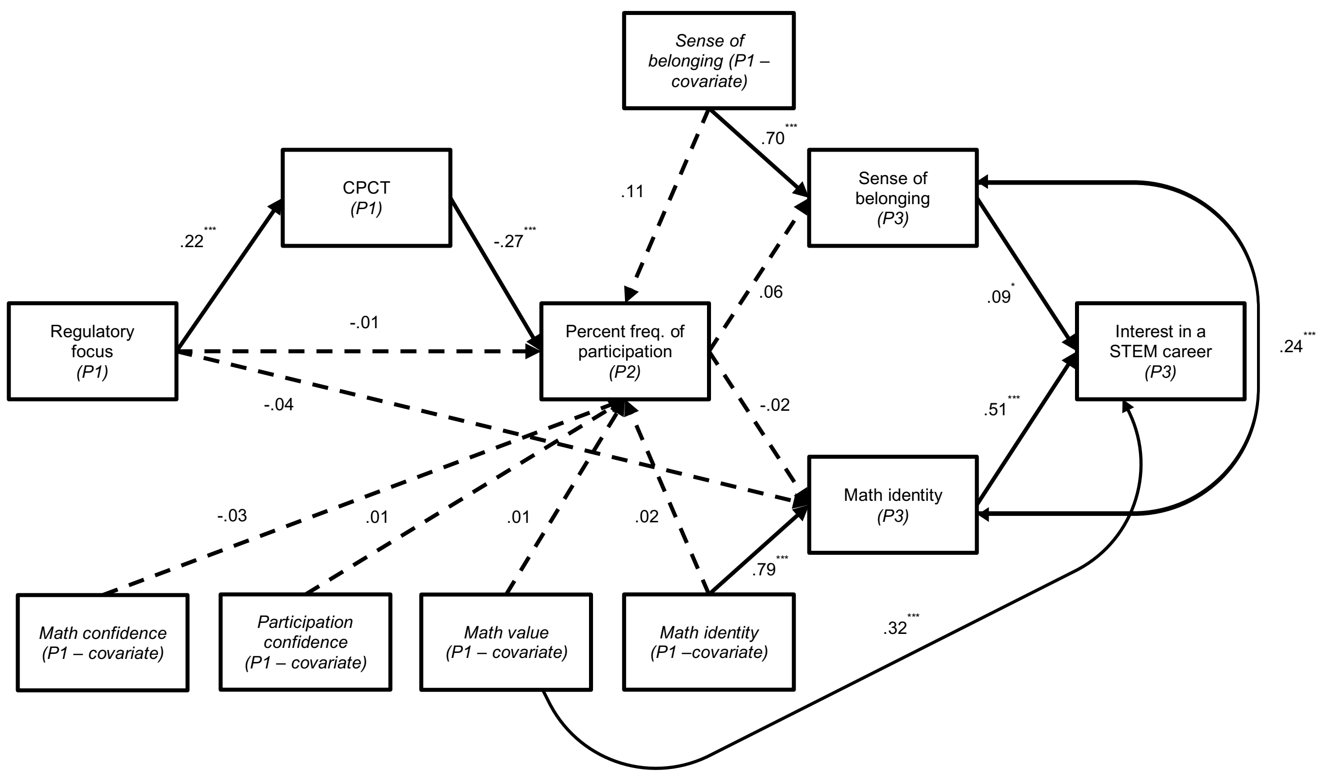


Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .754 and RMSEA = .192 CI [.160, .225]. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

In the next step of the analysis, I removed stereotype threat susceptibility from the model, due to the theoretical concerns with this construct, as well as in alignment with my effort to keep the Study 2 models consistent with those established in Study 1. Next, the modification indices suggested the addition of a pathway from regulatory focus to math identity, which mirrored the results from Study 1 for facilitating comparison of the models. An advantage of Study 2 over Study 1 was a larger sample size, which allowed me to include covariates directly in the models, rather than in separate targeted analyses. Therefore, at this stage, I added several covariates—namely, the expectancy and value covariates included in the mediation models of Study 1 (domain confidence, participation confidence, and domain value). I also included Phase 1 math identity and sense of belonging, with the goal of establishing potential causality. Although I originally planned to include major (i.e., math major versus other) as a covariate as well, it was ultimately not included because supplementary analyses revealed that it was not a predictor of participation in math ($\beta = 0.04$, $p = .461$) or social science ($\beta = -0.09$, $p = .134$) classes. The fit indices of this model were somewhat improved from the initial model (CFI = .896, RMSEA = .112 CI [.091, .134]), although the χ^2 test was still significant, $\chi^2(24) = 104.683$, $p < .001$. The final adjustment I made to the model was based on a modification index reflecting the largest expected parameter change that was also theoretically sound: the addition of a pathway from math value to STEM career interest. One of the components of math value is interest in the domain, so it makes intuitive sense that this would be a significant predictor of interest in pursuing a related career. The final model for the math domain is displayed in Figure 11.

Figure 11

Final Path Model (Study 2 – Math)



Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. "P1" indicates "Phase 1" measure. Fit indices: CFI = .941 and RMSEA = .086 CI [.063, .110]. All covariates were allowed to covary. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

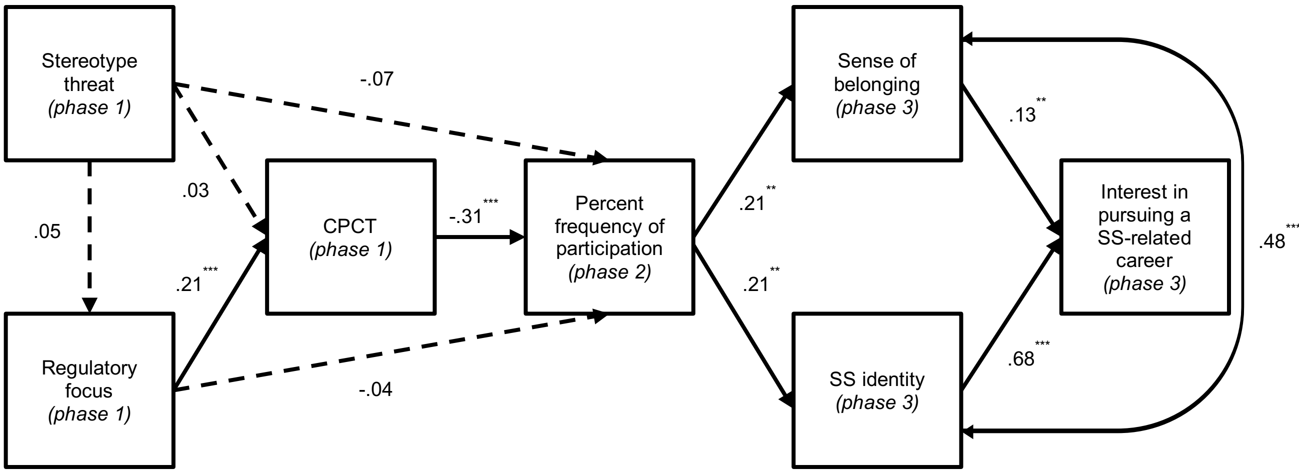
The final fit indices reflected a relatively good fit to the data, with CFI = .941 and RMSEA = .086 CI [.063, .110]. The χ^2 test was still significant, $\chi^2(23) = 69.013$, $p < .001$, but the ratio of the χ^2 to the degrees of freedom was equal to 3.00, which falls into an acceptable range for a complex model with a small sample size (Weston et al., 2008).

In the final path model, a prevention focus predicted a higher class participation confidence threshold ($\beta = 0.22$, $p < .001$). There was no direct effect of a prevention focus on frequency of participation, which mirrors the results of the targeted mediation analyses in Study 1 that included the expectancy and value covariates. Despite the inclusion of these covariates, a higher class participation confidence threshold remained a significant predictor of lower participation frequency ($\beta = -.27$, $p < .001$). Percent frequency of participation was no longer a predictor of Phase 3 sense of belonging and math identity when controlling for Phase 1 belonging and identity, which seemed to account for the majority of the variance as covariates. Lastly, both sense of belonging ($\beta = 0.09$, $p = .040$) and identification with the math domain ($\beta = 0.51$, $p < .001$) emerged as significant predictors of STEM career interest.

Social Science Domain. For social science, I also first tested the originally hypothesized model. Figure 12 depicts the results of this initial model in the social science domain.

Figure 12

Initial Path Model (Study 2 – Social Science)

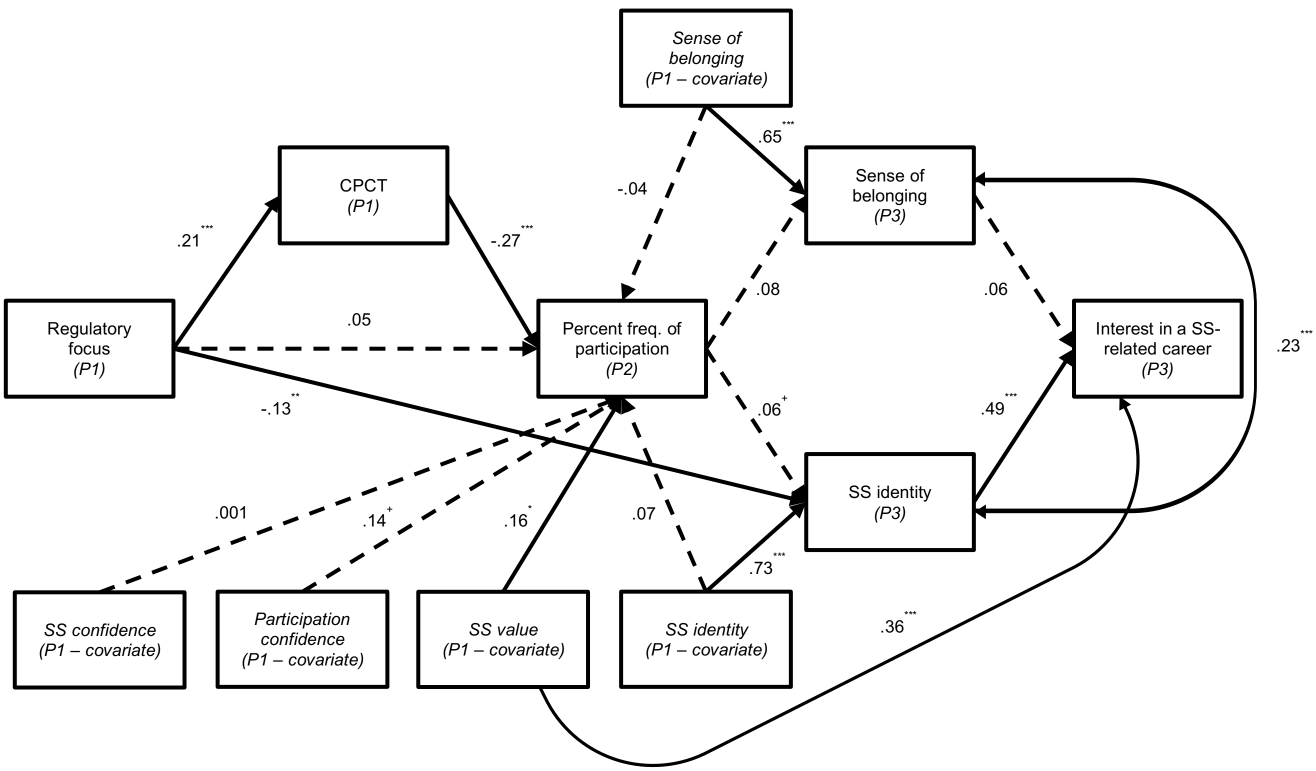


Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. Fit indices: CFI = .805 and RMSEA = .170 CI [.138, .204]. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

The fit indices for this initial model indicated a relatively poor fit to the data, with CFI = .805, RMSEA = .170 CI [.138, .204]. The χ^2 test was also significant, $\chi^2(10) = 87.458$, $p < .001$. The next steps of the path analyses closely resembled the changes made for the model in the math domain. More specifically, I again removed stereotype threat susceptibility from the model due to the theoretical and measurement-related concerns with the construct and the nonsignificant pathways leading from it to other variables in the model. As in the math domain, I also added a pathway from regulatory focus to identification with the social science domain, based on the modification indices. I also added the same covariates at this stage that were added to the math model: domain confidence, participation confidence, domain value, and Phase 1 social science identity and sense of belonging. The fit indices of this model were still poor but relatively better than those for the initial model (CFI = .908, RMSEA = .104 CI [.082, .126]), although the χ^2 test was still significant, $\chi^2(24) = 93.075$, $p < .001$. In one final adjustment to the model, I added a pathway from social science value to interest in a social-science related career, based on both a modification index and a desire to keep the models consistent across domains. The final model for social science is depicted in Figure 13.

Figure 13

Final Path Model (Study 2 – Social Science)



Note. All coefficients presented in the model are standardized. Solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. “P1” indicates “Phase 1” measure. Fit indices: CFI = .970 and RMSEA = .060 CI [.034, .086]. All covariates were allowed to covary. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

The fit indices for the final social science model reflected a good fit to the data: CFI = .970 and RMSEA = .060 CI [.034, .086]. Although the χ^2 test was still significant, $\chi^2(23) = 45.561$, $p = .003$, the ratio of the χ^2 value to the degrees of freedom was equal to 1.98, indicating adequate fit (Weston et al., 2008). The results of the social science path analyses closely resembled the results in the math domain. As in the math domain, a prevention focus predicted a higher class participation confidence threshold ($\beta = 0.21$, $p < .001$). Also mirroring the results in the math domain, there was no significant effect of a prevention focus on frequency of participation, which could be due to the inclusion of the expectancy and value covariates. However, class participation confidence threshold remained a significant predictor of lower participation frequency ($\beta = -.27$, $p < .001$), while percent frequency of participation was no longer a predictor of Phase 3 sense of belonging and social science identity when controlling for Phase 1 belonging and identity. Although sense of belonging was a predictor of career interest in the math domain, this relation was not significant for social science; social science identity, however, was still a significant predictor, $\beta = 0.49$, $p < .001$. Another difference from the final path model for math was that regulatory focus did remain a significant predictor of identity in the social science domain ($\beta = -.13$, $p = .003$), while this pathway became nonsignificant with the addition of Phase 1 identity in the math domain.

Research Aim 3: Mediation Effects

In Study 2, I was able to address Research Aim 3 (i.e., *to explore potential mediating mechanisms underlying the relations identified by the investigation of Research Aim 2*) in a more straightforward manner than in Study 1. While the small sample size of the first study necessitated separate mediation analyses of the primary portion of the model, Study 2 allowed for investigation of the indirect effects within the final models of each domain. I will note that because a direct pathway from regulatory focus to sense of belonging was not included in these final models, I was not able to investigate indirect effects from regulatory focus to belonging as I did in Study 1. Investigations of the relevant indirect effects from the Study 2 final path models are summarized in Table 13.

Results revealed a significant indirect effect of regulatory focus on percent participation frequency via CPCT-2 in the math domain, β = -.06, *SE* = .003, *p* = .005. This indirect effect was also significant in the social science domain, β = -.06, *SE* = .003, *p* = .006. This finding is particularly interesting due to its replication across domains and studies, despite the inclusion of covariates that are typically considered the most critical predictors of participation (i.e., confidence and value). I also examined indirect effects from regulatory focus to domain identity. Although, as described previously, there was a significant direct effect in the social science domain, β = -.13, *SE* = .04, *p* = .003, the indirect effect was not significant. This result could be partially due to the inclusion of identity at Phase 1 as a covariate in the model.

Table 13

Indirect Effects in Final Path Models of Study 2

	Direct effect		Indirect effect		Total effect	
	β	SE	*β*	SE	*β*	SE
Math						
Regulatory focus → percent participation frequency	-.01	.01	-.06**	.003	-.07	.01
Regulatory focus → domain identity [a]	-.04	.05	.001	.004	-.04	.05
Social Science						
Regulatory focus → percent participation frequency	.05	.01	-.06**	.003	-.01	.01
Regulatory focus → domain identity [a]	-.13**	.04	-.0003	.004	-.13**	.04

Note. All models include the following covariates: domain confidence, participation-specific confidence, domain value, Phase 1 belonging, and Phase 1 identity. [a] Indirect effect is via CPCT and participation. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Exploratory Analyses

In addition to the analyses intended to explore the primary research aims of the present dissertation, a series of more exploratory analyses was also conducted to investigate additional related research questions.

Gender-Related Variations in the Path Models

The present work was based on the hypothesis that the proposed theoretical model (Figure 1) would operate similarly for men and women, despite the posited mean-level differences in the constructs. However, as an exploratory investigation, I tested whether the relations among the variables differed for men and women by running a multiple group comparison, which allows for a more rigorous test of gender differences than including gender as a covariate in the model (Ohannessian et al., 2016).

In order to conduct the multiple group analysis, I first estimated an unconstrained model, with freely estimated path coefficients, variances, and covariances, for both men and women. Then, following Vandenberg and Lance (2000), I compared this set of models to another one in which parameters were constrained to be equal for both genders. Because of the reduced power inherent in exploring the model in smaller groups, the specific models I tested were the primary mediation models with covariates (as presented in Study 1). Specifically, I explored the mediation models linking regulatory focus to sense of belonging and domain identity via confidence threshold and percent participation frequency, including covariates of domain confidence, participation confidence, and domain value (but not Phase 1 belonging and identity).

Results of the omnibus comparison of the models with sense of belonging as the primary outcome for men and women in the math domain revealed model variance, $\chi^2(8) = 15.55$, $p = .049$, suggesting that gender differences exist in the nature of the model pathways. The omnibus test in the social science domain similarly suggested gender-related variance between the models, $\chi^2(8) = 21.67$, $p = .006$. The omnibus comparison of mediation models with domain identity as the outcome measure also revealed gender-related variance in both the math domain, $\chi^2(8) = 15.49$, $p = .050$, and social science domain, $\chi^2(8) = 16.84$, $p = .032$.

Next, I conducted more specific investigations of group variance of parameters. For both the model with sense of belonging and the one with domain identity, in the math domain, the parameters exhibiting a significant difference for men versus women were the paths from regulatory focus to percentage of participation ($\chi^2(1) = 5.911$, $p = .015$) and confidence threshold to percentage of participation ($\chi^2(1) = 5.531$, $p = .019$). For men, regulatory focus was a significant predictor of participation ($\beta = -0.22$, $p = .023$), but confidence threshold was not ($\beta = -0.11$, $p = .190$). For women, on the other hand, confidence threshold appeared to be the primary predictor of participation ($\beta = -0.38$, $p < .001$), while the direct effect of regulatory focus was not significant ($\beta = 0.08$, $p = .341$). In the social science domain, the only parameter suggesting a significant difference was for the path from regulatory focus to confidence threshold ($\chi^2(1) = 6.884$, $p = .009$). More specifically, this pathway was not significant for men ($\beta = 0.05$, $p = .580$) but was for women ($\beta = 0.32$, $p < .001$).

Gender Differences in Confidence Threshold Predicting Gender Differences in Participation

In addition to testing variation in the relations among variables for men versus women, I also investigated whether gender differences in certain variables in the model were associated with gender differences in other variables. Because no significant gender differences were found in regulatory focus—which emerged as the most distal factor in the final theoretical model—I was not able to test gender as a predictor in the model overall. Thus, in more targeted mediation analyses, I explored whether gender differences in confidence threshold predicted gender differences in participation frequency. Analyses revealed a significant indirect effect of gender on participation frequency via confidence threshold in the math domain, $\beta = -.07$, $SE = .007$, $p = .003$. Similarly, in social science, gender indirectly predicted participation frequency via confidence threshold, $\beta = -.04$, $SE = .007$, $p = .038$. Other proximal variables in the theoretical model did not exhibit the same pattern of gender differences and were therefore not included in this exploratory analysis.

Changes in Variables Across the Semester

I also examined potential changes in mean levels of variables from the beginning to the end of the semester. The only variables that were measured in both Phases 1 and 3 were sense of belonging, domain identity, and career interest. In order to investigate these possible changes, I conducted a series of paired samples *t*-tests comparing this set of six variables (three in each domain) across the semester. Sense of belonging significantly increased from Phase 1 to Phase 3 in both domains. In the math domain, students' reported sense of belonging at the beginning of the semester ($M = 4.03$, $SE = .05$) was significantly lower than their sense of belonging at the end of the semester ($M = 4.18$, $SE = .05$), $t(268) = -3.41$, $p < .001$. Similarly, sense of belonging in social science significantly increased over the course of the semester ($M = 4.32$, $SE = .05$ in Phase 1 to $M = 4.52$, $SE = .05$ in Phase 3), $t(268) = -4.49$, $p < .001$.

The other two variables—identity and career interest—only exhibited significant changes in the math domain. Students' math identity significantly increased from Phase 1 ($M = 3.42$, $SE = .08$) to Phase 3 ($M = 3.60$, $SE = .08$), $t(268) = -3.12$, $p = .002$. On the other hand, identification with social science did not change significantly across the semester ($M = 3.47$, $SE = .08$ in Phase 1 to $M = 3.51$, $SE = .07$ in Phase 3), $t(268) = -0.69$, $p = .493$. STEM career interest actually decreased across the semester ($M = 4.40$, $SE = .08$ to $M = 4.15$, $SE = .08$), $t(268) = 3.94$, $p <$

.001, whereas interest in a social science-related career remained relatively stable (M = 4.40, SE = .08 to M = 4.36, SE = .08), $t(268)$ = 0.53, p = .597.

Directionality of the Relation Between Belonging and Participation

Finally, in an effort to examine the potential reciprocal relationship between belonging and participation (e.g., Skinner & Belmont, 1993), I conducted time-lagged regression analyses. In the path models, controlling for belonging at the start of the semester had eliminated the effect of participation throughout the semester on belonging at the end of the semester. However, I was also interested in exploring changes from week to week within the daily diaries phase of the study. Thus, these analyses were intended to decipher, on a more granular level, the directionality of the relation between percent participation frequency and sense of belonging within Phase 2.

First, I examined whether while controlling for frequency of participation in Week 1 of Phase 2, belonging in Week 1 predicted participation in Week 2. Results in the math domain revealed that Week 1 belonging was not a predictor of Week 2 participation, whether controlling for Week 1 participation (β = -.001, $t(268)$ = -.009, p = .993) or not (β = .037, $t(268)$ = .501, p = .617). Similar results occurred in the social science domain: Week 1 belonging was not a predictor of Week 2 participation frequency, whether controlling for participation at Week 1 (β = .068, $t(268)$ = .950, p = .343) or not (β = .140, $t(268)$ = 1.930, p = .055).

Next, I explored the opposite potential directionality—i.e., whether participation was a predictor of belonging, as posited by the theoretical model introduced in Figure 1. More specifically, I tested whether controlling for belonging in Week 1, Week 1 participation frequency predicted sense of belonging in Week 2. Results revealed null effects when controlling for Week 1 sense of belonging in both math (β = .021, $t(268)$ = .391, p = .696) and social science

(β = .026, $t(268)$ = .465, p = .642). Notably, however, *without* controlling for sense of belonging in Week 1, percent participation frequency (Week 1) was a significant predictor of Week 2 sense of belonging for both the math (β = .152, $t(268)$ = 2.39, p = .017) and social science (β = .160, $t(268)$ = 2.53, p = .012) domains.

Discussion

Study 2 was intended to be an extension of the first study that allowed for additional exploration of relations among constructs. Along with measuring students' participation behaviors in a more naturalistic and valid manner than in the first study, the longitudinal design allowed for more rigorous causal investigations than could be conducted in the first study. The results of Study 2 both replicated and extended the findings from Study 1. Again, the below discussion provides a summary of this study's results; further interpretation will be included in the following chapter.

Research Aim 1: Gender and Domain Differences in Constructs

The majority of the findings related to gender and domain differences in Study 2 corroborated those of Study 1. Across phases, women exhibited higher levels of stereotype threat susceptibility, higher confidence thresholds, and a greater frequency of changing their mind about participating than men in both domains. Moreover, further mirroring Study 1, women's levels of belonging, career interest, confidence, and value were lower in math (and in some cases, higher in social science) than men's.

With respect to participation frequency, Study 2 provided an important clarification. In Study 1, the raw measure of participation frequency exhibited no gender difference, but a gender difference was present when (in an exploratory analysis) participation was measured as a ratio out of perceived participation opportunities. In Phase 2 of the second study, participation was operationalized as a percentage of participation frequency out of total opportunities. Analysis of this variable confirmed the results of the exploratory analysis of Study 1, showing that given the same number of opportunities, men tended to participate more than women in both domains. Further, in Phase 1, where opportunities to participate were not taken into account, analyses still revealed a higher participation frequency for men than for women in both domains.

In line with my expectations (and mirroring Study 1 findings), participants perceived a higher quantity of participation opportunities in social science than in math courses. An interesting departure from my hypotheses was that in Phase 1, women reported a higher perceived number of participation opportunities than men, whereas I did not expect to find a significant gender difference. There are two potential explanations for this finding. First, it could be a function of the types of courses within each domain in which men versus women tend to enroll, such that women tend to select courses that elicit more participation in the form of class conversation. In the present study, many students were enrolled in unique math and social science courses, which precluded the opportunity to investigate whether men and women perceived a different number of participation opportunities within the same courses. Another possibility is that because women generally have more hesitation about participating (as evidenced by the consistent gender differences in confidence thresholds), this reluctance may lead to heightened awareness of the number of opportunities they receive. In other words, students who are more reticent to participate may be more attuned to the number of opportunities they are failing to seize compared to students who do not have the same level of anxiety about participating.

Some other Study 2 findings were similar to Study 1 in one phase but different in another. For instance, in Phase 1, women again displayed higher social science identity and lower math identity than men. However, in Phase 3, participants overall appeared to be more identified with math than social science, and there were no significant gender differences. This is an interesting result that aligns with one of the exploratory analyses I conducted, in which I examined how math and social science identity changed across the semester. As math identity increased for participants on average, while social science identity remained stable, it is possible that the general increase in math identity throughout the semester for all participants contributed to reducing the gender differences observed at the start of the semester.

Another result that was different across phases was with respect to regulatory focus. Phase 1 analyses of regulatory focus were aligned with my hypotheses: women exhibited higher levels of prevention in math than in social science, while men did not exhibit any difference in prevention focus across domains. In contrast to my hypotheses, however, no gender or domain differences in prevention focus were present in Phase 2. A potential explanation for this null finding is measurement error. The Phase 2 regulatory focus scale attempted to assess participants' regulatory orientation on a daily basis and was ultimately averaged into a composite score. While I expected the aggregate measure to reflect the same regulatory orientation that would be measured by the more general regulatory focus measure, it is possible that the aggregation of the daily measures, combined with the operationalization as a difference score, eliminated meaningful variance.

Research Aim 2: Relations Between Constructs in the Theoretical Model

The zero-order correlations between constructs in Study 2 largely aligned with both the hypothesized theoretical model and the results of the correlations in Study 1. In contrast to Study 1, the scenario-based measure of stereotype threat (ST-1) was characterized by some significant associations with relevant variables. Although it was still not related to regulatory focus, it was positively associated with both measures of confidence threshold in the math domain only, suggesting the existence of some potential math-specific stereotype effects. ST-2 again emerged as more highly correlated with relevant variables, with stronger associations with confidence thresholds and a marginally significant negative relation to percent participation frequency in the math domain. Further discussion of the findings involving stereotype threat susceptibility will be in the final chapter of this dissertation.

An advantage of the path models of Study 2 over those of the first study was that the temporal ordering of the variables aligned with the directionality of the model. As a reminder, the predictors and covariates in the model were measured in Phase 1, participation was evaluated in Phase 2, and the outcome variables were assessed in Phase 3. Testing whether predictor variables measured at the start of the semester actually predicted participation throughout the semester was an exciting methodological and analytical improvement that gave more credence to conclusions from the cross-sectional data of the pilot studies and Study 1.

The final models for math and social science were quite similar to each other, with the exception of two parameters. First, sense of belonging was a significant predictor of career interest in the math domain, whereas the relation was nonsignificant for social science. This was an interesting contrast to Study 1, where the opposite result emerged: sense of belonging was a significant predictor of career interest only in social science. It is notable that the magnitude of the standardized effect in math was the same as it was in Study 1, where it was not significant, indicating that the lack of significance could have been related to a lack of power in the first study. The second difference between the final math and social science models was that while regulatory focus was a significant direct predictor of social science identity, it did not emerge as a predictor of math identity. In Study 1, this pathway was significant in both domains.

There were some additional differences from the Study 1 path models. First, the relation from participation to sense of belonging became nonsignificant in Study 2 in both domains once Phase 1 belonging was included as a covariate, as it most likely subsumed the majority of the variance of this relation. Similarly, participation and social science identity were also no longer related after the inclusion of Phase 1 identity as a covariate. In the case of math identity, its relation with participation frequency was nonsignificant both prior to and after the addition of the Phase 1 covariate. This lack of significance suggests the possibility that participation, as measured in the present study, may not meaningfully contribute to students' math identity. The final difference from Study 1 was the lack of direct effect from regulatory focus to participation frequency in social science. This can likely be explained by the inclusion of the expectancy and value covariates in the Study 2 final model.

It is worth noting that even with the addition of the Phase 1 covariates, the standardized effect sizes in each domain were relatively similar and not completely negligible, especially for the path from participation to sense of belonging. Thus, it is possible that significance would be detected with greater power. It is also important to point out that the time between Phases 1 and 3 was not very long, and the correlations between sense of belonging at both times was quite high. Thus, future analyses should explore mixed models that include the individual timepoints from the daily diaries portion of the study. On a more granular level, it is possible that I would find subtle changes in belonging from one class session to the next, as well as relations between participation and belonging between class sessions or across consecutive weeks.

Research Aim 3: Mediation Effects

Despite the nonsignificant result of the direct effect of regulatory focus on participation frequency in the final models for both domains, the indirect effects via confidence threshold did emerge as significant (as in Study 1). Also, in line with the results of Study 1, the relation between regulatory focus and identity was not characterized by a significant indirect effect in either domain. Because the final models of Study 2 did not measure a direct effect of regulatory focus on sense of belonging, I was not able to assess the indirect effects as I did in Study 1.

Exploratory Analyses

The exploratory analyses I conducted for Study 2 also revealed some interesting findings. First, although I expected the path models to operate similarly for men and women, results revealed some gender differences in the models. In the math domain, regulatory focus (but not confidence threshold) significantly predicted participation frequency for men, and confidence threshold (but not regulatory focus) predicted participation frequency for women. In social science, regulatory focus did not predict confidence threshold for men but was a significant predictor for women. These results are somewhat perplexing, but they seem to suggest that the confidence threshold construct perhaps is not as influential for men as it is for women. Future studies should follow up on these gender-related explorations with a larger sample size that includes more men. Since all of the pilot studies were done primarily with women, and both dissertation samples had more women than men, future investigations should be conducted regarding whether women are, for some reason, more attuned to the class participation confidence threshold variable.

The second exploratory analysis involved mediation analyses to investigate gender differences in confidence thresholds potentially predicting gender differences in participation. Results showed that in both the math and social science domains, the relation between gender and participation frequency was indeed mediated by confidence thresholds. This suggests that gender differences in confidence threshold might be responsible for gender differences in participation frequency. The results provide additional support for the explanatory role of confidence thresholds in predicting students' frequencies of participation in class, and particularly gender differences in the participation frequency.

Another exploratory analysis I conducted was examining how participants' levels of belonging, identity, and career interest changed across the semester. Belonging increased in both domains, which makes sense when considering how students might form greater connections to their courses and their peers in the courses after a longer period of time. With respect to identification with the domain, math identity increased, while social science identity remained stable. This is a fascinating result that could perhaps be explained by a potential effect of belonging on identity that is unique to the math domain. Despite the increase in both math sense of belonging and math identity, interest in a STEM career *decreased* across the semester. This result suggests that domain identity is likely not the only predictor of career interest, despite the robustness of this pathway across both studies.

The final exploratory analyses attempted to investigate the directionality of the relation between belonging and participation across both studies, based on empirical evidence that there is perhaps a reciprocal relation between the two constructs (e.g., Skinner & Belmont, 1993). As the addition of Phase 1 variables had nullified the association between participation and belonging in the final path models, I attempted to explore the relation between weeks within Phase 2 only. Unfortunately, these analyses did not provide any further information about the potential nature of this relation; when controlling for Week 1 variables, the relations were again eliminated. This result is not necessarily surprising, as the measures were aggregated within each week, and the weeks were also close together (typically only about two weeks apart). Furthermore, the weeks of Phase 2 occurred over a particular five-week period in the second half of the semester. It is possible that participation and belonging as measured only over this period might not reflect the entire extent of the relations between these variables, especially given that they change over the course of the semester.

In the *Results* section above, I noted that without controlling for sense of belonging in Week 1, percent participation frequency in Week 1 was a significant predictor of Week 2 sense of belonging in both math and social science. This finding is somewhat interesting, given that the opposite relation did not hold (i.e., Week 1 belonging was not a predictor of Week 2 participation frequency). In light of the other nonsignificant results, it could potentially lend some support to the nature of relation posited by the theoretical model (i.e., that participation predicts belonging, rather than the other direction). However, it is more likely that this result is merely a reflection of the significant correlation between Week 1 belonging and participation. In future studies, multilevel analyses examining the relation between sense of belonging and participation—without any aggregation across weeks or the semester—would help to tease apart how these variables influence each other throughout the semester.

Limitations

As with Study 1, the results of Study 2 must be considered in the context of a number of limitations. First, although the sample size of Study 2 was improved from the first study—with a 67% increase in sample size from Study 1—the number of variables that could be included in the path models was still somewhat limited. For instance, including career interest in Phase 1 as a covariate in the final models would have allowed for a further test of causality. However, to adhere to recommendations regarding sample size for path models (Kline, 2015), I was not able to include any additional constructs in the models.

Another limitation of Study 2 was the non-normality—and more specifically, the zero inflation—of the Phase 2 participation frequency variable. Unfortunately, I do not think there was a better way to measure this variable, as it reflected naturalistic count data and aligned with the reality that many students within a course never participate during the semester. In future work, I would correct for this issue analytically by conducting supplementary analyses consisting of alternate regression distributions (e.g., binomial or Poisson regression) in order to account for this limitation.

Finally, my use of aggregate measures for the Phase 2 (i.e., daily diaries) variables introduced additional limitations for consideration. Importantly, the primary research aims of the present dissertation were not affected by this decision, as the purpose of the diaries was mostly to gather a more naturalistic measure of students' participation behaviors. However, this operationalization did limit the potential for some further analyses and also presumably eliminated the variability necessary to detect some significant effects. More specifically, I was no longer able to explore how students' motivation and participation changed on a day-to-day basis throughout the semester. Future analyses should assess within-person relationships between variables by applying a multilevel model to account for variance across timepoints. For example, do students' confidence thresholds change on a daily basis, or are they relatively stable from student to student? If they do fluctuate, how do these changes influence their participation levels? By employing a multilevel model in which I have not collapsed across timepoints within Phase 2, I would be able to answer questions related to the within-person associations between variables in the model.

CHAPTER 6: GENERAL DISCUSSION AND CONCLUSIONS

Motivated by the existence of gender inequalities in the pursuit of STEM majors and careers (National Science Foundation, 2019), as well as the positive outcomes associated with active engagement in class (Christenson et al., 2001; O'Connor, 2013), the present dissertation has sought to investigate potential antecedents and consequences of gender-related disparities in classroom participation behaviors. The primary objectives were to explore possible gender differences in each of the constructs in my theoretical model (see Figure 1) and to investigate the hypothesized relations between these constructs. Specifically, the model posits that women participate less frequently than men in math classrooms because gender-based stereotype threat induces a prevention focus, prompting an increase in participation confidence thresholds and a reduction in participation frequency, which eventually impacts women's sense of belonging and identity in the math domain.

The dissertation consists of two studies, both of which were designed with the goal of probing associations in the theoretical model in samples of undergraduate students. Study 1 was a cross-sectional investigation of students' self-reported participation tendencies in their current math and social science classes. Study 2 used a daily diary methodology to investigate students' reports of their actual participation behavior throughout the semester.

In the following concluding chapter, I first address the major findings across both studies and interpret them in the context of both the objectives of the present work and prior research. In the second section, I discuss the primary limitations of this dissertation and suggest recommendations for how they should be remedied in future work. Finally, I present the chief contributions and implications of this research.

General Discussion of Findings

Preliminary summaries of the findings of each study are included in the *Discussion* sections of Chapters 4 and 5. The following general discussion addresses the key patterns of results that were consistent across both studies and therefore represent the principal conclusions of this dissertation.

Gender Differences in Constructs

The first aim of the present dissertation was focused on exploring gender differences in classroom participation frequency, as well as psychological mechanisms potentially related to students' participation. I hypothesized that women would report lower levels of participation frequency than men in math classes (but not necessarily in social science classes), and that this difference would be associated with women's higher stereotype threat susceptibility, prevention focus, and confidence thresholds, as well as with their lower levels of belonging, identity, and career interest in the math domain. The results partially supported this hypothesis, as they showed gender differences in the expected direction in math; however, a similar pattern of gender differences also emerged in the domain of social science for stereotype threat and confidence threshold. With respect to the hypothesized outcomes of classroom participation, the pattern of gender findings was domain-specific. In line with my hypotheses, women, in comparison to men, exhibited lower levels of belonging, identity, and career interest in math, but equal or greater levels of these variables in social science. Below, I discuss these findings in the context of existing literature on gender differences. Table 14 presents key patterns of gender and domain differences across both studies.

Table 14

Key Patterns of Gender and Domain Differences (Across Study 1 and Study 2/Phase 1)

	Overall Gender Differences	Gender Differences by Domain		Overall Domain Differences	Domain Differences by Gender	
		Math	Social Science		Female	Male
Stereotype threat (ST-1)	Female > Male	No interaction gender x domain		–	No interaction gender x domain	
Stereotype threat (ST-2)	Female > Male	Female > Male	Female > Male	Math > SS	Math > SS	SS > Math [c]
Prevention focus	–	–	Male > Female	Math > SS [b]	Math > SS	–
CPCT (both measures)	Female > Male	No interaction gender x domain		Math > SS	No interaction gender x domain	
Partic. frequency (raw)	Male > Female [c]	No interaction gender x domain		SS > Math	No interaction gender x domain	
Partic. frequency (of opps.) [a]	Male > Female	No interaction gender x domain		SS > Math	No interaction gender x domain	
Freq. of changing mind	Female > Male	Female > Male [d]	Female > Male [d]	–	–	SS > Math [d]
Participation opportunities	Female > Male [c]	No interaction gender x domain		SS > Math	No interaction gender x domain	
Sense of belonging	–	Male > Female	–	SS > Math	SS > Math	–
Domain identity	–	Male > Female [b]	Female > Male	–	SS > Math [e]	Math > SS [e]
Career interest	–	Male > Female	Female > Male	–	SS > Math	Math > SS
Domain confidence	–	Male > Female [b]	Female > Male [e]	SS > Math [c]	SS > Math	–
Domain value	–	Male > Female [b]	Female > Male	–	SS > Math [e]	Math > SS
Participation confidence	Male > Female [c]	No interaction gender x domain		–	No interaction gender x domain	
Confidence diff. score	Male > Female	No interaction gender x domain		SS > Math	No interaction gender x domain	

Note. Dashed line indicates no significant difference. [a] Measured in Study 2 (S2)/Phase 2 and explored in a supplemental analysis in Study 1 (S1). [b] Significant in S1, but not in S2. [c] Significant in S2, but not in S1. [d] = Only investigated in S1, as there was no significant interaction to probe in S2. [e] = Significant in S1, but only marginally significant in S2.

Gender Differences in Predictors of Participation and Participation Frequency

The findings with respect to stereotype threat susceptibility were consistent with research positing that women display greater susceptibility to stereotype threat in math than men. For the second stereotype threat measure (ST-2), the size of the gender difference in math was greater than the size of the difference in social science, which is consistent with prior research suggesting that gender differences in stereotype threat are more prominent in the math domain (e.g., Inzlicht & Ben-Zeev, 2000). However, the present findings showing that female students reported higher levels of stereotype threat in *both* math and social science are particularly noteworthy, given that I made an effort to select social science courses that presumably are not affected by gender stereotypes. To further investigate the nature of this finding, I added math major as a factor to the analysis. I expected that female math majors would exhibit higher levels of stereotype threat than both non-math major female students and male math majors, based on prior research suggesting the importance of identification with the domain in influencing stereotype threat susceptibility (Steele et al., 2002). However, this analysis revealed no significant effect of major or interaction involving major with respect to stereotype threat.

As highlighted in the *Literature Review*, the field of stereotype threat research is currently characterized by extensive deliberation about the nature and measurement of this phenomenon, particularly with respect to women in math contexts (Inzlicht, 2016). The present findings contribute another piece to this puzzle. In particular, while major is commonly used as a proxy for identification with a domain, the null result with respect to major is consistent with some other research that has attempted to explore the unique influence of stereotype threat on math-identified women outside of a laboratory setting (e.g., Cullen et al., 2006). As Cullen et al. suggested, when investigating effects of stereotype threat in naturalistic contexts, there are a multitude of additional factors that may affect the strength of potential moderators such as major. Another possible explanation for this result is that these effects are relatively small, and the present studies did not have sufficient power to detect them. Future research should make an effort to recruit a larger sample size overall, including a comparably sized group of math majors, in order to more carefully explore these effects.

Domain-specific gender differences in regulatory focus appeared in both studies, such that women displayed higher levels of prevention focus in math than in social science, while there was no domain difference for men. Further, women exhibited lower levels of prevention focus in social science than men, which aligns with prior research suggesting that women may display a unique predisposition to a prevention focus in math contexts (Coffman & Klinowski, 2020; Tannenbaum, 2012). To my knowledge, the present dissertation is the first set of studies to compare regulatory focus across both gender and domain in the same set of participants. Thus, these results provide evidence that women are particularly prevention-focused in math settings.

Across studies, confidence thresholds and participation frequency generally displayed gender differences in both domains. Across both academic domains, women reported higher confidence thresholds for participation, lower participation frequencies (in Study 2), and higher frequencies of changing their minds about participating. Although much of the research on gender differences in participation frequency has focused on STEM domains (e.g., Carter et al., 2018; Streitmatter, 1997), some other studies have identified lower participation for female students across multiple domains (Grover Aukrust, 2008). Thus, although I hypothesized that these differences would appear primarily in the math domain, the aggregate results of these studies suggest that women and girls perhaps participate less frequently than men and boys in general, regardless of domain.

It is noteworthy that, while I hypothesized that gender differences would exist primarily in the math domain and not necessarily in social science, many of the disparities in predictor variables were reflected in social science as well. Women reported higher stereotype threat susceptibility, higher confidence thresholds, lower participation frequency (in Study 2), and a greater frequency of changing their minds about participating than men in both domains, not just in math. While regulatory focus was characterized by a gender-by-domain interaction—such that women exhibited significantly lower prevention focus than men in social science—stereotype threat and confidence thresholds remained high for women across both domains. These results perhaps reflect effects of feminine gender norms, which were not measured for the present dissertation. In other words, perhaps the math domain is not the only setting in which women display a reticence to contribute; instead, this hesitation might permeate even more aspects of their lives via feminine norms such as modesty and aversion to risk taking (e.g., Mahalik et al., 2003, 2005). Pilot Study 1 included measures of these gender norms, but the analyses yielded no significant gender differences and no significant associations between these norms and frequency of participation. However, it is possible that the middle school students who were participants in the study had not yet internalized these norms. These gender norms were again revisited in Pilot Study 2, which included a sample of college students. While the gender imbalance in the sample precluded me from testing gender differences, the norm of risk-taking—which is traditionally associated with males (Mahalik et al., 2003)—was significantly negatively associated with both measures of confidence threshold in the math domain and one of the threshold measures in social science.

Along with gender norms, another possible explanation for overarching gender differences in these constructs is gender disparities in the personality trait of conscientiousness.

Small gender differences in conscientiousness, favoring women, have been reliably identified across a number of studies (Costa et al., 2001; Feingold, 1994; Schmitt et al., 2008). It is possible that conscientiousness—particularly with its relation to self-control—is associated with caution in how and when one chooses to participate (e.g., not wanting to blurt out the first thing that comes to mind and instead ensure that it is correct before contributing). Pilot Study 2 examined each of the personality traits comprising the Big Five Inventory using an abridged measure by Rammstedt and John (2007). Gender differences could not be explored due to the small number of male participants, and no relations emerged between conscientiousness and participation frequency in this sample. Although there is perhaps limited support for the impact of these variables, future iterations of this work could investigate the potential impacts of gender norms and personality differences in the theoretical model in a more representative sample.

Gender Differences in Outcomes of Participation

In contrast to the predictors of participation and frequency of participation itself, the outcome variables in the model generally exhibited math-specific gender differences. Specifically, across both studies, women generally reported lower levels of belonging, identity, and career interest than men in math, and equal or higher levels of these variables compared to men in social science. These findings are consistent with literature that has explored these constructs among women in STEM settings (e.g., Lock et al., 2013; Murphy et al., 2007; Thoman et al., 2014). However, to my knowledge, there has been limited research on these constructs for men in non-STEM domains, and I am not aware of any other research that has compared these constructs for both genders across academic domains. Thus, the domain-specific nature of these findings provides a novel contribution to the literature.

The final set of variables in which gender differences were explored was the proposed covariates in the model (i.e., confidence and value). Across studies, women generally displayed lower confidence in and perceived value of math, and higher levels of both variables in social science. These results are consistent with research that has found similar gender differences in expectancies and values (e.g., Wang & Degol, 2013). With respect to participation-related confidence specifically, women demonstrated lower levels in both domains in Study 2, which makes sense in light of the domain-independence of their higher confidence thresholds and lower participation frequency exhibited in the same study.

Relations Among Predictors of Classroom Participation

Stereotype Threat Susceptibility

From the original conceptualization of the research questions of this dissertation, stereotype threat was posited as the first factor in the theoretical model, representing the proposed source of gender differences in the model. In other words, my original hypothesis was that gender differences in stereotype threat susceptibility would translate to further disparities in other related constructs, such as regulatory focus and confidence thresholds. As a reminder, while ST-2 was associated with a number of other constructs in both studies, ST-1 was included in the path models because it represented a more appropriate measure of stereotype threat susceptibility than the second measure. Ultimately, however, stereotype threat was eliminated from the final models due to a lack of influence on other variables in the model.

In examining potential reasons for the lack of the relation between stereotype threat and its posited correlates, I considered the possibility that the measures of stereotype threat susceptibility used in the present study were not appropriate. Discussions among researchers have recently been focused on how stereotype threat has historically been operationalized; the first stereotype threat measure, ST-1, was created specifically to address these concerns. Scholars have called for a return to how stereotype threat was originally conceptualized by Steele (1997) as a "threat in the air," or the felt experience of participants who are concerned about being evaluated (e.g., Lewis & Sekaquaptewa, 2016). For this reason, with ST-1, I attempted to capture students' affective experiences in math and social science classrooms of various gender compositions. Despite these intentions, it is possible that the measure made gender too salient—in other words, it may have been too explicit in cueing students to consider their gender, introducing systematic bias into the results (e.g., if participants were aware of how they "should" be answering the prompt and were compelled to answer in a way that was not aligned with their actual experiences). It is worth noting that I was cognizant of this potential concern as I was creating the ST-1 measure, but it was challenging—particularly in a survey format—to ensure that participants would reflect on their experiences of stereotype threat when completing the subsequent measures if gender was not explicitly mentioned.

A final possible explanation for why ST-1 did not exhibit predictive influence in the path models is that the samples did not include enough highly-identified female math majors, who are theoretically more likely to display high levels of stereotype threat susceptibility (Steele et al., 2002). It is worth mentioning that in addition to exploring stereotype threat among all participants, I also examined correlations among female participants only, and the relevant associations remained nonsignificant (other than those with confidence threshold that were already present in the math domain in Study 2). In both studies, exploration of the influence of major on stereotype threat yielded null results, which could be due to the relatively low number of female math majors in the sample (15.5% in Study 1 and 10.4% in Study 2). It is possible that in samples with higher proportions of math majors (and particularly female math majors), significant associations between stereotype threat and other variables would be detected.

It is important to note that although the scenario-based measure of stereotype threat susceptibility, ST-1, was included in the path models, the Likert-type scale, ST-2, actually was characterized by some significant correlations with relevant variables. As a reminder, ST-2 was not used in the path models because of its focus on awareness of gender-related stereotypes, rather than the feeling “in the air” representing the currently accepted conceptualization of stereotype threat. ST-2 was *not* correlated with regulatory focus in either study or domain; however, in both studies, it was positively correlated with the confidence threshold measures in both domains and negatively associated with participation frequency in the math domain.

These associations elicit the question: if ST-2 does not work through regulatory focus to increase confidence thresholds and decrease participation frequency, via what mechanisms does it operate? Based on the inclusion of words such as “pressure,” “afraid,” and “fear” in the ST-2 items, I would have expected it to correlate with a predominant prevention focus. However, perhaps ST-2, with its focus on awareness of gender-based stereotypes, was a measure of worry associated with the social costs of doing poorly in math, whereas the domain-specific measure of regulatory focus captured a broader set of concerns, including the personal costs of not meeting one’s goals in the math domain. Although stereotype threat was removed from the final models for this dissertation, both of the measures employed in these studies exhibit some interesting characteristics that warrant further investigation in future studies.

Regulatory Focus

With the removal of stereotype susceptibility from the path models, regulatory focus (operationalized as tendency toward a prevention focus) became the most distal factor in the final models. Taken together, the results of both studies suggest that students' domain-specific regulatory focus is associated with their classroom behavior—and more specifically, with their frequency of participation in class via their internal confidence thresholds. Participants who were inclined to approach particular educational contexts (e.g., math classes) in a state of vigilance were presumably more likely to view participation as something that could result in loss and were therefore significantly more risk averse with respect to this form of engagement.

This finding is particularly compelling because, although regulatory focus theory has been explored extensively in social psychology research, it has been relatively underutilized in education research (Molden & Rosenzweig, 2016) and not studied at all with respect to classroom participation. Combined with the domain- and gender-related differences in prevention focus, such that women are more likely to exhibit a prevention orientation in math contexts than in social science contexts, the present dissertation perhaps offers an explanation for how this motivational state can explain gender differences in frequency of participation.

In addition to its associations with confidence thresholds and participation frequency, regulatory focus was also a direct predictor of domain-specific identity in the majority of the final path models (other than in the math domain in Study 2, when Phase 1 identity was included as a covariate). As described in the *Results* sections for both studies, this novel pathway was added based on a modification index identified during the model revision process. Although it was not originally hypothesized as a primary path in the theoretical model, its theoretical justification made it a reasonable addition to the models. Because regulatory focus is measured as a difference score of prevention minus promotion, it could be considered as a marker of relatively *low promotion focus* in addition to *high prevention focus*. It makes sense that students who do not feel eager, excited, or hopeful in their math courses would not view themselves as highly identified with math, especially in western cultures where identity is closely associated with self-actualization and individualistic aspiration (Jetten et al., 2002). It is also possible that this association could operate in the opposite direction. In other words, if domain identity was moved to the beginning of the model, perhaps results would show that a consequence of low identification with the subject is less of a promotion focus (i.e., fewer hopes and aspirations in the domain).

Confidence Thresholds

Perhaps the most robust set of findings of the present dissertation involves the novel measure of class participation confidence threshold. Across both studies and domains, confidence thresholds (particularly when measured with Likert items—i.e., CPCT-2) were positively predicted by prevention orientation and negatively predicted frequency of classroom participation. Furthermore, in both studies and for both math and social science, mediation analyses identified an indirect effect of regulatory focus on frequency of participation via confidence threshold. These findings suggest that being predominately prevention focused may cause students to participate less frequently because it leads them to set a higher internal threshold for tolerance for making a mistake.

In the educational psychology literature, Expectancy-Value Theory (Wigfield & Eccles, 2000) has historically been used to explain variation in educational behaviors, such as enrolling in particular courses (Durik et al., 2006), selecting a college major (Musu-Gillette et al., 2015), and procrastinating on academic tasks (Wu & Fan, 2017). In prior research, both expectations of success and task values have been found to influence students' choices, persistence, and engagement. In the present dissertation studies, even when controlling for these motivational variables traditionally thought to predict classroom behaviors—namely, confidence and value—

confidence threshold still emerged as a significant predictor of participation frequency (i.e., a form of engagement). This compelling result suggests that this novel variable may have a unique effect on participation that does not operate through the motivational variables that have traditionally been the focus of educational psychology. Relatedly, in the correlation results for both studies, participation frequency was consistently associated with confidence thresholds more strongly than it was with confidence and value. Thus, not only was the confidence threshold a novel motivational variable accounting for unique variance in students' participation behaviors; it actually accounted for much *more* overall variance than any other factors assessed in the present work. These results as a whole suggest that the magnitude of students' predominant prevention focus may influence this judgment threshold, which may ultimately impact participation, independently of one's overall confidence and value in the domain.

Relations Among Outcomes of Classroom Participation

Whereas classroom participation served as an outcome of stereotype threat susceptibility, prevention focus, and confidence thresholds in the theoretical model, it was hypothesized that participation itself might predict broad psychological outcomes such as sense of belonging, domain identification, and career interest. In this section, I discuss the constructs that were examined as potential outcomes of classroom participation.

Sense of Belonging

A significant association between participation and sense of belonging emerged in many of the final path models. However, in Study 2, this association was eliminated in both domains when Phase 1 belonging was added as a covariate. Although prior work has identified an impact of active participation on student belonging (Christenson et al., 2001; Goldstein & Benassi, 1994), the present work did not confirm whether frequency of participation positively impacted students' sense of belonging.

As first suggested in the *Literature Review*, it is possible that the relation between belonging and participation frequently operates in the opposite direction—with sense of belonging predicting participation frequency—or that the two have reciprocal effects on each other, in a positive feedback loop with accumulating effects (Skinner & Belmont, 1993; Yeager & Walton, 2011). Importantly, a recursive process like this would only thrive in environments in which students are encouraged to participate even if their responses are incorrect, and the support they receive after providing an incorrect answer is enough to further boost their sense of belonging. The results of the present dissertation did not address this potential reciprocal relation, despite attempts to control for belonging earlier in the semester, and even within the daily diaries period specifically. In future work, I plan to conduct multilevel analyses to investigate this relation across individual class days—i.e., on a more granular level, as opposed to with the coarse measures used in the present analyses. This further investigation might illuminate the nature of the relation(s) between belonging and participation.

Domain Identity

While the association between frequency of participation and sense of belonging was relatively robust across domain and study—despite the uncertainty about the causal nature of the relation—the influence of participation on identity appeared to be more domain-specific. Participation consistently emerged as a predictor of identity in social science, but not in math.

Potential reasons for this lack of a consistent association between math identity and frequency of participation can be more intuitively explored when considering the relation in the opposite direction from how it is posed in the theoretical model (i.e., the influence of identity on participation). More specifically, the lack of a consistent association could reflect opposing ways in which identity potentially impacts participation. For some students, a strong sense of math identity could be associated with increased participation because they consider engaging in a particular domain to be an activity that aligns with their sense of self. In other cases, students' increased math identity may lead them to feel anxious due to heightened stakes they personally associate with the domain, as stereotype threat research has shown that stereotype threat effects are particularly salient for highly identified individuals (Steele et al., 2002). In this case, taking math particularly seriously may raise the stakes of participation, leading the student to raise their hand less often. That is, there may be suppression effect at work, whereby identity simultaneously impacts participation in opposite directions, such that the overall correlation between the variables is very small.

Career Interest

In addition to exploring the effects of participation on belonging and identity, the final models also examined how sense of belonging and identity within each domain impacted students' interest in pursuing a career in a related field. As expected, across both studies and domains, the cross-sectional association between identity and career interest remained stable and highly significant. This result is consistent with prior evidence suggesting that domain identity may impact career choice (Cribbs et al., 2020; Lock et al., 2013).

With respect to belonging, the results were less reliable—in math, the effect was only significant in Study 2 (when controlling for identity at the same timepoint), whereas in social science, it was only significant in Study 1. Notably, however, the size of the standardized effect was the same in both studies in the math domain, indicating that the nonsignificant result in the first study may have been due to a lack of power. Thus, the results do seem to suggest that participants' sense of belonging within a domain may impact their interest in related careers, which aligns with findings from other work on the impact of belonging in career choice (e.g., Thoman et al., 2014). A critical caveat is that I did not examine the effects of identity and sense of belonging on career interest while controlling for career interest at the beginning of the semester, and therefore I cannot make any strong claims about the direction of these associations.

It is worth noting that the final models did not explore the possibility of a direct effect of participation on career interest (i.e., one that was not mediated by belonging or identity). It is possible that becoming more engaged in a class could generate interest in the course material, which could bolster interest in the field overall. Future analyses of Study 2 could explore the potential existence of this direct effect.

General Commentary on the Outcome Variables

Overall, the findings with respect to the second half of the theoretical model, containing the outcome variables, were less consistent than the results related to the first portion of the model. One reason for this difference may be that so many factors contribute to these outcomes that a single factor (e.g., participation) might not account for a substantial proportion of their variance. Relatedly, if participation does have a small but reliable impact on belonging, identity, or career interest, this impact may be more readily detected by assessing participation throughout the entire semester, rather than across just two weeks (as in Study 2 of the present dissertation), especially when averaging participation to create a composite measure. Future analyses that explore relations between participation and belonging and identity or belonging on a day-to-day basis would also be informative.

Another, more substantial, consideration is that the present dissertation was based on the underlying assumption that, overall, participation is a positive thing. In other words, it was assumed that the students were operating within a supportive environment, such that participation would elicit stronger engagement, belonging, and identification. However, the present studies did not assess the extent to which the classes from which the participants were sampled actually supported their participation. In environments where students did not perceive their comments as being validated or positively reinforced by instructors and peers, it is possible that participation may have been a neutral or aversive experience. In fact, students who were in an unsupportive environment may have experienced reactions from their instructors and peers that undermined their sense of belonging and identity. Thus, it is possible that while some students experienced increases in positive outcomes from participating, these effects were counteracted by the negative experiences of other students who did not consider the responses of their instructor to be positively reinforcing.

It is also important to note that participation in the present dissertation was operationalized as frequency of raising one's hand, whether or not the student was selected by the instructor to actually answer the question. Participation was measured in this way because it aligned well with the first portion of the model: regardless of being selected to respond or not, raising one's hand is a measurable indicator of surpassing one's confidence threshold enough to prompt behavior. However, without measuring if students were actually selected to participate, it was not possible to determine how many of instances of *intent* to participate actually manifested in sharing a response in front of the class. Thus, it is possible that many of the instances of raising one's hand did not equate to actual participation. Of course, if participants were not actually selected to participate, an improvement in belonging and identity would not reasonably be expected. In fact, the opposite effect is possible: a student who raises their hand consistently but is never selected by the instructor may experience a *decline* in sense of belonging. In sum, it is possible that my initial assumption regarding the average educational environment at the participating universities—i.e., that faculty generally try to support the participation of all students who are willing to participate—was overly optimistic.

A final thought regarding the outcomes of participation is that there may be individual differences in the experience of actively participating in class, even within a given classroom environment. In cases where participation is taken into account for students' grades, for example, being evaluated could be viewed as a burden by some students and an incentive for others. Some students might find that participating—and perhaps even the process of planning how to participate—helps to keep them engaged in class. In contrast, other students might be so consumed by thoughts of how to contribute that the pressure to participate (whether for grades or otherwise) undermines their classroom experience. If it is the case that participation is a positive experience for some students and an aversive one for others, the overall effects of participation on outcomes like belonging, identity, and career interest could be reduced by the conflicting influences.

In sum, with respect to the inconsistent results regarding relations among the outcomes of participation, it is important to note that participation does generally seem to be associated with positive outcomes in the research literature. It is certainly possible that contrary to my expectations, participating more frequently—especially in the form studied for this dissertation—does not necessarily lead to substantial increases in belonging and identity. However, even if it does not directly impact students' identification with or belonging in an academic domain in all of its forms, prior research suggests that it is influential in improving students' understanding of the material, increasing grades (even while controlling for prior achievement), fostering critical thinking, and nurturing students' sense of agency in their learning (Böheim et al., 2020b; Dallimore et al., 2004; Dixon et al., 2009; Webb et al., 2014).

Limitations and Future Directions

The analyses conducted for this dissertation represent an initial examination of the rich set of data collected across the two studies. Consequently, there are some limitations to consider in interpreting the findings. In addition to the limitations specifically pertaining to the individual studies (as detailed in the *Discussion* sections of Chapters 4 and 5), the dissertation as a whole is characterized by some further limitations.

Methodological Limitations

Sample Considerations

As discussed previously, samples of both studies were perhaps not an adequate size to appropriately conduct the path analyses that were performed (Kline, 2015). Both samples also displayed an unequal gender breakdown, although the second study approached a more equivalent balance than the first study. During participant recruitment for both studies, I attempted to maintain some balance between group sizes while also retaining as many participants as possible. As Levene's tests were generally nonsignificant for both studies, this inequality did not appear to affect the robustness of the results. However, it is possible that the comparatively small number of male participants in each study led to a lack of power to detect some significant effects. Based on the exploratory analysis that suggested that there may be differences between the path models for men and women, it is also possible that a comparatively low number of male participants contributed to some misleading path analysis results (when collapsing across gender).

Further, math majors constituted a greater percentage of male participants than female participants in both studies, despite efforts to recruit math majors of both genders. This could be a problem because it potentially introduced identity and stereotype threat susceptibility differences that were not exclusively related to gender. Thus, future research should attempt to replicate the present findings in a larger sample with equal numbers of men and women and more evenly distributed percentages of math majors of each gender.

Another sample limitation relates to which participants were excluded due to missing data. While participants were not removed from the study for failing to complete the Phase 3 survey, they were required to complete a minimum threshold of daily diaries surveys to be considered as final participants. Thus, even though I used missing data analytical techniques to retain as many participants as possible, there was potential for selection bias, as the motivational characteristics of the excluded participants may have been somewhat different from those of the participants who completed enough surveys to be considered in the final sample.

Race as a Missing Variable

Another limitation is that both studies' samples were overwhelmingly racially homogenous, with approximately 90% of the participants in each study identifying as White or Asian. Thus, the samples of the present dissertation are not representative of college students nationally. Furthermore, race was not considered as a relevant variable in the present dissertation, despite its potential to influence participation, belonging, and identity as much as gender. With respect to stereotype threat in particular, the findings regarding women in STEM are the most tenuous in light of the field's recent reckoning (e.g., Inzlicht, 2016), while many scholars continue to endorse the existence of race-related stereotype threat effects (e.g., Lewis & Sekaquaptewa, 2016).

Another way that race should be considered in this line of work is in terms of its intersectionality with gender. One Asian female participant of Study 2 sent me an email after completing the final survey that highlighted her conflicted feelings on sense of belonging in her courses with respect to her gender and race. She wrote: “I am aware of the stereotype that women are bad at math/science, and that does hover over me, especially in my 5:1 male to female [math] class. However, I often feel like I fit right in during [name of math class] because that is my only majority-Asian class this semester. So often, my mind is more on the stereotype/expectation that I will excel as an Asian American student rather than the stereotype that I will fail as a female student. That is part of why I feel more like I belong there than in my [social science] class.” The qualitative experience of this participant is reflected in research on the stereotype threat susceptibility of students whose identities align with multiple, potentially conflicting stereotypes (e.g., Shih et al., 1999). Thus, future work that takes participants’ race into account might illuminate the intersectional influence of race in the theoretical model of this dissertation.

Context of Data Collection: COVID-19 Pandemic

The historical context of the present dissertation also presented a number of limitations, as data collection was conducted during the COVID-19 pandemic for both studies. Since the onset of the pandemic in 2020, college students have been particularly affected by the ensuing disruption. More specifically, they have reported increased academic stress, depression, and difficulty coping (Clabaugh et al., 2021), as well as attention struggles and increased challenges with both physical and mental health (Elharake et al., 2022).

Beyond its impact on participants, the pandemic also directly affected the structure of the universities at which data collection was being conducted. During Study 1 in particular, the majority of university courses were conducted online. This is a critical consideration, especially given that the present dissertation was focused on classroom participation. Naturally, active engagement in a Zoom classroom is a qualitatively different experience than participating in in-person settings (Joia & Lorenzo, 2021), and this likely had an effect on the frequency and nature of students' participation in class. In Study 2, according to participants' responses on the daily surveys, classes were mostly held in person; out of 864 total math classes and 804 total social science classes attended by participants throughout the semester, only 29 were held online synchronously. However, as described previously, courses in Study 2 were characterized by frequent last-minute class cancellations and both students and instructors facing mandatory quarantining. Thus, the interpretation of the results of both studies should be characterized by careful consideration of the impact of the pandemic.

Given the ways in which both participants and universities were impacted by COVID-19 during data collection, it is possible that the results of the present dissertation are not generalizable beyond pandemic times. However, it is worth noting that many of the findings align with those of the pilot studies, which were mostly conducted before the onset of the pandemic.

Instructor- and Course-Level Differences

Another set of methodological limitations of the present dissertation is related to the specific courses participants chose to focus on throughout the semester. Although I did review each participant's list of current courses to determine which ones qualified as math and/or social science courses, I was not able to ensure that students ultimately selected the most appropriate course from the list of options. For example, I did not have access to information about class size, so I was not able to account for course size in deciding which courses were acceptable.

Thus, some of the final courses may have been held in lecture halls, which could have contributed to nearly half of the participants reporting having raised their hand zero times in the math domain in Study 2. In future studies, I would ensure that only courses with enrollment sizes under a certain threshold (e.g., maximum 40 students) were eligible for selection.

Relatedly, courses varied in the types of participation offered. Some students may have chosen to focus on courses that offered a type of participation opportunities that was not assessed in the present dissertation (e.g., clicker responses in large lectures or small group discussion). These forms of participation represent a low possibility of "loss" associated with confidence and acceptance, since they are more anonymous than participation in front of the whole class. As an attempt to mitigate this issue, I encouraged participants to consider which courses offered the most opportunities for participation specifically via hand-raising while selecting their relevant courses. However, given that 15 participants (5.6%) in Study 2 reported not having any opportunities to participate in at least one of their courses, it is clear that some students may not have had a satisfactory course option available to them. For this reason, I included opportunities for participation in the final models (as a covariate in Study 1 and inherently incorporated in the percent frequency of participation variable in Study 2), as this was an important consideration in evaluating students' participation frequency.

Another potential course-level difference is the extent to which participation counted toward students' final grades. In cases where participation constituted a significant portion of students' grades, a prevention focus could potentially lead students to actually participate *more*, due to increased vigilance about avoiding loss of points toward their final course grades. However, given that instructors seemed to offer many opportunities for participating in each class period (i.e., an average of 16 opportunities per math class and 26 opportunities per social science session in Study 2), I suspect that most participants viewed participating in class in a promotion-focused manner, rather than viewing each missed opportunity as a loss of points. In other words, I believe that a prevention focus may have led these students to participate some minimal amount in order to secure a sufficient number of points, rather than to participate a lot (i.e., in a gain-oriented manner). Although there are likely a number of conflicting sources of motivation for students (e.g., avoiding public shame and loss of confidence, while also avoiding loss of points for participation), the focus of the present dissertation is gender differences and which motivators might uniquely affect women. Presumably all students in a course would be considering their grades in determining whether to participate, but women might be more subject to concerns about what they would lose if they answered a question incorrectly.

Analytical Limitations

Approach to Path Analysis

As previously outlined, my general approach to conducting path analysis in both studies was to prioritize ease of comparison of models across domains and studies, while balancing parsimony, theory, and fit within the constraints of the data. In other words, my goal was essentially to test the theoretical model, not to find a new model explaining the relations between constructs. Even though the final models in Study 2 achieved reasonably adequate fit, that does not necessarily mean they are “correct,” as there are likely better-fitting models that were not explored (Streiner, 2005). Future work should test models that include alternate pathways that were not explicitly investigated as part of the present dissertation. For example, in future models, I would be interested in investigating the direct effect of regulatory focus on career interest. This relation makes theoretical sense because of the affective component of regulatory focus, which posits that negative emotions such as agitation are frequently associated with a prevention focus, particularly when students are having difficulty achieving their goals in a prevention focus (Higgins, 1997). Thus, students might be more likely to select a career in a domain that they associate with more positive emotions.

Another ramification of this parsimonious approach to testing path models was that I could not include all the variables I would have liked incorporate or test all potential pathways that are theoretically reasonable and worth exploring in the future. For instance, as mentioned in the discussion for Study 2, there were not adequate degrees of freedom to include career interest in Phase 1 as a covariate in the final models, despite the value it would have added as a further test of causality. It is also possible that the relation between belonging and participation could itself be mediated by confidence thresholds. Future analyses will investigate other potential configurations of the theoretical model.

Use of Difference Scores

A final potential analytical limitation of this dissertation is the use of difference scores—specifically in measuring both stereotype threat, regulatory focus, and the aptly named confidence difference score (i.e., a measure of one's typical confidence level over and above their confidence threshold). Difference scores are often criticized, particularly due to their low reliability (Cafri et al., 2010). On the other hand, some researchers have defended difference scores, particularly for use in inferential statistics (Thomas & Zumbo, 2012; Trafimow, 2015). Regardless, their limitations have been investigated in depth, and recommendations based on statistical results from difference-score constructs should be made with caution.

Conclusions

In sum, the present dissertation identified gender differences in all constructs in the theoretical model in Figure 1 and found support for many of the theoretical model's relations. In both domains, women generally displayed higher levels of the predictor variables than men, as well as lower levels of participation, especially when taking number of opportunities into account. With respect to the outcome variables of belonging, identity, and career interest, findings depended on the domain: women exhibited lower levels than men in math (although, only in one study for identity) and equal or higher levels in social science. Tests of the hypothesized model showed that a prevention focus negatively predicted frequency of participation, as mediated by class participation confidence threshold. Tests of the second half of the model (i.e., the outcomes of participation) were less robust, with some domain-dependent findings, although the associations between participation and belonging, and identity and career interest, were generally established. Together, these results suggest that gender differences in regulatory focus in the math domain may lead women to set higher confidence thresholds and participate less frequently. The findings did not confirm whether this less frequent participation was causally related to women's lower belonging, identity, and STEM career interest.

Contributions and Implications of the Present Research

This dissertation offers a number of potential contributions to the extant literature on gender differences in academic contexts. First, the two reported studies were unique in their focus on *classroom participation* as a construct of interest in the exploration of potential reasons for gender disparities in pursuit of STEM occupations. To my knowledge, no prior studies have explored how gender differences in participation might contribute to other downstream gender disparities. This is a missed opportunity, as participation is a malleable, observed behavior that could presumably be responsive to intervention. Furthermore, the present dissertation extends existing literature in attempting to explain underlying reasons for gender differences in participation behaviors, rather than simply focusing on establishing their existence. While a number of predictors and outcomes of classroom participation have been addressed in an emerging line of research (e.g., Böheim et al., 2020a, 2020b; Webb et al., 2014), the present dissertation examined a set of predictors and outcomes not frequently investigated in relation to participation. Thus, the present studies contribute to this burgeoning research on antecedents and consequences of classroom participation, while simultaneously serving as a bridge with literature on gender differences.

Another contribution of this dissertation is its introduction of a novel psychological construct—*class participation confidence thresholds*—in an attempt to explain why general confidence in a given domain cannot solely account for gaps in participation behaviors. The robust effect of confidence thresholds on participation frequency is compelling, especially given that it still holds even when controlling for other motivational measures from Expectancy-Value Theory (Wigfield & Eccles, 2000). Furthermore, confidence thresholds introduce a mechanism by which certain motivational variables that have been underexplored in educational psychology (e.g., regulatory focus) influence educationally-relevant behaviors.

Moreover, this novel construct could also potentially apply to the threat-related decisions people make in contexts other than the classroom. For instance, a frequently referenced internal report from Hewlett-Packard discovered that while men applied for promotions when they satisfied about 60% of the job requirements, women generally did not apply unless they met 100% of the qualifications (Clark, 2014). With respect to women's persistence in STEM specifically, research shows that women enroll in fewer upper-level math classes than men (National Science Foundation, 2019). Much of the work exploring women's selection of major and enrollment in math-related courses from the perspective of Expectancy-Value Theory has focused on women's confidence in their math abilities and enjoyment of the subject. However, it is possible that there is a mechanism similar to confidence thresholds dictating women's perceptions of how much confidence and value is required in order to persist in STEM, and that these thresholds are influenced by a range of more distal variables that have not been carefully explored. Thus, future research should examine how confidence thresholds apply in other motivational contexts that may further explain women's reticence to pursue careers in STEM fields.

A third contribution of the present work relates to the inclusion of *regulatory focus* in the theoretical model. Although some of the relations in the model have been established in prior literature, others, such as the association between regulatory focus and participation, were investigated for the first time in this research. Regulatory focus is a concept that is popular in the social psychology literature but has not been widely adopted in educational psychology (Molden & Rosenzweig, 2016). These studies represent a preliminary examination of the impact of regulatory focus on classroom participation, which is made even more compelling by the consistent significance of the relation between these variables (as mediated by confidence thresholds) across both studies and both domains.

Taken together, the key takeaway of these findings is that a prevention focus in a classroom context can lead to less frequent participation, via increased class participation confidence thresholds. Although this relation seems to hold across domains, women are significantly more likely to exhibit a predominant prevention focus in a math setting compared to a social science setting, report higher confidence thresholds than men, and participate less frequently than men (based on results from Study 2). Based on this relation and supplementary exploratory analyses, gender differences in confidence thresholds could reasonably be responsible for differences in participation frequency. This is concerning because participation is associated with a number of important outcomes, both based on prior work (Böheim et al., 2020b; Christenson et al., 2001; Dallimore et al., 2004) and the present set of studies (although not necessarily in a causal manner).

Initially, it may seem that the primary implication of this work is for educators to help prevention-focused students become more promotion-focused in order to increase their frequency of participation. Alternatively, given the robust gender differences in confidence threshold—but not regulatory focus—observed in the math domain, it may also make sense to direct intervention efforts toward confidence thresholds by encouraging women to contribute even if they are not completely certain that their response is correct. In some cases, encouraging modifications to women's regulatory focus and confidence thresholds may be suitable solutions. However, as previously discussed, simply increasing frequency of participation may not necessarily be helpful in certain classroom contexts. In fact, depending on the context, a prevention orientation (and perhaps a high confidence threshold) can actually be adaptive (Scholer & Higgins, 2012). For example, in a hostile classroom environment, acting conservatively by participating less frequently could serve as a protective mechanism that ensures a student's sense of belonging does not decrease even more.

Thus, it is important to interrogate the implicit assumption that if women merely behaved more like men in math classrooms by participating more frequently, it would inherently be a positive thing. Without taking the particular classroom environment into consideration, it is naïve to declare that encouraging a promotion focus with respect to participation would always lead to beneficial outcomes, particularly given that many college math classrooms are characterized by a "chilly climate" for women (Hall & Sandler, 1982; Lee & McCabe, 2021). Therefore, directing change toward college math classroom environments, in order to ensure that they offer safe spaces for participation, may be the most direct and impactful first step before encouraging women to participate more frequently in the existing contexts. In fact, a positive change in math classroom environments may itself minimize women's prevention focus with respect to participation, when it is not adaptive. Once the environment is made to be a positive one, more female students may adopt a promotion focus, resulting in reduced confidence thresholds and more frequent participation from previously more reticent students.

References

Abdullah, M. Y., Bakar, N. R. A., & Mahbob, M. H. (2012). Students' participation in classroom: What motivates them to speak up? *Procedia - Social and Behavioral Sciences, 51*, 516–522.

Aguillon, S. M., Siegmund, G. F., Petipas, R. H., Drake, A. G., Cotner, S., & Ballen, C. J. (2020). Gender differences in student participation in an active-learning classroom. *CBE—Life Sciences Education*, *19*(2), ar12.

Allen, K., Kern, M. L., Vella-Brodrick, D., Hattie, J., & Waters, L. (2018). What schools need to know about fostering school belonging: A meta-analysis. *Educational Psychology Review*, *30*(1), 1-34.

Alon, S., & Gelbgiser, D. (2011). The female advantage in college academic achievements and horizontal sex segregation. *Social Science Research*, *40*(1), 107-119.

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance, *Psychological Science, 12*, 385-390.

Anderson, R. (2007). Being a Mathematics Learner: Four Faces of Identity. *Mathematics Educator*, *17*(1), 7-14.

Aronson, J., & Salinas, M. F. (1997). *Stereotype threat, attributional ambiguity, and Latino underperformance.* Unpublished manuscript, University of Texas.

Atkins, W. J., Leder, G. C., O'Halloran, P. J., Pollard, G. H., & Taylor, P. (1991). Measuring risk taking. *Educational Studies in Mathematics*, *22*(3), 297-308.

Ayalon, H. (1995). Math as a gatekeeper: Ethnic and gender inequality in course taking of the sciences in Israel. *American Journal of Education*, *104*(1), 34-56.

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science*, *60*(2), 434-448.

Bauer, D. J., & Curran, P. J. (2012). *Introduction to structural equation modeling.* Personal collection of Curran-Bauer Analytics, LLC, Durham, NC.

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General, 130,* 701–725.

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened?. *Personality and Social Psychology Bulletin*, *32*(8), 1059-1071.

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, *133*(4), 584-600.

Belanger, A. L., Joshi, M. P., Fuesting, M. A., Weisgram, E. S., Claypool, H. M., & Diekman, A. B. (2020). Putting belonging in context: Communal affordances signal belonging in STEM. *Personality and Social Psychology Bulletin*, *46*(8), 1186-1204.

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*, *6*(1), 1-62.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389-391.

Böheim, R., Knogler, M., Kosel, C., & Seidel, T. (2020a). Exploring student hand-raising across two school subjects using mixed methods: An investigation of an everyday classroom behavior from a motivational perspective. *Learning and Instruction*, *65*, 101250.

Böheim, R., Urdan, T., Knogler, M., & Seidel, T. (2020b). Student hand-raising as an indicator of behavioral engagement and its role in classroom learning. *Contemporary Educational Psychology*, *62*, 101894.

Browman, A. S., Destin, M., & Molden, D. C. (2017). Identity-specific motivation: How distinct identities direct self-regulation across distinct situations. *Journal of Personality and Social Psychology*, *113*(6), 835-857.

Cafri, G., Van Den Berg, P., & Brannick, M. T. (2010). What have the difference scores not been telling us? A critique of the use of self—ideal discrepancy in the assessment of body image and evaluation of an alternative data-analytic framework. *Assessment*, *17*(3), 361-376.

Carter, A. J., Croft, A., Lukas, D., & Sandstrom, G. M. (2018). Women's visibility in academic seminars: Women ask fewer questions than men. *PloS one*, *13*(9), e0202743.

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*(6), 1045-1060.

Christenson, S. L., Sinclair, M. F., Lehr, C. A., & Godber, Y. (2001). Promoting successful school completion: Critical conceptual and methodological guidelines. *School Psychology Quarterly*, *16*(4), 468-484.

Clabaugh, A., Duque, J. F., & Fields, L. J. (2021). Academic stress and emotional well-being in United States college students following onset of the COVID-19 pandemic. *Frontiers in Psychology*, *12*, 628787.

Clark, N. (2014, April 28). Act now to shrink the confidence gap. *Forbes*. https://www.forbes.com/sites/womensmedia/2014/04/28/act-now-to-shrink-the-confidence-gap/

Coffman, K. B., & Klinowski, D. (2020). The impact of penalties for wrong answers on the gender gap in test scores. *Proceedings of the National Academy of Sciences*, *117*(16), 8794-8803.

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, *324*(5925), 400-403.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330-351.

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81,* 322–331.

Cribbs, J., Hazari, Z., Sonnert, G., & Sadler, P. M. (2020). College students' mathematics-related career intentions and high school mathematics pedagogy through the lens of identity. *MEdRJ*.

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, *24*(6), 588-594.

Crombie, G., Pyke, S. W., Silverthorn, N., Jones, A., & Piccinin, S. (2003). Students' perceptions of their classroom participation and instructor as a function of gender and context. *The Journal of Higher Education*, *74*(1), 51-76.

Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes*, *69*(2), 117-132.

Cullen, M. J., Waters, S. D., & Sackett, P. R. (2006). Testing stereotype threat theory predictions for math-identified and non-math-identified students by gender. *Human Performance*, *19*(4), 421–440.

Dallimore, E. J., Hertenstein, J. H., & Platt, M. B. (2004). Classroom participation and discussion effectiveness: Student-generated strategies. *Communication Education*, *53*(1), 103-115.

Deemer, E. D., Lin, C., Graham, R., & Soto, C. (2016). Development and validation of a measure of threatening gender stereotypes in science: A factor mixture analysis. *Journal of Career Assessment*, *24*(1), 145-161.

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science*, *21*(8), 1051-1057.

Dixon, J. K., Egendoerfer, L. A., & Clements, T. (2009). Do they really need to raise their hands? Challenging a traditional social norm in a second grade mathematics classroom. *Teaching and Teacher Education*, *25*(8), 1067-1076.

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology, 98*(2), 382-393.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215-225.

Eddy, S. L., Brownell, S. E., & Wenderoth, M. P. (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. *CBE—Life Sciences Education*, *13*(3), 478-492.

Ehrtmann, L., & Wolter, I. (2018). The impact of students' gender-role orientation on competence development in mathematics and reading in secondary school. *Learning and Individual Differences*, *61*, 256-264.

Elharake, J. A., Akbar, F., Malik, A. A., Gilliam, W., & Omer, S. B. (2022). Mental health impact of COVID-19 among children and college students: A systematic review. *Child Psychiatry & Human Development*, 1-13.

Eliasson, N., Sørensen, H., & Karlsson, K. G. (2016). Teacher–student interaction in contemporary science classrooms: is participation still a question of gender?. *International Journal of Science Education*, *38*(10), 1655-1672.

Ellison, G., & Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: Evidence from the American Mathematics Competitions. *Journal of Economic Perspectives*, *24*(2), 109-128.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, *136*(1), 103-127.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430-457.

Fassinger, P. A. (1995). Understanding classroom interaction: Students' and professors' contributions to students' silence. *The Journal of Higher Education*, *66*(1), 82-96.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116,* 429 – 456.

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*(1), 25-44.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109.

Freeman, T. M., Anderman, L. H., & Jensen, J. M. (2007). Sense of belonging in college freshmen at the classroom and campus levels. *The Journal of Experimental Education*, *75*(3), 203-220.

Freitas, A. L., & Higgins, E. T. (2002). Enjoying goal-directed action: The role of regulatory fit. *Psychological Science*, *13*(1), 1-6.

Fritschner, L. M. (2000). Inside the undergraduate college classroom: Faculty and students differ on the meaning of student participation. *The Journal of Higher Education*, *71*(3), 342-362.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*(1), 148-162.

Galyon, C. E., Blondin, C. A., Yaw, J. S., Nalls, M. L., & Williams, R. L. (2012). The relationship of academic self-efficacy to class participation and exam performance. *Social Psychology of Education*, *15*(2), 233-249.

George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update* (10a ed.) Boston: Pearson.

Gödöllei, A. F., & Beck, J. W. (2020). Development and validation of the state regulatory focus scale. *Human Performance*, 1-26.

Goldstein, G. S., & Benassi, V. A. (1994). The relation between teacher self-disclosure and student classroom participation. *Teaching of Psychology*, *21*(4), 212-217.

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*(4), 700-717.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206-213.

Grover Aukrust, V. (2008). Boys' and girls' conversational participation across four grade levels in Norwegian classrooms: Taking the floor or being given the floor?. *Gender and Education*, *20*(3), 237-252.

Guinier, L., Fine, M., & Balin, J. (1994). Becoming gentlemen: Women's experiences at one Ivy League law school. *University of Pennsylvania Law Review*, *143*, 1-110.

Gunthert, K. C., & Wenze, S. J. (2012). Daily diary methods. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (p. 144–159). The Guilford Press.

Hall, R. M., & Sandler, B. (1982). The classroom climate: A chilly one for women? *Project on the Status of Women*. Washington, DC: Association of American Colleges.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach* (2nd ed.). Guilford Press.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, *52*(12), 1280-1300.

Higgins, E. T., Roney, C. J., Crowe, E., & Hymes, C. (1994). Ideal versus ought predilections for approach and avoidance distinct self-regulatory systems. *Journal of Personality and Social Psychology*, *66*(2), 276-286.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*(5888), 494-495.

Inzlicht, M. (2016, February 29). Reckoning with the past. *Getting better*. http://michaelinzlicht.com/getting-better/2016/2/29/reckoning-with-the-past

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

Jackson, R. C., Ashford, K. J., & Norsworthy, G. (2006). Attentional focus, dispositional reinvestment, and skilled motor performance under pressure. *Journal of Sport and Exercise Psychology*, *28*(1), 49-68.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children’s self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509-527.

Jansen, A. (2006). Seventh graders’ motivations for participating in two discussion-oriented mathematics classrooms. *The Elementary School Journal*, *106*(5), 409-428.

Jetten, J., Postmes, T., & McAuliffe, B. J. (2002). ‘We're all individuals’: Group norms of individualism and collectivism, levels of identification and identity threat. *European Journal of Social Psychology*, *32*(2), 189-207.

Joia, L. A., & Lorenzo, M. (2021). Zoom in, zoom out: The impact of the COVID-19 pandemic in the classroom. *Sustainability*, *13*(5), 2531.

Jones, S. M., & Dindia, K. (2004). A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research*, *74*(4), 443-471.

Karabenick, S. A. (2003). Seeking help in large college classes: A person-centered approach. *Contemporary Educational Psychology, 28*(1), 37–58.

Kier, M. W., Blanchard, M. R., Osborne, J. W., & Albert, J. L. (2014). The development of the STEM career interest survey (STEM-CIS). *Research in Science Education*, *44*(3), 461-481.

Kline. (2015). *Principles and Practice of Structural Equation Modeling, Fourth Edition*. Guilford Publications.

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, *53*(6), 2576-2590.

Land, K. C. (1969). Principles of path analysis. *Sociological Methodology*, *1*, 3-37.

Leder, G. C. (1988). Teacher-student interactions: The mathematics classroom. *Unicorn*, *14*(3), 161-166.

Lee, J. J., & McCabe, J. M. (2021). Who speaks and who listens: Revisiting the chilly climate in college classrooms. *Gender & Society*, *35*(1), 32-60.

Lesko, A. C., & Corpus, J. H. (2006). Discounting the difficult: How high math-identified women respond to stereotype threat. *Sex Roles*, *54*(1-2), 113-125.

Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262-265.

Lewis, N. A. Jr., & Sekaquaptewa, D. (2016). Beyond test performance: A broader view of stereotype threat. *Current Opinion in Psychology*, *11*, 40-43.

Lock, R. M., Hazari, Z., & Potvin, G. (2013). Physics career intentions: The effect of physics identity, math identity, and gender. *Physics Education Research Conference*, *1513*(1), 262-265.

Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive or negative role models: regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology*, *83*(4), 854-864.

Mahalik, J. R., Locke, B. D., Ludlow, L. H., Diemer, M. A., Scott, R. P., Gottfried, M., & Freitas, G. (2003). Development of the conformity to masculine norms inventory. *Psychology of Men & Masculinity*, *4*(1), 3-25.

Mahalik, J. R., Morray, E. B., Coonerty-Femiano, A., Ludlow, L. H., Slattery, S. M., & Smiler, A. (2005). Development of the conformity to feminine norms inventory. *Sex Roles*, *52*(7-8), 417-435.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, *38*(1), 299-337.

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, *82*(4), 623-636.

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*(4), 645-657.

Molden, D. C., Lee, A. Y., & Higgins, E. T. (2008). Motivations for promotion and prevention. In J. Y. Shah & W. L. Gardner (Eds.), *Handbook of motivation science* (pp. 169–187). New York, NY: Guilford Press.

Molden, D. C., & Rosenzweig, E. Q. (2016). The origins and educational implications of promotion-focused and prevention-focused achievement motivations. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school* (pp. 477-503). Routledge.

Molden, D. C., & Winterheld, H. A. (2013). Motivations for promotion or prevention in close relationships. In J. A. Simpson & L. Campbell (Eds.), Oxford library of psychology. The Oxford handbook of close relationships (p. 321–347). Oxford University Press.

Morrow, J., & Ackermann, M. (2012). Intention to persist and retention of first-year students: The importance of motivation and sense of belonging. *College Student Journal*, *46*(3), 483-491.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*(10), 879-885.

Musu-Gillette, L. E., Wigfield, A., Harring, J. R., & Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation*, *21*(4), 343-370.

Narayan, J. S., Heward, W. L., Gardner, R., Courson, F. H., & Omness, C. K. (1990). Using response cards to increase student participation in an elementary classroom. *Journal of Applied Behavior Analysis*, *23*(4), 483-490.

National Science Foundation, National Center for Science and Engineering Statistics. (2019). Women, minorities, and persons with disabilities in science and engineering (NSF 11 309). https://ncses.nsf.gov/pubs/nsf19304/digest.

Neuville, E., & Croizet, J. C. (2007). Can salience of gender identity impair math performance among 7–8 years old girls? The moderating role of task difficulty. *European Journal of Psychology of Education*, *22*(3), 307-316.

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*(6), 1314-1334.

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

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., . . . Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 106*(26), 10593–10597.

O'Connor, K. (2013). Class participation: Promoting in-class student engagement. *Education*, *133*(3), 340-344.

O'Dea, R. E., Lagisz, M., Jennions, M. D., & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications*, *9*(1), 3777-3784.

Oh, Y. J., Jia, Y., Lorentson, M., & LaBanca, F. (2013). Development of the educational and career interest scale in science, technology, and mathematics for high school students. *Journal of Science Education and Technology*, *22*(5), 780-790.

Ohannessian, C. M., Flannery, K. M., Simpson, E., & Russell, B. S. (2016). Family functioning and adolescent alcohol use: A moderated mediation analysis. *Journal of Adolescence*, *49*, 19-27.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement?. *Contemporary Educational Psychology*, *26*(3), 291-310.

Oyserman, D., Uskul, A. K., Yoder, N., Nesse, R. M., & Williams, D. R. (2007). Unfair treatment and self-regulatory focus. *Journal of Experimental Social Psychology*, *43*(3), 505-512.

Rainey, K., Dancy, M., Mickelson, R., Stearns, E., & Moller, S. (2018). Race and gender differences in how sense of belonging influences decisions to major in STEM. *International Journal of STEM Education*, *5*(1), 10-23.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203-212.

Ramos, I., & Lambating, J. (1996). Gender Difference in Risk-Taking Behavior and their Relationship to SAT-Mathematics Performance. *School Science and Mathematics*, *96*(4), 202-207.

Reardon, S. F., Kalogrides, D., Fahle, E. M., Podolsky, A., & Zárate, R. C. (2018). The relationship between test item format and gender achievement gaps on math and ELA tests in fourth and eighth grades. *Educational Researcher*, *47*(5), 284-294.

Rosenzweig, E. Q., & Miele, D. B. (2016). Do you have an opportunity or an obligation to score well? The influence of regulatory focus on academic test performance. *Learning and Individual Differences*, *45*, 114-127.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.

Sadker, M., & Sadker, D. (1986). Sexism in the classroom: From grade school to graduate school. *The Phi Delta Kappan*, *67*(7), 512-515.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*(2), 336-356.

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology, 94,* 168–182.

Scholer, A. A., & Higgins, E. T. (2012). Too much of a good thing? Trade-offs in promotion and prevention focus. In R. M. Ryan (Ed.), *The Oxford Handbook of Human Motivation* (pp. 65–84). Oxford University Press.

Scholer, A. A., Miele, D. B., Murayama, K., & Fujita, K. (2018). New directions in self-regulation: The role of metamotivational beliefs. *Current Directions in Psychological Science*, *27*(6), 437-442.

Seibt, B., & Förster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology*, *87*(1), 38-56.

Shah, J., Higgins, T., & Friedman, R. S. (1998). Performance incentives and means: how regulatory focus influences goal attainment. *Journal of Personality and Social Psychology*, *74*(2), 285-293.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*(1), 80-83.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*(4), 571-581.

Smith, J. L., Lewis, K. L., Hawthorne, L., & Hodges, S. D. (2013). When trying hard isn't natural: Women's belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns. *Personality and Social Psychology Bulletin*, *39*(2), 131-143.

Smyth, E., & Steinmetz, S. (2008). Field of study and gender segregation in European labour markets. *International Journal of Comparative Sociology*, *49*(4-5), 257-281.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: A critical review. *American Psychologist*, *60*(9), 950-958.

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

Ståhl, T., Van Laar, C., & Ellemers, N. (2012). The role of prevention focus under stereotype threat: Initial cognitive mobilization is followed by depletion. *Journal of Personality and Social Psychology*, *102*(6), 1239-1251.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613-629.

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

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In *Advances in experimental social psychology* (Vol. 34, pp. 379-440). Academic Press.

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement?. *Review of General Psychology*, *16*(1), 93-102.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, *100*(2), 255-270.

Stowell, J. R., & Nelson, J. M. (2007). Benefits of electronic audience response systems on student participation, learning, and emotion. *Teaching of Psychology*, *34*(4), 253-258.

Streiner, D. L. (2005). Finding our way: an introduction to path analysis. *The Canadian Journal of Psychiatry*, *50*(2), 115-122.

Streitmatter, J. (1997). An exploratory study of risk-taking and attitudes in a girls-only middle school math class. *The Elementary School Journal*, *98*(1), 15-26.

Subotnik, R. F., & Strauss, S. M. (1995). Gender Differences in Classroom Participation and Achievement: An Experiment Involving Advanced Placement Calculus Classes. *Journal of Secondary Gifted Education*, *6*(2), 77-85.

Tannenbaum, D. I. (2012). *Do gender differences in risk aversion explain the gender gap in SAT scores? Uncovering risk attitudes and the test score gap.* Unpublished manuscript, University of Chicago.

Thoman, D. B., Arizaga, J. A., Smith, J. L., Story, T. S., & Soncuya, G. (2014). The grass is greener in non-science, technology, engineering, and math classes: Examining the role of competing belonging to undergraduate women's vulnerability to being pulled away from science. *Psychology of Women Quarterly*, *38*(2), 246-258.

Thomas, B., & Costello, J. (1988). Identifying attitudes to mathematics. *Mathematics Teaching*, *122*, 62-64.

Thomas, D. R., & Zumbo, B. D. (2012). Difference scores from the point of view of reliability and repeated-measures ANOVA: In defense of difference scores for data analysis. *Educational and Psychological Measurement*, *72*(1), 37-43.

Thomas, N., Harel, O., & Little, R. J. (2016). Analyzing clinical trial outcomes based on incomplete daily diary reports. *Statistics in Medicine*, *35*(17), 2894-2906.

Trafimow, D. (2015). A defense against the alleged unreliability of difference scores. *Cogent Mathematics*, *2*(1), 1064626.

Turner, J. C., & Patrick, H. (2004). Motivational influences on student participation in classroom learning activities. *Teachers College Record*, *106*(9), 1759-1785.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-70.

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, *78*(3), 516-551.

Walton, G. M., & Brady, S. T. (2017). *The many questions of belonging.* In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (p. 272–293). The Guilford Press.

Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review*, *33*(4), 304-340.

Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N., & Turrou, A. C. (2014). Engaging with others' mathematical ideas: Interrelationships among student participation, teachers' instructional practices, and learning. *International Journal of Educational Research, 63*, 79-93.

Weston, R., Gore, P. A., Chan, F., & Catalano, D. (2008). An introduction to using structural equation models in rehabilitation psychology. *Rehabilitation Psychology*, *53*(3), 340-356.

Wickens, T. D., & Keppel, G. (2004). *Design and analysis: A researcher's handbook*. Upper Saddle River, NJ: Pearson Prentice-Hall.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68-81.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., & Blumenfeld, P. C. (1997). Changes in children's competence beliefs and subjective task values across the elementary school years: A three-year study. *Journal of Educational Psychology*, *89*(3), 451–469.

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. In *Advances in Experimental Social Psychology* (Vol. 46, pp. 55-123). Academic Press.

Wu, F., & Fan, W. (2017). Academic procrastination in linking motivation and achievement-related behaviours: a perspective of expectancy-value theory. *Educational Psychology*, *37*(6), 695-711.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*(2), 267-301.

Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology, 29*(5), 703–709.

Zou, X., Scholer, A. A., & Higgins, E. T. (2020). Risk preference: How decision maker's goal, current value state, and choice set work together. *Psychological Review*, *127*(1), 74-94.

Zumbrunn, S., McKim, C., Buhs, E., & Hawley, L. R. (2014). Support, belonging, motivation, and engagement in the college classroom: A mixed method study. *Instructional Science*, *42*(5), 661-684.

Appendix A
Pilot Study Materials

Table A1

Measures Included in Pilot Studies 1–4

Measure	Description & Sample Items	Pilot 1	Pilot 2	Pilot 3	Pilot 4
Stereotype threat susceptibility (1)	3-item Likert-type scale adapted from Marx and Goff (2005) e.g., “I worry that my ability to perform well on math tasks is affected by my gender.”	X	X		
Stereotype threat susceptibility (2)	Scenario-based measure in which participants are instructed to imagine they are sitting in a class and are surrounded by an instructor and peers all of a given gender (male or female). Followed by 3 Likert-type agreement items adapted from Marx and Goff (2005). Operationalized as a difference score of responses in opposite-gender scenario and same-gender scenario e.g., “I would worry about what my professor’s perception of my math ability would be if I performed poorly on the exam.”				X
Stereotype threat susceptibility (3)	7-item Stereotype Threat in Math Scale by Deemer et al. (2016) e.g. “I am afraid that if I do poorly in this class, it will confirm the stereotype that members of my gender group cannot be good in math.”				X
Regulatory focus orientation	6-item scale adapted from Browman et al. (2017) – 3 items measuring prevention and 3 assessing promotion e.g., “I have important academic standards in my math class that I focus on maintaining.”	X	X	X	X
CPCT-1 (instructor-prompted version)	Vignette description of a classroom situation in which an instructor poses a question to the class, followed by a continuous slider ranging from 0 to 100%, with answer options of whole percentages, on which participants declared the level of confidence they would require in their response before raising their hand to participate	X	X	X	X

Measure	Description	Pilot 1	Pilot 2	Pilot 3	Pilot 4
CPCT-2 (instructor-prompted version)	3-item Likert-type measure of willingness to raise one's hand (based on a scenario in which an instructor poses a question to the class) e.g., "I tend to raise my hand even if I am not at all sure that I know the correct answer."		X	X	X
CPCT-1 (general version)	Similar to the specific version of CPCT-1 (slider measure), updated to incorporate a wider range of participation behaviors				X
CPCT-2 (general version)	Similar to the specific version of CPCT-2; prompt updated to incorporate a wider variety of participation behaviors e.g., "I usually participate only when I am absolutely certain that my contribution is valid."				X
Self-reported participation behaviors (1)	2 Likert-type items measuring frequency of participating and frequency of changing one's mind about participating e.g., "How often do you raise your hand to answer a question in math class?"; "On occasions when you decide that you do want to raise your hand to answer a question in your math class, how often do you change your mind?"	X	X	X	
Self-reported participation behaviors (2)	6 Likert-type items measuring frequency of participating and frequency of changing one's mind about participating (intended to encompass a wider variety of participation behaviors than were presented in Pilot Studies 1-3, including participation in small groups) e.g., "When no question has been asked, how often do you raise your hand to make a comment or provide an explanation in your math class?"				X
Self-confidence in domain (1)	6 items from the Academic Self-Description Questionnaire (Marsh, 1990) e.g., "Work in math classes is easy for me."	X	X	X	X
Self-confidence in domain (2)	Percentage (slider) format of confidence in domain-specific ability e.g., "In general, how confident are you in your math ability? For example, 0% confident = not at all confident, 75% confident = mostly confident, 100% confident = completely confident."		X		

Measure	Description	Pilot 1	Pilot 2	Pilot 3	Pilot 4
Participation confidence (1)	Designed to align with CPCT-1 (slider) measure of confidence threshold. Asks participants to rate how confident they have generally felt in the past about knowing the correct answer when their instructor has posed a question to the class e.g., “In the past, when your math instructor has asked a question to the class about a problem on the board, how confident did you generally feel about knowing the correct answer? … Please mark your response on the slider below.”		X	X	
Participation confidence (2)	Designed to align with CPCT-2 (Likert-type) measure of confidence threshold. Measure of level of confidence in knowing the correct answer of a question posed by one’s instructor (instructor-prompted version) and of feeling confident in the validity of one’s contribution (general version) e.g., “I tend to feel very certain that I know the right answer” [instructor-prompted version]; “I tend to feel very certain that my contribution is valid.”				X
Perceived domain-specific value	6-item scale adapted from Eccles and Wigfield (1995) to measure utility value, intrinsic value, and importance e.g., “How important is it to know math to get a good job?”	X	X	X	X
Perceived costs and value of participation	6-item Likert-type scale intended to measure perceived costs and value associated with classroom participation; updated to include two additional items related to value for Pilot 4 e.g., “Participating in math class is important to me.”	X	X		X
Sense of belonging in domain	18-item Math Sense of Belonging Scale (Good et al., 2012); also adapted for the psychology domain in Pilot 4 e.g., “When I am in a math setting, I feel like I am part of the mathematics community.”			X	X
STEM career interest	12-item STEM Career Interest Survey (STEM-CIS), Mathematics subscale (Kier et al., 2014) e.g., “I plan to use mathematics in my future career.”			X	

Measure	Description	Pilot 1	Pilot 2	Pilot 3	Pilot 4
Educational career interest in STEM	Educational Career Interest Scale in Science, Technology, and Mathematics (Oh et al., 2013). All 4 subscales included in Pilot 3; only math subscale included in Pilot 4 e.g., "I am interested in working in a career that allows me to use mathematics-related skills or knowledge."			X	X
Domain-specific identity	6 items adapted from Lesko and Corpus (2006) e.g., "Being good at math is *not* an important part of who I am."			X	X
State regulatory focus	8-item State Regulatory Focus Scale by Gödöllei & Beck (2020) specifying what participants are focused on at the moment of responding (included as a manipulation check) e.g., "At this moment, I am focused on minimizing losses."			X	
Chronic (general) regulatory focus	11-item Regulatory Focus Questionnaire (RFQ) by Higgins et al. (2001) (included as a manipulation check) e.g., "Not being careful enough has gotten me into trouble at times."			X	
Conformity to gender norms	6 items adapted from the Modesty and Risk-Taking scales from the Conformity to Masculine Norms Inventory and Conformity to Feminine Norms Inventory (Mahalik et al., 2003, 2005) e.g., "I always downplay my achievements" [modesty]; "I enjoy taking risks" [risk-taking]	X	X		
Help-seeking behaviors	12 items from a help-seeking scale by Karabenick (2003) e.g., "I would feel like a failure if I needed help in a math class."		X		
Personality	10 items comprising the Big 5 Inventory-10 by Rammstedt & John (2007) e.g., "I see myself as someone who is reserved."		X		

Note: Unless otherwise specified, all measures were presented with respect to both the math and contrasting domains.

Regulatory Focus Manipulation Instructions
Pilot Study 3

General Induction

Promotion Condition:

Hopes and Aspirations

For this task, I would like you to think about how your current hopes and aspirations are different now from what they were when you were growing up. In other words, what accomplishments would you ideally like to meet at this point in your life? What accomplishments did you ideally want to meet when you were a child? In the space on the next screen, please write a brief essay describing how your hopes and aspirations have changed from when you were a child to now.

You will have at least three minutes to reflect and write. After three minutes, the "Continue" arrow will appear on your screen, and you can continue to the next page of the survey. You can stay on the page for a maximum of five minutes, at which time the website will automatically advance you to the next screen.

Prevention Condition:

Duties and Obligations

For this task, I would like you to think about how your current duties and obligations are different now from what they were when you were growing up. In other words, what responsibilities do you think you ought to meet at this point in your life? What responsibilities did you think you ought to meet when you were a child? In the space on the next screen, please write a brief essay describing how your duties and obligations have changed from when you were a child to now.

You will have at least three minutes to reflect and write. After three minutes, the "Continue" arrow will appear on your screen, and you can continue to the next page of the survey. You can stay on the page for a maximum of five minutes, at which time the website will automatically advance you to the next screen.

Academic-Specific Induction

Promotion Condition:

Now that you have written about your general hopes and aspirations, I would like you to think about the academic aspirations you have as a college student. You will have approximately 2 minutes to write about these more specific aspirations. If you already addressed your academic aspirations on the previous screen, please briefly reiterate them here.

You will have at least two minutes to reflect and write. After two minutes, the "Continue" arrow will appear on your screen, and you can continue to the next page of the survey. You can stay on the page for a maximum of three minutes, at which time the website will automatically advance you to the next screen.

[Next screen]: Please think about something you ideally would like to accomplish in your classes as a BC student. In other words, think about an academic hope or aspiration that you currently have. Reflect on this aspiration below.

Prevention Condition:
Now that you have written about your general duties and obligations, I would like you to think about the academic obligations you have as a college student. You will have approximately 2 minutes to write about these more specific obligations. If you already addressed your academic obligations on the previous screen, please briefly reiterate them here.

You will have at least two minutes to reflect and write. After two minutes, the "Continue" arrow will appear on your screen, and you can continue to the next page of the survey. You can stay on the page for a maximum of three minutes, at which time the website will automatically advance you to the next screen.

[Next screen]: Please think about something you think you ought to do in your classes as a BC student. In other words, think about an academic duty or obligation that you currently have. Reflect on this obligation below.

Appendix B
Pilot Study Results

Table B1

Correlations Between Pilot Study 1 Variables (Math)

	1	*2*	*3*	*4*	*5*
Stereotype threat					
1. Stereotype threat in math contexts	1				
Regulatory focus					
2. Math-specific RF	.173*	1			
Participation behaviors					
3. Frequency of participation	-.095	-.247**	1		
4. Frequency of changing one's mind	.330***	.240**	-.119	1	
Class participation confidence thresholds					
5. CPCT-1 (slider)	-.188*	-.040	-.083	.035	1

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

Table B2

Correlations Between Pilot Study 1 Variables (Social Studies)

	1	*2*	*3*	*4*	*5*
Stereotype threat					
1. Stereotype threat in SS contexts	1				
Regulatory focus					
2. SS-specific RF	.100	1			
Participation behaviors					
3. Frequency of participation	-.011	-.253**	1		
4. Frequency of changing one's mind	.281**	.212**	-.136+	1	
Class participation confidence thresholds					
5. CPCT-1 (slider)	-.056	-.047	-.153+	.099	1

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

Table B3

Correlations Between Pilot Study 2 Variables (Math)

	1	*2*	*3*	*4*	*5*	*6*
Stereotype threat						
1. Stereotype threat in math contexts	1					
Regulatory focus						
2. Math-specific RF	.085	1				
Participation behaviors						
3. Frequency of participation	.065	-.277***	1			
4. Frequency of changing one's mind	-.070	.235***	-.285***	1		
Class participation confidence thresholds						
5. CPCT-1 (slider)	.011	.123	-.504***	.158*	1	
6. CPCT-2 (Likert)	.040	.327***	-.619***	.267***	.668***	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table B4

Correlations Between Pilot Study 2 Variables (Psychology)

	1	*2*	*3*	*4*	*5*	*6*
Stereotype threat						
1. Stereotype threat in psych contexts	1					
Regulatory focus						
2. Psych-specific RF	.077	1				
Participation behaviors						
3. Frequency of participation	-.053	-.256**	1			
4. Frequency of changing one's mind	.143	.125	-.315***	1		
Class participation confidence thresholds						
5. CPCT-1 (slider)	.026	.274***	-.378***	.143+	1	
6. CPCT-2 (Likert)	.061	.247**	-.549***	.253**	.646***	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table B5

Correlations Between Pilot Study 3 Variables (Math)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Regulatory focus								
1. Math-specific RF	1							
Participation behaviors								
2. Frequency of participation	$-.169^{+}$	1						
3. Frequency of changing one's mind	$.230^{*}$	$-.472^{***}$	1					
Class participation confidence thresholds								
4. CPCT-1 (slider)	$.160^{+}$	$-.512^{***}$	$.326^{***}$	1				
5. CPCT-2 (Likert)	$.243^{**}$	$-.516^{***}$	$.409^{***}$	$.813^{***}$	1			
Outcome measures								
6. Math identity	$-.334^{***}$	$.222^{*}$	-.099	.043	.012	1		
7. Sense of belonging	$-.387^{***}$	$.588^{***}$	$-.514^{***}$	$-.243^{**}$	$-.287^{**}$	$.473^{***}$	1	
8. Math-related career interest	$-.436^{***}$	$.184^{*}$	-.071	-.008	-.094	$.689^{***}$	$.466^{***}$	1

Note. $^{+} p < .10$, $^{*} p < .05$, $^{**} p < .01$, $^{***} p < .001$

Table B6

Correlations Between Pilot Study 4 Variables (Math)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Stereotype threat												
1. Math ST (difference score)	1											
2. Math ST (Deemer et al. measure)	.361***	1										
Regulatory focus												
3. Math-specific RF	-.005	.097	1									
Participation behaviors												
4. Frequency (answer question)	-.095	-.099	-.335**	1								
5. Freq. change mind (answer ques.)	.035	.116	.169*	-.366***	1							
6. Frequency (comment/explanation)	.022	-.062	-.400**	.629***	-.359***	1						
7. Freq. change mind (comm./explan.)	.009	.103	-.012	-.128+	.439***	-.200**	1					
Class participation confidence thresholds ("general" version)												
8. CPCT-1 (slider)	-.001	.177**	.271**	-.451***	.293***	-.472***	.154*	1				
9. CPCT-2 (Likert)	.013	. 184**	.291**	-.542***	.339***	-.593***	.192**	.622***	1			
Outcome measures												
10. Math identity	-.063	-.018	-.326**	.434***	-.117**	.289***	-.017	-.153*	-.149*	1		
11. Sense of belonging	-.074	-.285**	-.507***	.596***	-.297***	.486***	-.061	-.357***	-.437***	.622***	1	
12. Math-related career interest	.051	-.006	-.414***	.382***	-.110	.279***	.041	-.216**	-.199**	.641***	.541***	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table B7

Correlations Between Pilot Study 4 Variables (Psychology)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Stereotype threat												
1. Psych ST (difference score)	1											
2. Psych ST (Deemer et al. measure)	$.272^{***}$	1										
Regulatory focus												
3. Psych-specific RF	.023	$.154^{*}$	1									
Participation behaviors												
4. Frequency (answer question)	-.056	$-.119^{+}$	$-.257^{***}$	1								
5. Freq. change mind (answer ques.)	.039	$.182^{**}$	.102	$-.318^{***}$	1							
6. Frequency (comment/explanation)	-.045	-.090	$-.298^{***}$	$.626^{***}$	$-.221^{***}$	1						
7. Freq. change mind (comm./explan.)	.071	$.198^{**}$	.098	$-.118^{+}$	$.440^{***}$	-.112	1					
Class participation confidence thresholds ("general" version)												
8. CPCT-1 (slider)	-.048	-.005	.077	$-.292^{***}$	.103	$-.383^{***}$	.050	1				
9. CPCT-2 (Likert)	-.096	$.164^{*}$	$.254^{***}$	$-.527^{***}$	$.313^{***}$	$-.479^{***}$	$.169^{*}$	$.587^{**}$	1			
Outcome measures												
10. Psych identity	-.071	.082	$-.216^{**}$	$.178^{**}$	-.072	$.118^{+}$	-.072	.006	$-.152^{*}$	1		
11. Sense of belonging	-.001	$-.197^{**}$	$-.360^{***}$	$.426^{***}$	$-.270^{***}$	$.404^{***}$	-.109	$-.264^{**}$	$-.432^{***}$	$.528^{***}$	1	
12. Psych-related career interest	-.070	.099	$-.292^{***}$	$.291^{***}$	-.019	$.202^{**}$	.001	-.044	$-.161^{*}$	$.686^{***}$	$.449^{***}$	1

Note. $^{+} p < .10$, $^{*} p < .05$, $^{**} p < .01$, $^{***} p < .001$

Table B8

Correlations Between General and Instructor-Prompted Variables (Pilot Study 4 – Math)

	1	*2*	*3*	*4*	*5*	*6*
CPCT-1						
1. CPCT-1 (general)	1					
2. CPCT-1 (instructor-prompted)	**.796*****	1				
CPCT-2						
3. CPCT-2 (general)	.622***	.591***	1			
4. CPCT-2 (instructor-prompted)	.592***	.687***	**.752*****	1		
Participation-specific confidence						
5. General	-.253***	-.211**	-.221**	-.271***	1	
6. Instructor-prompted	-.300***	-.297***	-.246***	-.279***	**.680*****	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table B9

Correlations Between General and Instructor-Prompted Variables (Pilot Study 4 – Psychology)

	1	*2*	*3*	*4*	*5*	*6*
CPCT-1						
1. CPCT-1 (general)	1					
2. CPCT-1 (instructor-prompted)	**.740*****	1				
CPCT-2						
3. CPCT-2 (general)	.587***	.526***	1			
4. CPCT-2 (instructor-prompted)	.504***	.632***	**.723*****	1		
Participation-specific confidence						
5. General	-.244***	-.169*	-.375***	-.333***	1	
6. Instructor-prompted	-.121^{+}	-.075	-.206**	-.150*	**.591*****	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Appendix C
Study 1 Measures

ENROLLMENT FORM

Thank you for your interest in participating in this study! Please enter your information below. If you are selected to participate, you will receive an email containing a link to the study survey in several weeks.

- First Name: ____________
- Last Name: ____________
- BC email address: ____________
- BC Eagle ID number: ____________
- Gender: ____________
- Major(s): ____________
- Class year: *Freshman, Sophomore, Junior, Senior, Other (please explain)*
- Are you 18 years or older? *Yes, No*
- Have you participated in past versions of the *Class Participation* study via the Lynch School Sona pool research participation requirement? (Please note that if you have participated in a prior version of this study, you will not be eligible to participate in this iteration.) *Yes, No*
- Please list your current courses below:

	Course Title	**Course Number**
Course #1	*e.g., Finite Probability and Applications*	*e.g., MATH1004*
Course #2		
Course #3		
Course #4		
Course #5		
Course #6		

Thank you! We will contact you in several weeks if you are selected to participate in this study.

MAIN SURVEY

Selection of Courses
In this survey, you will be asked to reflect on your experiences in a math class or a social science class in which you are currently enrolled.

Based on the survey you completed several weeks ago, we have determined which of your courses qualify as math and/or social science courses.

[If I determined that participants were enrolled in only one eligible math course and one eligible social science course]:

- **Math course:** [Name of math course selected by me]

- **Social science course:** [Name of social science course selected by me]

  As you are completing this survey, please keep your experiences within these two courses in mind.

[If participants were currently enrolled in more than one math course and/or more than one eligible social science course, I invited them to select which course(s) they would like to reference throughout the study]:

Before completing the following survey, please identify which **MATH [SOCIAL SCIENCE]** course you would like to refer to throughout the survey. When selecting a course, **please consider which courses offer the most opportunities for classroom participation via hand-raising**.

Please select which math course you would like to refer to: *[options presented as multiple choice]*
Please select which social science course you would like to refer to: *[options presented as multiple choice]*

As you are completing this survey, please keep your experiences within these two courses in mind.

[Next page]: For the next set of questions, please think about your experiences in your **MATH [SOCIAL SCIENCE] class, [name of course].**

Stereotype Threat Susceptibility – Scenario-Based Measures (ST-1)
Response scale: Strongly disagree (1) to Strongly agree (6)

[Male scenario:]
Imagine that you are sitting in your **math [social science]** class, **[name of course]**, and are about to take a midterm. You are surrounded mostly by **male** peers and have a **male** professor. Please rate the extent to which you agree or disagree with the statements below.

[Female scenario:]
Imagine that you are sitting in your **math [social science]** class, **[name of course]**, and are about to take a midterm. You are surrounded mostly by **female** peers and have a **female** professor. Please rate the extent to which you agree or disagree with the statements below.

1. I would worry about my ability to perform well on the exam.
2. I would worry about what my professor's perception of my math [social science] ability would be if I performed poorly on the exam.
3. I would worry about what my peers' evaluations of my math [social science] ability would be if I performed poorly on the exam.

Stereotype Threat in Math [Social Science] (ST-2)

<u>Response scale</u>: Strongly disagree (1) to Strongly agree (6)

1. I feel pressure to do what I can to improve the image of my gender group in math [social science].
2. I am afraid that if I do poorly in this class, it will confirm the stereotype that members of my gender group cannot be good in math [social science].
3. I fear that members of the opposite gender in this class will judge me as being incompetent in math [social science] if I do not do well on the assignments and exams.
4. I feel pressure to do what I can to change the negative stereotype about my gender being generally weak in math [social science].
5. I am afraid of being negatively evaluated by members of the opposite gender in this class.
6. I am afraid of confirming the stereotype that members of my gender group do not have the skills to be mathematicians [social scientists].
7. I feel pressure to represent my gender group in math [social science] because there are so few of us in the field.

Domain-Specific Regulatory Focus

<u>Response scale</u>: Strongly disagree (1) to Strongly agree (6)

1. I worry about making mistakes in my math [social science] work when I am taking a test or doing homework. *[prevention]*
2. I frequently think about how I can prevent failures in my math [social science] classes. *[prevention]*
3. I frequently think about making progress toward being successful in math [social science]. *[promotion]*
4. I frequently imagine how I will achieve my math [social science]-related hopes and aspirations. *[promotion]*
5. I have important academic standards in my math [social science] class that I focus on maintaining. *[prevention]*
6. When I see an opportunity for achieving a math [social science] goal, I get excited right away. *[promotion]*

Class Participation Confidence Threshold – Slider Measure (CPCT-1)

In your **math [social science]** class, **[name of course]**, there are a number of ways in which you might participate. For instance, you might raise your hand to answer a question posed by the instructor, make a comment, share your experience, or provide an explanation.

Some students will choose to participate even if they are ***not at all sure*** that their contribution is valid. Other students will only raise their hand if they are ***completely sure*** that their contribution is valid.

What about you? If you had an opportunity to contribute in your **math [social science]** class, how sure of your contribution would you need to be in order to raise your hand? For example:

- **0% sure:** I would choose to participate even if I was **not at all sure** that my contribution was valid.
- **25% sure:** I would choose to participate as long as I was **a little bit sure** that my contribution was valid.
- **50% sure:** I would choose to participate as long as I was **half sure** that my contribution was valid.
- **75% sure:** I would choose to participate as long as I was **mostly sure** that my contribution was valid.
- **100% sure:** I would choose to participate only if I was **completely sure** that my contribution was valid.

Please mark your response on the slider: *[Display slider with five anchors above marked; slider will notify participant of selected percentage from 0-100]*

Class Participation Confidence Threshold – Likert-Type Measure (CPCT-2)
Response scale: Strongly disagree (1) to Strongly agree (6)

In your **math [social science]** class, **[name of course]**, there are a number of ways in which you might participate. For instance, you might raise your hand to answer a question posed by the instructor, make a comment, share your experience, or provide an explanation. Please rate the extent to which you agree or disagree with the statements below.

1. I generally do NOT participate unless I feel completely sure that my contribution is valid.
2. I tend to participate even if I am not at all sure that my contribution is valid.
3. I usually participate only when I am absolutely certain that my contribution is valid.

Self-Reported Class Participation Behaviors
Response scale: Never, Very rarely, Rarely, Occasionally, Often, Very often

1. How often do you raise your hand to answer a question posed by the instructor in your math [social science] class?
2. When no question has been asked, how often do you raise your hand to make a comment, share your experience, or provide an explanation in your math [social science] class?
3. On occasions when you decide that you want to raise your hand to answer a question posed by the instructor in your math [social science] class, how often do you change your mind?
4. On occasions when no question has been asked, and you decide that you want to raise your hand to make a comment, share your experience, or provide an

explanation in your math [social science] class, how often do you change your mind?

Opportunities for Participation

Response scale: Never, Very rarely, Rarely, Occasionally, Often, Very often

1. How often does the instructor of your math [social science] course pose questions to the class?
2. How often do you think students in your math [social science] course have an opportunity to participate?

Self-Confidence in Math [Social Science]

Response scale: False, Mostly false, More false than true, More true than false, Mostly true, True

1. Compared to others, I am good at math [social science].
2. Work in math [social science] classes is easy for me.
3. I learn things quickly in math [social science].
4. I get good grades in math [social science].
5. I'm hopeless when it comes to math [social science].
6. I have always done well in math [social science].

Participation-Specific Confidence in Math [Social Science]

Response scale: Strongly disagree (1) to Strongly agree (6)

Imagine that you are sitting in **math [social science]** class, **[name of course]**, and you have an opportunity to contribute (e.g., to answer a question posed by the instructor, make a comment, share your experience, or provide an explanation). Please rate the extent to which you agree or disagree with the statements below.

1. I generally do NOT feel certain that my contribution is valid.
2. I tend to feel very certain that my contribution is valid.
3. I am usually absolutely certain that my contribution is valid.

Perceived Value of Math [Social Science]

1. How useful is math [social science] – how much can it help you do better outside of class? *(Not at all useful, Not that useful, A little bit useful, Useful, Very useful)*
2. How important is it to know math [social science] in order to get a good job? *(Not at all important, Not that important, A little bit important, Important, Very important)*
3. How important is math [social science] compared to other school subjects? *(The least important subject, Less important than most other subjects, Equally important to most other subjects, More important than most other subjects, The most important subject)*
4. How important is it to be good at math [social science]? *(Not at all important, Not that important, A little bit important, Important, Very important)*

5. How interesting is math [social science] for you? *(Not at all interesting, Not that interesting, A little bit interesting, Interesting, Very interesting)*
6. How much do you enjoy doing math [social science]? *(I do not enjoy it at all, I do not enjoy it that much, I enjoy it a little bit, I enjoy it, I enjoy it a lot)*

Costs and Value of Participating in Class
Response scale: Strongly disagree (1) to Strongly agree (6)

1. Participating in math [social science] class is important to me.
2. I find participating in math [social science] class to be enjoyable.
3. I find participating in math [social science] class to be useful.
4. Participating in math [social science] class makes me feel anxious.
5. In math [social science] class, I'm concerned that if I answer the professor's question incorrectly, my professor will think I am dumb.
6. In math [social science] class, I'm concerned that if I answer the professor's question incorrectly, other students in my class will think I am dumb.
7. Raising my hand in math [social science] class requires too much effort.
8. In math [social science] class, I am able to pay attention to the lesson more when I am not participating.

Math [Social Science] Sense of Belonging
Response scale: Strongly disagree (1) to Strongly agree (6)

When I am in a **math [social science]** setting…

1. …I feel that I belong to the math [social science] community.
2. …I consider myself a member of the math [social science] world.
3. …I feel like I am part of the math [social science] community.
4. …I feel a connection with the math [social science] community.
5. …I feel like an outsider.
6. …I feel accepted.
7. …I feel respected.
8. …I feel disregarded.
9. …I feel valued.
10. …I feel neglected.
11. …I feel appreciated.
12. …I feel excluded.
13. …I feel like I fit in.
14. …I feel insignificant.
15. …I wish I could fade into the background and not be noticed.
16. …I try to say as little as possible.
17. …I enjoy being an active participant.
18. …I wish I were invisible.

Math [Social Science] Identity
Response scale: Strongly disagree (1) to Strongly agree (6)

1. Being good at math [social science] is *not* an important part of who I am.
2. I consider myself to be a math [social science] person.
3. Math [social science] is an important part of my identity.

Mathematics [Social Science] Career Interest

<u>Response scale</u>: *Not at all true (1) to Very true (6)*

1. I am interested in taking courses that help me learn more about mathematics [social science].
2. I am interested in working in a career that allows me to use mathematics-related [social science-related] skills or knowledge.
3. I would like to learn mathematics-related [social science-related] knowledge and skills because they can be useful to help me be prepared for a career.

Demographic Questions

- What grades do you typically receive in **<u>math</u>** classes? *(Mostly As, Mostly Bs, Mostly Cs, Mostly Ds, Mostly Fs)*
- What grades do you typically receive in **<u>social science</u>** classes? *(Mostly As, Mostly Bs, Mostly Cs, Mostly Ds, Mostly Fs)*
- What grades do you typically receive in general? *(Mostly As, Mostly Bs, Mostly Cs, Mostly Ds, Mostly Fs)*
- What is your gender? _______________
- What is your major? _______________
- What is your date of birth? _______________
- Are you Hispanic or Latino? (i.e., a person of Cuban, Mexican, Puerto Rican, South or Central American descent, or other Spanish culture or origin, regardless of race)? *(Yes, No)*
- Please indicate your race (select all that apply). *(White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other – please specify)*
- What is the highest level of education either of your parents have achieved? *(Some high school; High school; Associate's degree; Bachelor's degree; Master's degree; Med school, law school, or PhD)*

Attention & Suspicion Check

The next couple of questions refer to your experience and level of engagement during the time you worked on this questionnaire. **<u>Please respond honestly</u>.** Your responses will in no way affect your credit for participating today.

1. How distracted were you while filling out this questionnaire? *(Not at all distracted, A little distracted, Somewhat distracted, Very distracted, Extremely distracted)*
2. How seriously did you fill out this questionnaire? *(Not at all seriously, A little seriously, Somewhat seriously, Very seriously, Extremely seriously)*
3. Did you find anything in this survey to be strange or difficult to answer? *[open response]*

Debrief
Thank you for your participation in this study. Your participation is sincerely appreciated, and we hope that you have found your experience to be interesting.

As we told you initially, the purpose of this study was to develop a better understanding of students' confidence and motivation in particular school subjects and how they may relate to class participation in those subjects. We would now like to share some more specific information with you about the hypothesis we were investigating. We are interested in how society's expectations for men versus women in different subjects may influence students' confidence, motivation, and/or class participation in these subjects.

Your participation in this project will help our efforts in understanding how gender expectations may impact students' class participation behaviors. In the long term, this research might have implications for future studies of how student characteristics can impact engagement and confidence in school.

So, that's a basic description of what the study is about. It is very important for other participants to come in without knowing what we are studying. For this reason, **please do not talk about this study with other students who may participate**. Prior expectations may influence the findings unintentionally and thus make our efforts (and yours) potentially less useful and informative.

If you have any additional questions, comments, or concerns, please feel free to email Meghan Coughlan, the principal investigator, at coughlmp@bc.edu. Thank you again for your participation. We truly appreciate it!

FOLLOW-UP SURVEY
Note: Data in this survey was collected in order to address research questions that are unrelated to the present dissertation.

Instructions
For the next set of questions, please think about your experiences in your **MATH [SOCIAL SCIENCE] class, [name of course]**.

Self-Reported Class Participation Behaviors
Response scale: Never, Very rarely, Rarely, Occasionally, Often, Very often

1. How often do you raise your hand to answer a question posed by the instructor in your math [social science] class?
2. When no question has been asked, how often do you raise your hand to make a comment, share your experience, or provide an explanation in your math [social science] class?
3. On occasions when you decide that you want to raise your hand to answer a question posed by the instructor in your math [social science] class, how often do you change your mind?

4. On occasions when no question has been asked, and you decide that you want to raise your hand to make a comment, share your experience, or provide an explanation in your math [social science] class, how often do you change your mind?

Mathematics [Social Science] Career Interest
Response scale: Not at all true (1) to Very true (6)

1. I am interested in taking courses that help me learn more about mathematics [social science].
2. I am interested in working in a career that allows me to use mathematics-related [social science-related] skills or knowledge.
3. I would like to learn mathematics-related [social science-related] knowledge and skills because they can be useful to help me be prepared for a career.

Cognitive Strategies

1. Please take two or three minutes to describe how you study for your **math [social science]** course. *[open response]*
2. For your **math [social science]** course, how often do you use the following study activities? (*Response scale: Never, Rarely, Sometimes, Often, Always)*
    - Space out your study sessions over multiple days
    - Test yourself with questions or practice problems
    - Use flashcards
    - Reread chapters, articles, notes, etc.
    - Make outlines
    - Underline or highlight while reading
    - Make diagrams, charts, or pictures
    - Study with friends
    - “Cram” a lot of information the night before a test
    - Ask questions or verbally participate during class
    - Watch/listen to recordings from the instructor or from an outside source (Kahn Academy, YouTube, etc.)
    - Implement feedback from prior assignments (including exams and papers)
    - Study concepts from multiple topics/units during the same study session
    - Rewrite your notes
    - Explain the material or problem solution to yourself
    - Compare or contrast information
    - Keep track of how well you understand the material
    - Identify which concepts or information you do not understand
    - Take verbatim notes from the book, lecture notes, etc.
    - Test yourself before you begin studying to see what you already know

Appendix D
Study 2 Measures

PHASE 0: ENROLLMENT FORM

Thank you for your interest in participating in this study! Please enter your information below. If you are selected to participate, you will receive an email containing a link to the study survey in the next couple of weeks.

- First Name: ____________
- Last Name: ____________
- University email address: ____________
- Student ID number: ____________
- Are you 18 years or older? *Yes, No*
- Gender: ____________
- Major(s): ____________
- Class year: *Freshman, Sophomore, Junior, Senior, Other (please explain)*
- Please list your current courses below, entering the course titles as precisely as possible (e.g., "Finite Probability and Applications").
  Course #1: ____________
  Course #2: ____________
  Course #3: ____________
  Course #4: ____________
  Course #5: ____________
  Course #6: ____________
  Course #7: ____________
  Course #8: ____________

This study will involve completing some short surveys throughout the semester, so we would like to know a bit about your availability.

Are there weeks (or parts of weeks) that you anticipate being off campus, unavailable to participate in the study, or not attending classes this semester? Check all that apply. *[List of weeks in semester provided here as a checklist.] Final option in list:* I do not plan to be off campus, out of classes, or unavailable to participate in the study during any of the above weeks.

PHASE 1: INITIAL SURVEY

Selection of Courses
In this survey, you will be asked to reflect on your experiences in a math class or a social science class in which you are currently enrolled.

Based on the survey you completed, we have determined which of your courses qualify as math and/or social science courses.

[If I determined that participants were enrolled in only one eligible math course and one eligible social science course]:

- **Math course:** [Name of math course selected by me]
- **Social science course:** [Name of social science course selected by me]

As you are completing this survey, please keep your experiences within these two courses in mind.

[If participants were currently enrolled in more than one math course and/or more than one eligible social science course, I invited them to select which course(s) they would like to reference throughout the study]:

Before completing the following survey, please identify which **MATH [SOCIAL SCIENCE]** course you would like to refer to throughout the survey. When selecting a course, <u>**please consider which courses offer the most opportunities for classroom participation via hand-raising**</u>.

Please select which math course you would like to refer to: *[options presented as multiple choice]*
Please select which social science course you would like to refer to: *[options presented as multiple choice]*

As you are completing this survey, please keep your experiences within these two courses in mind.

[Next page]: For the next set of questions, please think about your experiences in your <u>**MATH**</u> **[<u>SOCIAL SCIENCE</u>] class, [name of course].**

Stereotype Threat Susceptibility – Scenario-Based Measures (ST-1)
<u>Response scale</u>: Strongly disagree (1) to Strongly agree (6)

[Male scenario:]
Imagine that you are sitting in your **math [social science]** class, **[name of course]**, and are about to take a midterm. You are surrounded mostly by <u>**male**</u> peers and have a <u>**male**</u> professor. Please rate the extent to which you agree or disagree with the statements below.

[Female scenario:]
Imagine that you are sitting in your **math [social science]** class, **[name of course]**, and are about to take a midterm. You are surrounded mostly by <u>**female**</u> peers and have a <u>**female**</u> professor. Please rate the extent to which you agree or disagree with the statements below.

1. I would worry about my ability to perform well on the exam.
2. I would worry about what my professor's perception of my math [social science] ability would be if I performed poorly on the exam.

3. I would worry about what my peers' evaluations of my math [social science] ability would be if I performed poorly on the exam.

Stereotype Threat in Math [Social Science] (ST-2)
<u>Response scale</u>: *Strongly disagree (1) to Strongly agree (6)*

1. I feel pressure to do what I can to improve the image of men [women] in math [social science].
2. I am afraid that if I do poorly in this class, it will confirm the stereotype that men [women] cannot be good in math [social science].
3. I fear that men [women] in this class will judge me as being incompetent in math [social science] if I do not do well on the assignments and exams.
4. I feel pressure to do what I can to change the negative stereotype about men [women] being generally weak in math [social science].
5. I am afraid of being negatively evaluated by men [women] in this class.
6. I am afraid of confirming the stereotype that men [women] do not have the skills to be successful in math- [social science-] related careers.
7. I feel pressure to represent men [women] in math [social science] because there are so few of us in the field.

Domain-Specific Regulatory Focus
<u>Response scale</u>: *Strongly disagree (1) to Strongly agree (6)*

1. I worry about making mistakes in my math [social science] work when I am taking a test or doing homework. *[prevention]*
2. I frequently think about how I can prevent failures in my math [social science] classes. *[prevention]*
3. I frequently think about making progress toward being successful in math [social science]. *[promotion]*
4. I frequently imagine how I will achieve my math- [social science-] related hopes and aspirations. *[promotion]*
5. I have important academic standards in my math [social science] class that I focus on maintaining. *[prevention]*
6. When I see an opportunity for achieving a math [social science] goal, I get excited right away. *[promotion]*

Class Participation Confidence Threshold – Slider Measure (CPCT-1)
In your **math [social science]** class, **[name of course]**, there are a number of ways in which you might participate. For instance, you might raise your hand to answer a question posed by the instructor, make a comment, share your experience, or provide an explanation.

Some students will choose to participate even if they are ***not at all sure*** that their contribution is valid. Other students will only raise their hand if they are ***completely sure*** that their contribution is valid.

What about you? If you had an opportunity to contribute in your **math [social science]** class, how sure of your contribution would you need to be in order to raise your hand? For example:

- **<u>0% sure</u>:** I would choose to participate even if I was **<u>not at all sure</u>** that my contribution was valid.
- **<u>25% sure</u>:** I would choose to participate as long as I was **<u>a little bit sure</u>** that my contribution was valid.
- **<u>50% sure</u>:** I would choose to participate as long as I was **<u>half sure</u>** that my contribution was valid.
- **<u>75% sure</u>:** I would choose to participate as long as I was **<u>mostly sure</u>** that my contribution was valid.
- **<u>100% sure</u>:** I would choose to participate only if I was **<u>completely sure</u>** that my contribution was valid.

Please mark your response on the slider: *[Display slider with five anchors above marked; slider will notify participant of selected percentage from 0-100]*

Class Participation Confidence Threshold – Likert-Type Measure (CPCT-2)
<u>Response scale</u>: Strongly disagree (1) to Strongly agree (6)

In your **math [social science]** class, **[name of course]**, there are a number of ways in which you might participate. For instance, you might raise your hand to answer a question posed by the instructor, make a comment, share your experience, or provide an explanation. Please rate the extent to which you agree or disagree with the statements below.

1. I generally do NOT participate unless I feel completely sure that my contribution is valid.
2. I tend to participate even if I am not at all sure that my contribution is valid.
3. I usually participate only when I am absolutely certain that my contribution is valid.

Self-Reported Class Participation Behaviors
<u>Response scale</u>: Never, Very rarely, Rarely, Occasionally, Often, Very often

1. How often do you raise your hand to <u>answer a question posed by the instructor</u> in your math [social science] class?
2. When no question has been asked, how often do you raise your hand to <u>make a comment, share your experience, or provide an explanation</u> in your math [social science] class?
3. On occasions when you decide that you want to raise your hand to <u>answer a question posed by the instructor</u> in your math [social science] class, how often do you change your mind?
4. On occasions when no question has been asked, and you decide that you want to raise your hand to <u>make a comment, share your experience, or provide an</u>

explanation in your math [social science] class, how often do you change your mind?

Opportunities for Participation
Response scale: Never, Very rarely, Rarely, Occasionally, Often, Very often

1. How often does the instructor of your math [social science] course pose questions to the class?
2. How often do you think students in your math [social science] course have an opportunity to participate?

Self-Confidence in Math [Social Science]
Response scale: False, Mostly false, More false than true, More true than false, Mostly true, True

1. Compared to others, I am good at math [social science].
2. Work in math [social science] classes is easy for me.
3. I learn things quickly in math [social science].
4. I get good grades in math [social science].
5. I'm hopeless when it comes to math [social science].
6. I have always done well in math [social science].

Participation-Specific Confidence in Math [Social Science]
Response scale: Strongly disagree (1) to Strongly agree (6)

Imagine that you are sitting in **math [social science]** class, **[name of course]**, and you have an opportunity to contribute (e.g., to answer a question posed by the instructor, make a comment, share your experience, or provide an explanation). Please rate the extent to which you agree or disagree with the statements below.

1. I generally do NOT feel certain that my contribution is valid.
2. I tend to feel very certain that my contribution is valid.
3. I am usually absolutely certain that my contribution is valid.

Perceived Value of Math [Social Science]

1. How useful is math [social science] – how much can it help you do better outside of class? *(Not at all useful, Not that useful, A little bit useful, Useful, Very useful)*
2. How important is it to know math [social science] in order to get a good job? *(Not at all important, Not that important, A little bit important, Important, Very important)*
3. How important is math [social science] compared to other school subjects? *(The least important subject, Less important than most other subjects, Equally important to most other subjects, More important than most other subjects, The most important subject)*
4. How important is it to be good at math [social science]? *(Not at all important, Not that important, A little bit important, Important, Very important)*

5. How interesting is math [social science] for you? *(Not at all interesting, Not that interesting, A little bit interesting, Interesting, Very interesting)*
6. How much do you enjoy doing math [social science]? *(I do not enjoy it at all, I do not enjoy it that much, I enjoy it a little bit, I enjoy it, I enjoy it a lot)*

Costs and Value of Participating in Class
Response scale: Strongly disagree (1) to Strongly agree (6)

1. Participating in math [social science] class is important to me.
2. I find participating in math [social science] class to be enjoyable.
3. I find participating in math [social science] class to be useful.
4. Participating in math [social science] class makes me feel anxious.
5. In math [social science] class, I'm concerned that if I answer the professor's question incorrectly, my professor will think I am dumb.
6. In math [social science] class, I'm concerned that if I answer the professor's question incorrectly, other students in my class will think I am dumb.
7. Raising my hand in math [social science] class requires too much effort.
8. In math [social science] class, I am able to pay attention to the lesson more when I am not participating.

Math [Social Science] Sense of Belonging
Response scale: Strongly disagree (1) to Strongly agree (6)

When I am in a **math [social science]** setting…
1. …I feel that I belong to the math [social science] community.
2. …I consider myself a member of the math [social science] world.
3. …I feel like I am part of the math [social science] community.
4. …I feel a connection with the math [social science] community.
5. …I feel like an outsider.
6. …I feel accepted.
7. …I feel respected.
8. …I feel disregarded.
9. …I feel valued.
10. …I feel neglected.
11. …I feel appreciated.
12. …I feel excluded.
13. …I feel like I fit in.
14. …I feel insignificant.

Math [Social Science] Identity
Response scale: Strongly disagree (1) to Strongly agree (6)

1. Being good at math [social science] is *not* an important part of who I am.
2. I consider myself to be a math [social science] person.
3. Math [social science] is an important part of my identity.

Mathematics [Social Science] Career Interest
Response scale: Not at all true (1) to Very true (6)

1. I am interested in taking courses that help me learn more about mathematics [social science].
2. I am interested in working in a career that allows me to use mathematics-related [social science-related] skills or knowledge.
3. I would like to learn mathematics-related [social science-related] knowledge and skills because they can be useful to help me be prepared for a career.

Scheduling Questions
Please indicate (by checking the appropriate boxes) whether you will need to engage in any of the following academic tasks during each of the weeks listed below. Please check your syllabi to confirm.

- In-class midterm exam
- In-class final exam
- In-class presentation

In my **math [social science]** class, **[name of course**], I have one or more of the above in-class tasks on the following weeks: *[List of weeks in semester provided here as a checklist.] Final option in list:* I do not have any of the above in-class tasks during the weeks listed.

Cognitive Strategies
Note: This measure was included in order to address research questions that are unrelated to the present dissertation.

1. Please take two or three minutes to describe how you study for your **math [social science]** course, **[name of course]**. *[open response]*
2. For your **math [social science]** course, **[name of course],** how often do you use the following study activities? (*Response scale: Never, Rarely, Sometimes, Often, Always)*
    - Space out your study sessions over multiple days
    - Test yourself with questions or practice problems
    - Use flashcards
    - Reread chapters, articles, notes, etc.
    - Make outlines
    - Underline or highlight while reading
    - Make diagrams, charts, or pictures
    - Study with friends
    - “Cram” a lot of information the night before a test
    - Ask questions or verbally participate during class
    - Watch/listen to recordings from the instructor or from an outside source (Kahn Academy, YouTube, etc.)
    - Implement feedback from prior assignments (including exams and papers)
    - Study concepts from multiple topics/units during the same study session

- Rewrite your notes
- Explain the material or problem solution to yourself
- Compare or contrast information
- Keep track of how well you understand the material
- Identify which concepts or information you do not understand
- Take verbatim notes from the book, lecture notes, etc.
- Test yourself before you begin studying to see what you already know

Demographic Questions

- What grades do you typically receive in **math** classes? *(Mostly As, Mostly Bs, Mostly Cs, Mostly Ds, Mostly Fs)*
- What grades do you typically receive in **social science** classes? *(Mostly As, Mostly Bs, Mostly Cs, Mostly Ds, Mostly Fs)*
- What grades do you typically receive in general? *(Mostly As, Mostly Bs, Mostly Cs, Mostly Ds, Mostly Fs)*
- What is your gender? _______________
- What is your date of birth? _______________
- Are you Hispanic or Latino? (i.e., a person of Cuban, Mexican, Puerto Rican, South or Central American descent, or other Spanish culture or origin, regardless of race)? *(Yes, No)*
- Please indicate your race (select all that apply). *(White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other – please specify)*
- What is the highest level of education either of your parents have achieved? *(Some high school; High school; Associate's degree; Bachelor's degree; Master's degree; Med school, law school, or PhD)*

Attention & Suspicion Check

The next couple of questions refer to your experience and level of engagement during the time you worked on this questionnaire. **Please respond honestly.** Your responses will in no way affect your credit for participating today.

1. How distracted were you while filling out this questionnaire? *(Not at all distracted, A little distracted, Somewhat distracted, Very distracted, Extremely distracted)*
2. How seriously did you fill out this questionnaire? *(Not at all seriously, A little seriously, Somewhat seriously, Very seriously, Extremely seriously)*
3. Did you find anything in this survey to be strange or difficult to answer? *[open response]*

PHASE 2: DAILY DIARIES

Attendance Questions

1. Did you attend **[name of math course]** today?
    - Yes, I attended in person.

- Yes, I attended online <u>synchronously</u> (i.e., live, at the same time as the instructor and other students).
- Yes, I attended online <u>asynchronously</u> (i.e., not live, in a self-paced manner).
- No, I did not attend this class today.

2. Did you attend **[name of social science course]** today?
   - Yes, I attended in person.
   - Yes, I attended online <u>synchronously</u> (i.e., live, at the same time as the instructor and other students).
   - Yes, I attended online <u>asynchronously</u> (i.e., not live, in a self-paced manner).
   - No, I did not attend this class today.

[The following questions were only displayed for the course(s) that participants attended in person OR online synchronously on the given day. All participants were presented with the cognitive strategies items at the end of the survey.]

Daily Version – Stereotype Threat
<u>Response scale</u>: Strongly disagree (1) to Strongly agree (6)

Immediately before or during **[name of math/social science course]** today…
1. I felt pressure to improve the image of [males/females] in math [social science].
2. I felt afraid that if I do poorly in this class, it will confirm the stereotype that [males/females] cannot be good in math [social science].
3. I felt afraid that [males/females] in this class view me as incompetent in math [social science].
4. I felt afraid of confirming the stereotype that [males/females] do not have the skills to be successful in math- [social science-] related careers.

Daily Version – Regulatory Focus
<u>Response scale</u>: Strongly disagree (1) to Strongly agree (6)

Immediately before or during **[name of math/social science course]** today, I was focused on…
1. …not making mistakes. *[prevention]*
2. …minimizing failure in this class. *[prevention]*
3. …making progress toward being successful in this class. *[promotion]*
4. …achieving my hopes and aspirations for this class. *[promotion]*
5. …maintaining the standards that I set for myself in this class. *[prevention]*
6. …opportunities to achieve my goal for this class. *[promotion]*

Daily Version – Class Participation Confidence Threshold
<u>Response scale</u>: Strongly disagree (1) to Strongly agree (6)

In **[name of math/social science course]**, there are a number of ways in which you might participate. For instance, you might raise your hand to answer a question posed by the instructor, make a comment, share your experience, or provide an explanation.

Please rate the extent to which you agree or disagree with the statements below.

1. Throughout class today, when I was deciding whether to participate, I felt that I needed to be completely sure that each contribution was valid.
2. I was willing to participate in class today even when I was not at all sure that my potential contribution was valid.
3. I was only willing to participate in class today when I was absolutely certain that my potential contribution was valid.

Daily Version – Self-Reported Class Participation Behaviors
Response format: Open response boxes

Below, please enter numbers only. If you don't know the exact number, give your best estimate.

Today in **[name of math/social science course]**, how many times did you…

1. …raise your hand to answer a question posed by the instructor?
2. …raise your hand to make a comment, share your experience, or provide an explanation when no question had been asked?
3. …change your mind after originally deciding that you wanted to answer a question posed by the instructor?
4. …change your mind after originally deciding that you wanted to make a comment, share your experience, or provide an explanation?

Opportunities for Participation
Response format: Open response box

1. Approximately how many times did your **[name of math/social science course]** instructor pose questions to the class today? Please enter a number.

Daily Version – Sense of Belonging
Response scale: Strongly disagree (1) to Strongly agree (6)

During **[name of math/social science course]** today…

1. …I felt that I belonged.
2. …I felt like a member of the math [social science] world.
3. …I felt like an outsider.
4. …I felt accepted.
5. …I felt excluded.
6. …I felt like I fit in.

Cognitive Strategies
Note: This measure was included in order to address research questions that are unrelated to the present dissertation.

1. Did you study for **[name of math/social science course]** today? (*Yes, No)*

[The item below was displayed only for the course(s) for which participants selected "yes" above, indicating that they had studied for the course that day.]

2. Please check all the ways in which you studied for **[name of math/social science course]**:
   - ☐ Tested yourself with questions or practice problems
   - ☐ Used flashcards
   - ☐ Reread chapters, articles, notes, etc.
   - ☐ Made outlines
   - ☐ Underlined or highlighted while reading
   - ☐ Made diagrams, charts, or pictures
   - ☐ Studied with friends
   - ☐ "Crammed" for a test or quiz
   - ☐ Watched/listened to recordings from the instructor or from an outside source (Kahn Academy, YouTube, etc.)
   - ☐ Implemented feedback from prior assignments (including exams and papers)
   - ☐ Studied concepts from multiple topics/units during the same study session
   - ☐ Rewrote your notes
   - ☐ Explained the material or problem solution to yourself
   - ☐ Compared or contrasted information
   - ☐ Kept track of how well you understood the material
   - ☐ Identified which concepts or information you did not understand
   - ☐ Took verbatim notes from the book, lecture notes, etc.
   - ☐ Tested yourself before you began studying to see what you already knew
   - ☐ Other (please explain)

<u>PHASE 3: FINAL SURVEY</u>

Course Schedule

1. On which days of the week do you have **[name of math course]**?
   - Mondays, Wednesdays, and Fridays
   - Tuesdays and Thursdays
   - Other (please explain) __________

2. On which days of the week do you have **[name of social science course]**?
   - Mondays, Wednesdays, and Fridays
   - Tuesdays and Thursdays
   - Other (please explain) __________

Gender Composition of Courses

<u>Response scale</u>: Mostly male, About evenly split, Mostly female

1. How would you describe the gender composition of your <u>math</u> class, **[name of math course]**?
2. How would you describe the gender composition of your <u>social science</u> class, **[name of social science course]**?

Course Reminder
Math course: [Name of math course]
Social science course: [Name of social science course]

As you are completing this survey, please consider your experiences within these two courses.

For the next set of questions, please think about your experiences in your **MATH [SOCIAL SCIENCE] class, [name of course]**.

Math [Social Science] Sense of Belonging Scale
Response scale: Strongly disagree (1) to Strongly agree (6)

When I am in a **math [social science]** setting…

1. …I feel that I belong to the math [social science] community.
2. …I consider myself a member of the math [social science] world.
3. …I feel like I am part of the math [social science] community.
4. …I feel a connection with the math [social science] community.
5. …I feel like an outsider.
6. …I feel accepted.
7. …I feel respected.
8. …I feel disregarded.
9. …I feel valued.
10. …I feel neglected.
11. …I feel appreciated.
12. …I feel excluded.
13. …I feel like I fit in.
14. …I feel insignificant.

Math [Social Science] Identity
Response scale: Strongly disagree (1) to Strongly agree (6)

1. I consider myself to be a math [social science] person.
2. Math [social science] is an important part of my identity.
3. Being good at math [social science] is *not* an important part of who I am.

Mathematics [Social Science] Career Interest
Response scale: Not at all true (1) to Very true (6)

1. I am interested in taking courses that help me learn more about mathematics [social science].
2. I am interested in working in a career that allows me to use mathematics-related [social science-related] skills or knowledge.
3. I would like to learn mathematics-related [social science-related] knowledge and skills because they can be useful to help me be prepared for a career.

Cognitive Strategies

Note: This measure was included in order to address research questions that are unrelated to the present dissertation.

1. For your math [social science] course, **[name of course]**, how often have you "crammed" a lot of information the night before a test or quiz? *(Never, Rarely, Sometimes, Often, Always)*

[If participant did NOT select "Never" in #1, the following three questions were displayed.]

2. In general, why have you crammed the night before a test or quiz in your math [social science] course? *[Open response]*
3. In general, when you have crammed the night before a test or quiz in your math [social science] course, to what extent was it because you procrastinated? *(Not at all, To a small extent, To some extent, To a large extent, To a great extent)*
4. In general, when you have crammed the night before a test or quiz in your math [social science] course, to what extent was it because you wanted to make sure you studied everything that would be covered on the test? *(Not at all, To a small extent, To some extent, To a large extent, To a great extent)*

Debrief

Thank you for your participation in this study. Your participation is sincerely appreciated, and we hope that you have found your experience to be interesting.

As we told you initially, the purpose of this study was to develop a better understanding of students' confidence and motivation in particular school subjects and how they may relate to class participation in those subjects. We would now like to share some more specific information with you about the hypothesis we were investigating. We are interested in how society's expectations for men versus women in different subjects may influence students' confidence, motivation, and/or class participation in these subjects.

Your participation in this project will help our efforts in understanding how gender expectations may impact students' class participation behaviors. In the long term, this research might have implications for future studies of how student characteristics can impact engagement and confidence in school.

So, that's a basic description of what the study is about. It is very important for other participants to come in without knowing what we are studying. For this reason, **please do not talk about this study with other students who may participate**. Prior expectations may influence the findings unintentionally and thus make our efforts (and yours) potentially less useful and informative.

If you have any additional questions, comments, or concerns, please feel free to email Meghan Coughlan, the principal investigator, at coughlmp@bc.edu. Thank you again for your participation. We truly appreciate it!

Appendix E
Correlations Within Each Phase (Study 2)

Table E1

Correlations Between Study 2/Phase 1 Variables (Math)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Primary Constructs															
1. ST-1 (scenario)	1														
2. ST-2 (Likert)	$.374^{***}$	1													
3. Regulatory focus	.093	.010	1												
4. CPCT-1 (slider)	$.129^{*}$	$.231^{***}$	$.184^{**}$	1											
5. CPCT-2 (Likert)	$.197^{**}$	$.339^{***}$	$.220^{***}$	$.704^{***}$	1										
6. Participation frequency	$-.109^{+}$	$-.252^{***}$	$-.187^{**}$	$-.461^{***}$	$-.632^{***}$	1									
7. Freq. of changing mind	$.155^{*}$	$.296^{***}$	.074	$.284^{***}$	$.350^{***}$	$-.283^{***}$	1								
8. Sense of belonging	$-.101^{+}$	$-.329^{***}$	$-.374^{***}$	$-.184^{**}$	$-.309^{***}$	$.343^{***}$	$-.318^{***}$	1							
9. Domain identity	.055	.046	$-.502^{***}$	-.081	$-.125^{*}$	$.108^{+}$	-.068	$.480^{***}$	1						
10. Career interest	.017	.068	$-.473^{***}$	.015	-.049	$.119^{+}$	.004	$.376^{***}$	$.683^{***}$	1					
Covariates															
11. Participation opportunities	-.069	.035	-.015	-.072	$-.156^{*}$	$.277^{***}$	.028	.041	-.076	$-.047^{+}$	1				
12. Domain confidence	.098	-.028	$-.225^{***}$	$-.143^{*}$	$-.132^{*}$	$.187^{**}$	$-.113^{+}$	$.562^{***}$	$.604^{***}$	$.492^{***}$	-.006	1			
13. Participation confidence	$-.111^{+}$	$-.243^{***}$	-.093	$-.146^{*}$	$-.246^{***}$	$.381^{***}$	$-.369^{***}$	$.433^{***}$	.088	$.117^{+}$	.042	$.260^{***}$	1		
14. Conf. diff. score	$-.197^{**}$	$-.371^{***}$	$-.201^{**}$	$-.552^{***}$	$-.808^{***}$	$.648^{***}$	$-.455^{***}$	$.467^{***}$	$.136^{*}$	$.103^{+}$	$.128^{*}$	$.245^{***}$	$.770^{***}$	1	
15. Domain value	.018	.039	$-.492^{***}$	-.057	-.100	$.153^{*}$	-.010	$.416^{***}$	$.717^{***}$	$.763^{***}$	-.016	$.526^{***}$	$.135^{*}$	$.148^{*}$	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table E2

Correlations Between Study 2/Phase 1 Variables (Social Science)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Primary Constructs															
1. ST-1 (scenario)	1														
2. ST-2 (Likert)	.166**	1													
3. Regulatory focus	.055	-.063	1												
4. CPCT-1 (slider)	-.050	.175**	.044	1											
5. CPCT-2 (Likert)	.042	.239***	.209**	.634***	1										
6. Participation frequency	-.078	-.161**	-.194**	-.325***	-.604***	1									
7. Freq. of changing mind	.062	.285***	.054	.226***	.295***	-.266***	1								
8. Sense of belonging	-.099	-.168**	-.327***	-.057	-.262***	.389***	-.242***	1							
9. Domain identity	-.103^{+}	.159**	-.509***	.008	-.141*	.322***	-.014	.577***	1						
10. Career interest	-.013	.188**	-.502***	-.039	-.195**	.364***	.023	.532***	.742***	1					
Covariates															
11. Participation opportunities	.004	-.067	.171**	-.204**	-.190**	.225***	-.037	.031	-.121*	-.104	1				
12. Domain confidence	-.015	-.066	-.291***	-.022	-.198**	.331***	-.123*	.563***	.469***	.413***	.034	1			
13. Participation confidence	-.067	-.173**	-.064	-.133*	-.292***	.473***	-.313***	.502***	.240***	.231***	.124*	.411***	1		
14. Conf. diff. score	-.066	-.258***	-.176**	-.497***	-.832***	.674***	-.377***	.465***	.233***	.264***	.198**	.370***	.774***	1	
15. Domain value	-.009	.203**	-.425***	-.073	-.164**	.370***	-.026	.561***	.690***	.817***	-.066	.386***	.287***	.275***	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table E3

Correlations Between Study 2/Phase 2 Variables (Math)

	1	*2*	*3*	*4*	*5*
1. ST-2 (Likert)	1				
2. Regulatory focus	.127	1			
3. CPCT-2 (Likert)	.349***	.260**	1		
4. Percent freq. of participation	-.099	-.109	-.203**	1	
5. Career interest	-.459***	-.232**	-.336***	.190**	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table E4

Correlations Between Study 2/Phase 2 Variables (Social Science)

	1	*2*	*3*	*4*	*5*
1. ST-2 (Likert)	1				
2. Regulatory focus	.162*	1			
3. CPCT-2 (Likert)	.246***	.259***	1		
4. Percent freq. of participation	-.125+	-.095	-.296***	1	
5. Career interest	-.268***	-.309***	-.266***	.177**	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table E5

Correlations Between Study 2/Phase 3 Variables (Math)

	1	*2*	*3*
1. Sense of belonging	1		
2. Domain identity	$.390^{***}$	1	
3. Career interest	$.377^{***}$	$.689^{***}$	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table E6

Correlations Between Study 2/Phase 3 Variables (Social Science)

	1	*2*	*3*
1. Sense of belonging	1		
2. Domain identity	$.461^{***}$	1	
3. Career interest	$.423^{***}$	$.685^{***}$	1

Note. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$