DISSERTATION

THE ROLE OF WOMEN'S IDENTIFICATION WITH MATH AND ACADEMIC MAJOR IN WOMEN'S SUSCEPTIBILITY TO STEREOTYPE THREAT AND STEREOTYPE LIFT

Submitted by

Devi Deviyanti

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2015

Doctoral Committee:

Advisor: Silvia Sara Canetto
Co-Advisor: Paul Hernandez

Nathalie Kees
Lucy Troup

ABSTRACT

THE ROLE OF WOMEN'S IDENTIFICATION WITH MATH AND ACADEMIC MAJOR IN WOMEN'S SUSCEPTIBILITY TO STEREOTYPE THREAT AND STEREOTYPE LIFT

A stereotype threat (ST) occurs when individuals underperform in a domain, for example in math, as a result of exposure to a relevant negative stereotype. Women engaged in math-intensive tasks can experience ST when negative stereotypes about women's math ability are made salient, via for example, test instructions that allege superior math performance by men. Evidence regarding the role of ST test instructions on women's math performances has been mixed (e.g., Bell, Spencer, Iserman, & Logel, 2003; Schmader, 2002). While prior studies found that women underperform in ST conditions that emphasize the validity of a math test (i.e., when a math test is presented as indicative of math ability), no study has included a condition in which the validity of a math test is downplayed (i.e., "test not indicative of math ability" conditions). Studies examining conditions alleging men's superiority in math (i.e., "men perform better than women" conditions) have not included conditions that presented a math test as indicative of math ability (Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Johnson, Bernard-Brak, Saxon, & Johnson, 2012). Additionally, it is unclear which women are most vulnerable to math ST conditions. While ST is found to have greater impact on women who are highly identified with math relative to women with low identification with math (e.g., Aronson, Quinn, & Spencer, 1999), there is also evidence that women in math-intensive majors (e.g., engineering) have lower susceptibility to math ST than women not in math-intensive majors (e.g., psychology) (Crisp, Bache, & Maitner, 2009; Croizet et al., 2004). Furthermore, the roles of identification with math and academic major have been researched independently. The present study examines

the roles of women's identification with mathematics and college majors on their susceptibility to math underperformance under two ST conditions, one related to the validity of the math test and the other involving comparisons in math performance between women and men.

Women (n = 847), of whom 231 were in math-intensive majors and 616 were not in math-intensive majors at a large Mountain West state university, completed the Identification with Math Scale and reported their college majors five to seven days before completing a mathematics test. They were then randomly assigned to one of six math ST conditions in a 2 (Validity of Math Test Variable: test indicative of math ability, test not indicative of math ability) × 3 (Women-Men Math Performance Differences Variable: men perform better than women, no mention of differences in math performance, or women perform better than men) factorial design experiment.

It was hypothesized that women in the "men perform better than women" condition would underperform relative to women in the "no mention of differences in math performance" condition. It was also hypothesized that women high in identification with math who were assigned to the "test indicative of math ability" condition would experience greater math underperformance than women in the "test not indicative of math ability" condition. A significant interaction between the Women-Men Math Performance Differences Variable and the Identification with Math Variable was found. Women high in identification with math in the "men perform better than women" condition scored significantly lower than women in the "no mention of differences in math performance" condition. No such difference in performance was observed for women low in identification with math. Women in the "women perform better than men" condition performed better than women in other conditions regardless of their identification with math.

This study’s findings suggest that women who strongly identify with math may be especially vulnerable to ST, consistent with past findings (e.g., Steinberg, Okun, & Aiken, 2012). In support of findings from past studies (e.g., Johnson et al., 2012), this study also demonstrates that to do well in math tests women may benefit from exposure to information explicitly contradicting female math incompetence stereotypes. Current study’s findings have implications for intervention programs with highly math-identified women.

ACKNOWLEDGMENT

I would like to express gratitude to the members of my dissertation committee: Dr. Silvia Sara Canetto (Chair), Dr. Paul R. Hernandez (Co-chair), Dr. Nathalie Kees, and Dr. Lucy Troup for their guidance. I am also eternally grateful to my family, James Ney McDonald and Kristina Ochsner Armstrong for their unfailing support, encouragement, and insight.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Stereotypes are “beliefs about the characteristics, attributes, and behaviors of members of certain groups” (Hilton & von Hippel, 1996, p. 240). “The existence of such a stereotype means that anything one does or any of one’s features that conform to it make the stereotype more plausible as a self-characteristics in the eyes of others and perhaps in one’s own eyes” (Steele & Aronson, 1995, p. 797) and in some situation, stereotypes can influence individuals to make choices that are consistent with the stereotype (Renninger, 2000). The prevalence and permeating power of the stereotypes that math-intensive field is “a man’s domain” and that “women don’t do math” can hinder performance on critical tests, and limit the academic choices women make.

Imagine identical twins, a girl and a boy, sitting down to complete an important math test. Despite identical familial and educational backgrounds, and even assuming equal motivation to do well, the girl must contend with additional pressure resulting from a general expectation that she will do poorly due to her sex. She must contend with the knowledge that her performance could confirm that expectation. In short, she takes the test knowing that if she fails, she may validate the stereotype with her performance on the test. She may endure the anxiety that comes with this test, but such strain may undermine her performance on the test and come to influence her academic and professional decisions throughout her life. The experience of this individual is common, and, for many young girls, it is worse. A number of factors have been proposed to explain this experience. These include sex differences in spatial ability (Baenninger & Newcombe, 1995; Levine, Huttenlocher, Taylor, & Langrock, 1999; Terlecki, Newcombe & Little, 2007), brain development, and hormonal sex differences (Ardila, Rosselli, Matute, & Inozemtseva, 2011; Halpern, 1992; Wilder & Powell 1989). Nevertheless, what is known as the “different availability of aptitude” explanation does not explain why boys and girls only start to emerge at a certain developmental period, namely middle school (Ambady, Shih, Kim, & Pittinsky, 2001; Huguet & Regner, 2007, 2009; Muzzatti & Agnoli, 2007, Experiment 2) and in certain contexts, such as in high-stakes evaluative tests (e.g., Johnson et al., 2012; Kiefer & Sekaquaptewa, 2007; Steinberg et al., 2012; Thoman,White, Yamawaki, & Koishi, 2008). These findings suggest that a possible explanation for these discrepancies is the influence of social-cultural factors, such as the phenomena known as stereotype threat or ST (Ceci, Williams, & Barnett, 2009).

The purpose of this study was to investigate the situation-specific effects of ST on women’s math performances. ST has been observed to emerge in high-stakes math testing situations (e.g., Aronson et al., 1999) and research has shown that a specific feature of the situation, specifically the manner in which a math test is described via instruction prior to test-taking, can induce ST and contribute to math underperformance (or math ST; e.g., Johnson et al., 2012). However, not all studies found support for the negative effects of math ST (e.g., McFarland, Kemp, Viera, & Odin, 2003; Oswald & Harvey, 2000–2001; Schneeberger & Williams, 2003). These mixed findings suggest that the extent to which math ST influences women’s math performance may vary depending on individual difference factors (Nguyen & Ryan, 2008). Research has investigated the role of two important individual difference factors that influence women’s susceptibility to math ST: academic major and identification with mathematics and their interactions with math ST. However, their roles have been researched in isolation. Therefore, this study aimed to examine the moderating role of these individual difference factors and their interactions with the situation-specific contributors of math ST concurrently. This study first examined past research on ST and mechanisms by which ST affects performance, both in general and later directly pertaining to woman's math performance. It also examined previous research on the two aforementioned factors, test instructions and individual differences, and attempted to answer research questions pertaining to these contributors to math underperformance.

Background: Stereotype Threat Theory

Stereotypes are "shared beliefs about person attributes, usually personality traits, but often also behaviors, of a group of people" (Leyens, Yzerbyt, & Schadron, 1994, p. 11). As such, ST can be thought of as a situational predicament in which an individual experiences underperformance in the presence of shared beliefs pertaining to negative performance (Steele & Aronson, 1995; Steele, 1997). Numerous studies have demonstrated that situational pressure can make a relevant stereotype that alleges a lack of ability salient (e.g., Aronson et al., 1999; Johnson et al., 2012). Individuals need not believe or internalize the stereotype for it to influence behavior; the mere expectations of taking a relevant test when a negative stereotype is salient is adequate to induce ST (Steele, Spencer, & Aronson, 2002).

The concept of ST was introduced by Steele and colleagues in the 1990s where they observed underperformance of negatively stereotyped group members in high-stakes testing situations (Steele & Aronson, 1995). In general, ST affects performance when a stereotype was salient before the completion of a domain-relevant test. For instance, ST effects have been shown to result in underperformance when participants are asked to indicate their stereotyped demographic (ethnicity, SES, or sex; e.g., Croizet and Claire, 1998). However, a standard paradigm that was commonly used in ST research was first demonstrated by Claude Steele in a landmark ST study (Steele & Aronson, 1995). The study informed African Americans, who are negatively stereotyped with regard to math skills, that a test was indicative (diagnostic) of academic abilities or that the test was a laboratory problem-solving task (non-diagnostic or "not indicative") as part of the test instructions. African Americans in the ST conditions who were informed that the test was indicative of academic abilities performed worse relative to their European American counterparts. African Americans in the "indicative of ability" condition were also more likely to complete a word fragment task in a stereotype-confirming manner than those in the "test not indicative of ability" condition (i.e., by completing L O_ _ _ as LOSER; Steele & Aronson, 1995). This study demonstrates how a common feature of a high-stakes test (i.e., content of a test instruction) can evoke or make salient negative stereotypes that subsequently contribute to behaviors that are consistent with the stereotype.

Mechanisms Underlying ST Effects

ST appears to be a pervasive problem, hindering achievement in varying groups and domains. Latino/a students face challenges on math and analytical ability tasks (Gonzales, Blanton, & Williams, 2002). Individuals of a lower socioeconomic status (SES) encounter trouble with tasks measuring verbal abilities (Croizet & Claire, 1998). Even European American students, a non-minority, exhibit an achievement barrier relative to African American in tasks that involve "natural athletic ability" (Stone, Lynch, Sjomeling, & Darley, 1999). ST is not observed only across ethnicities in the United States. Of particular interest to this study is the ST effect found when women are in high-stakes math testing conditions (1999). How is it, though, that a problem defined by specific socio-cultural beliefs can be an almost universal dilemma? Based on past research, Schmader, Johns, and Forbes (2008) postulate a model that specifies how ST undermines math performance. Subsequent studies have utilized this model to explain how ST influences women's math performances. The sequence of processes that have been proposed to account for this interference in performance is outlined below.

Process I: Automatic Activation of ST

The model proposed by Schmader et al. (2008) posits that ST can affect performance without conscious awareness. Awareness of the relevant stereotype can be consciously experienced, but such awareness is not necessary. Many of the psychological, cognitive, and affective processes that are affected during ST happen automatically; resulting in outcomes that are in direct opposition to the person's conscious goals and intentions (Schmader, 2010; Schmader et al., 2008; Schmader & Beilock, 2012).

ST's effects on women's math performances. Schmader et al. (2008) postulates that women under ST experience a state of cognitive imbalance that stems from conflicts between a positive sense of oneself as a capable individual, and a situational cue that suggests poor performance and ability (Johns & Schmader, 2010). This state of imbalance is inconsistent with one's fundamental need for cognitive consistency (Festinger, 1957). Therefore, it sets into motion a series of physiological manifestations of stress, including cognitive and affective responses to cope with the experience (Major & O'Brien, 2005; Schmader et al., 2008). More specifically, a sense of uncertainty and self-doubt emerge in an attempt to resolve this state of imbalance (Beilock, Rydell, McConnell, 2007; Cadinu, Maass, Rosabiance, & Kiesner, 2005). A state of uncertainty emerges as ST raises psychological burden over two competing possible outcomes**:** to perform poorly as is consistent with the stereotype or to perform well and thus to the standards of one's own goals.

In an attempt to resolve these two competing outcomes, one searches for cues that can provide evidence against, or in support of, either outcome. Nevertheless, during this state of uncertainty, cues that may be innocuous, such as a simple error in calculations can be automatically misconstrued as a sign of failure (Schmader & Beilock, 2012). A study by Johns

Inzlicht, and Schmader (2008) demonstrates this process. Women in the ST condition showed increased vigilance toward anxiety-related words and showed subsequent math underperformance relative to women in a control condition (Johns et al., 2008). Another expression of uncertainty is self-doubt. A sense of self-doubt (even if activated outside of conscious awareness) colored women's interpretation in ways that disrupt optimal cognitive processing (Schmader, Forbes, Chang, & Mender, 2009). The above-mentioned research suggests that ST automatically triggers a sense of cognitive inconsistency. This sense of cognitive inconsistency includes uncertainty and increased vigilance towards cues that may signify the unwanted outcome of confirming the stereotype (Schmader & Beilock, 2012). These automatic processes are also accompanied by more controlled processes as one aims to manage one's behaviors, thoughts, and emotions.

Process II: Explicit Activation of ST

ST also contributes to increased task effort, decreased working memory capacity, and increased efforts to regulate thoughts and emotions, as well as, physiological stress and arousal.

Increased effort at the task. A core tenet of ST theory is ST's capability to increase one's motivation to disconfirm the stereotype. Nevertheless, this increase in motivation paradoxically leads to underperformance, depending on task difficulty (Schmader & Beilock, 2012). In a high-stakes testing situation, one's performance is at risk of being evaluated. Therefore, one's dominant response is to perform optimally, by exerting more effort. During tasks that are cognitively simplistic and well-learned, one's predominant response is to exert this extra effort.

ST's effects on women's math performances. Research has shown that women in a ST condition who were told that there was performance difference between women and men on a math test experienced a boost in speed on a relatively simple task: name-writing (Ben-Zeev, Fein, & Inzlicht, 2005). On the other hand, being in a ST condition impaired women's performance on a cognitively demanding task: writing one's name backwards as many times as possible for 20 seconds (Ben-Zeev et al., 2005). Another study demonstrates that increased motivation and effort when under ST can have negative effects on women's performance on a more difficult task (Schmader & Beilock, 2012). The study asked women to counter their automatic tendencies to focus on distracting cues that flashed on the periphery of a target. However, despite increased motivation, women in a ST condition who were told that the task was related to visuospatial and math ability before completing the task were less able to expend corrective effort to do well on the task (Jamieson & Harkins, 2007).

Decreased working memory capacity. ST has been shown to impair performance by taking away working memory resources (Beilock et al., 2007; Beilock, 2008; Schmader & Johns, 2003). Working memory capacity is essential to optimal performance because it allows focused attention on relevant tasks while inhibiting thoughts on irrelevant cues (Engle, 2002). Therefore, one's performance on a complex task can be undermined when working memory resources are temporarily depleted or are used to process other tasks (Schmader & Beilock, 2012).

ST's effects on women's math performances. The proposition that working memory depletion mediates the relationship between ST and women's underperformance in math is supported by research (Beilock et al., 2007; Croizet et al., 2004; Schmader & Johns, 2003). Women in the ST condition who were informed "women are poorer at math than men" experienced reduced working memory capacity (Schmader & Johns, 2003). Specifically, they recalled fewer words while they were completing math problems than women in the control condition who were not provided with information regarding differences in math performance between women and men. These studies supported the proposition that reduced working memory causes underperformance under ST.

One of the ways by which ST taxes working memory is via task ruminations (Schmader & Beilock, 2012). Steele et al. (2002) posit that ST contributes to “concerns about how one will be perceived, doubts about one’s ability, and thoughts about the stereotype” (p. 392). Beilock (2008) argued that the awareness of the negative stereotype saturated working memory resources with internal worrying over one’s performance. Indeed, research has shown that women under ST reported an increased number of negative thoughts. Moreover, an increase in the number of negative thoughts during the first half of a math test was associated with underperformance at the second half of the test (Cadinu et al., 2005). Nevertheless, another study demonstrated that worrying only partially explained (mediated) the relationship between ST and impaired performance (Beilock et al., 2007).

Additional research demonstrated that self-doubts undermined working memory when anxiety was present (Schmader et al., 2009). As mentioned above, a sense of cognitive imbalance under ST triggers a sense of self-uncertainty. Schmader and Beilock (2012) argued that efforts to interpret and resolve this sense of uncertainty require effortful and controlled processing, which depletes working memory and takes cognitive resources away from the task at hand. Overall, research demonstrated that ST contributes to a redirection of working memory into processing negative self-relevant stereotype information.

Increased efforts to regulate thoughts and emotions. As mentioned above, ST increases one's motivation to disconfirm the stereotype of math inability. However, it also triggers anxiety and self-doubt. Cognitive efforts that are expended to suppress these negative emotions are the same ones needed to do well at cognitive tasks.

ST's effects on women's math performances. An experiment using a dot probe task demonstrated that women actively suppressed their anxiety when informed that the task was meant to assess their anxiety under ST. This group of women was faster to identify a dot that was in the same position as a neutral word (instead of anxiety-related word). However, this inhibitory control lowered their working memory capacity, contributing to less than optimal performances on a subsequent task (Johns et al., 2008).

Physiological processes underlying ST. Several studies demonstrated that physiological stress and arousal also contributed to underperformance under ST (Ben-zeev et al., 2005; Inzlicht & Kang, 2010; O'Brien & Crandall, 2003; Osborne 2006, 2007; Vick et al., 2008). ST also affected the physiological processes of a group of individuals with a widely-known stereotype of "lower intelligence" (i.e., undergraduate psychology majors as compared with science majors). Psychology majors in the ST condition who were told that a math test was an indicator of general intellectual ability experienced higher heart rate variability (HRV) relative to when a test was described as "not diagnostic of any ability." This increase in HRV, an indirect indicator of increased mental workload, was followed by math underperformance (Croizet et al., 2004).

ST's effects on women's math performances. Women in the ST condition who were told that there were performance differences between women and men exhibited increased systemic vascular resistance, which is indicative of increased mental load, as compared to women in the control condition who were told that women and men did not differ in their performances on the test (Vick et al., 2008). Another study found that women experienced signs of increased stress and arousal such as increased heart rates, greater skin conductance, and greater sympathetic activation of the cardiovascular system in response to ST (Murphy, Steele, & Gross, 2007).

Women experiencing ST also exhibited lower levels of activation in brain regions associated with problem solving that are thought to be important for optimal math performance. Further, when compared to women in the control condition, those who were reminded of the stereotype of math inability experienced increased activation in their ventral anterior cingulate cortex, a region that is responsible for detecting and processing emotion-based information (Krendl, Richeson, Kelly, & Heatherton, 2008). Overall, there is converging evidence that women under ST experience negative emotions and physiological responses. Attempts to regulate these responses overload women's cognitive resources, negatively impacting their math performance.

CHAPTER 2: LITERATURE REVIEW

Among the most extensively researched topic in the ST literature is the effects of ST on women's math performances. ST has been found to undermine performance when an evaluative situation makes salient a relevant negative stereotype (i.e., in high-stakes math testing situation; Johnson et al., 2012). A woman experiences math underperformance in a high-stakes testing situation because "when women perform math, unlike men, they risk being judged by the negative stereotype that women have weaker math ability (Spencer et al., 1999, p. 4)."

A classic study by Spencer and colleagues (1999) demonstrated the effect of ST on women's math performance via test instruction. In the study, women and men of equal ability completed a difficult math test. Women who were in a ST condition (they were told of presumed differences in math performance between women and men) performed worse than men. However, in conditions where women were not told of differences in math performance between women and men, women performed equally to men. Additionally, another group of women in a ST condition who were informed of a difference in math performance between women and men performed worse than women who were placed in a "no mention" condition, where they were not told anything about differences in math performance (Spencer et al., 1999). Following Spencer and colleagues' (1999) seminal work, numerous follow-up studies demonstrated the same ST effect on women's math performance and situational factors influencing women's susceptibility to ST (Aronson et al., 1999, Study 1; Johnson et al., 2012; Keller, 2002; Keller & Dauenheimer, 2003; Rydell, McConnell, & Beilock, 2009; Schmader, 2002; Smith & White, 2002; Spencer et al., 1999; Steele et al., 2002; Steinberg et al., 2012).

ST induced by situational factors in a high-stakes context can partially explain why on average, U.S. women performed worse than their male counterparts on high-stakes math tests, while performing equally well or better than men using other metrics of math knowledge (Ceci et al., 2009; The College Board, 2011; Willingham & Cole, 1997). Indeed, a recent meta-analysis found that ST contributed to female math underperformance by a quarter of a standard deviation, which is approximately 33 points out of a possible 800 points below the predicted average score of female test takers on the quantitative section of the Graduate Record Examination or GRE-Q (Picho, Rodriquez, & Finnie, 2013).

Furthermore, some findings suggest that the extent to which situational factors in high-stakes math-testing situations, such as test instructions, influence women's math performance (i.e., math ST) can vary depending on important individual difference factors (e.g., McFarland et al., 2003; Oswald & Harvey, 2000–2001; Schneeberger & Williams, 2003). In support of Kurt Lewin's field theory that suggests that behavior is a function of environmental factor and personal characteristics (Lewin, 1943), these studies' findings suggest that "ST is a multifaceted situational predicament whose mediational path can be shaped by features of the person, the context, and their interaction" (Davies, Spencer, & Quinn, 2002, p. 1617). However, existing studies have examined the moderating role of these individual difference factors in isolation. By contrast, this study simultaneously investigated the role of these individual difference factors, situational factors, and their interaction. Specifically, this study investigated whether identification with math (i.e., the extent to which a woman identifies with math; Steele et al., 2002) and academic major (i.e., whether a woman is a math-intensive major or a not in a math-intensive major; Crisp et al., 2009; Croizet et al., 2004) increase and decrease women's susceptibility to the ST-induced test instructions. This literature review discusses under what conditions math ST occurs, focusing on the influence of two types of ST-specific test instructions and for whom, focusing on the individual difference factors that moderate women's susceptibility to math ST, as well as their interaction.

Conditions Under Which Math ST Effects Occur

Features of a math-testing situation. Research has demonstrated that ST is more likely to occur in high-stakes situations (e.g., in high-stakes math-testing situations). In this context, women have reported a fear of confirming the stereotype via their performances on the test, thereby, increasing their susceptibility to ST (Steele, 1997). Following the landmark 1995 study (Aronson & Steele, 1995), researchers have demonstrated the effects of ST in an experimental testing situation by manipulating the salience of the stereotype. According to subsequent research, several features of a math testing situation can be responsible for increasing the salience of the stereotype, thereby, inducing math ST (Spencer et al., 1999). Therefore, a majority of studies focus on identifying features in a math-testing environment that can induce ST and interfere with women's math performances (Aronson & Dee, 2012; Kiefer & Sekaquaptewa, 2007; Wout, Danso, Jackson, & Spencer, 2008).

Common ways to induce ST in a math-testing situation (or math ST) involve increasing the salience of stereotypes of math inability experimentally, as a part of a research design. Some studies increased the salience of the stereotype and induce ST in high-stakes math testing situation in an indirect and subtle way, by asking women to indicate their sex before completing a math test (e.g., Croizet & Claire, 1998; Schmader, 2002; Shih, Pittinsky, & Ambady, 1999; Steele & Aronson, 1995), or by varying the ratio of female to male test-takers in math examinations (Ben-Zeev, Fein, & Inzlicht, 2005; Inzlicht & Ben-Zeev, 2000, 2003; Murphy, et al., 2007; Sekaquaptewa & Thompson, 2002; 2003; Schmader & Johns, 2003). However, the most predominant way to induce math ST is to vary the manner in which a math test is described via instructions prior to test taking.

The role of test instructions. The two most common ways to induce math ST via test instructions are to inform women that a math test they are about to take is a valid indicator (diagnostic) of math ability (e.g., Kiefer & Sekaquaptewa, 2007) or to tell women that men perform better in that specific math test (Brown & Pinel, 2003; Rosenthal & Crisp, 2006; 2007). Below is a more in-depth review of the relevant studies focusing on the two most commonly utilized ST-specific test instruction manipulations.

ST-test instruction conditions that suggest the validity of a math test. Introducing a math test as one that is indicative of math ability can increase the salience of the stereotype and lead to ST and thus, underperformance (Aronson & Steele, 2005; Good & Aronson, 2008; Steele, 1997). This is because assigning women to a ST condition (i.e., by informing them that a math test as indicative of math ability) interferes with optimal cognitive processing needed to excel at the task at hand (Beilock et al., 2007; Osborne, 2007; Schmader & Johns, 2003). On the other hand, assigning women to a control condition (i.e., by presenting a test as not indicative of math ability) suggests that the stereotype of math inability is no longer applicable to the task; hence, women are less likely to experience math underperformance due to math ST (Cadinu et al., 2005). Numerous studies have supported the link between assigning women to a ST condition (i.e., presenting a math test as indicative of math ability) and women's math underperformance (Bell et al., 2003; Carr & Steele, 2009; Kiefer & Sekaquaptewa, 2007; Crisp et al., 2009; Croizet et al., 2004; Davies et al., 2002; Good et al., 2008; Martens, Johns, Greenberg, & Schimel, 2006; Nussbaum & Steele, 2007; Quinn & Spencer, 2001; Smith, Sansone, & White, 2007; Smith &

White, 2002; Steele et al., 2002; Steele & Aronson, 1995, Study 1; Wout et al., 2008, Wout, Shih, Jackson, & Sellers, 2009).

ST- test instruction conditions that compare women's and men's math performances. Math ST can also be induced into a math testing situation by alluding to the stereotype of woman's math inferiority via test instructions (e.g., by suggesting an unfavorable performance comparison with men's or by mentioning that there is no differences in performance; Steele et al., 2002). On the other hand, in conditions where a math test is described such that the negative stereotype is not relevant to it (e.g., when performance differences are not mentioned), women's math performances match that of the nonstereotyped group (e.g., Spencer et al., 1999).

Some studies induced math ST by mentioning a potential difference in math performance (i.e., by stating that the study's purpose is to examine the math performance difference between women and men; Brown & Pinel, 2003; Johns, Schmader, & Martens, 2005). Other studies induced math ST by informing women that the performance difference between women and men has been documented on the upcoming math test (e.g., O'Brien & Crandall, 2003) or that the test has been shown to produce sex differences (Keller, 2007; Keller & Molix, 2008; Spencer et al., 1999, study 2 & 3). Others have informed women that a difference in math performance exists, without specifying the direction of the difference (a "mention of difference" condition; McIntyre, Paulson, & Lord, 2003; Vick et al., 2008). In addition, the effect of reminding women of the stereotype is so strong that even mentioning that women had performed equally to men on the test (a "mention of no difference" condition) has been shown to induce math ST (Cadinu et al., 2003; Steinberg et al., 2012). However, the most common way to induce math ST via test instruction is to explicitly compare women's and men's performances in the test (i.e., a "men perform better than women" condition). Math underperformance has been shown among women who are exposed to this explicit instruction (Cadinu et al., 2003; Inzlicht & Schmader, 2012; Johnson et al., 2012; Keller, 2002; Keller & Dauenheimer, 2003; Kiefer & Sekaquaptewa, 2007; Rivardo, Rhodes, Camaione, & Legg, 2011; Rydel et al., 2009; Schmader, 2002; Smith et al., 2007; Smith & White, 2002; Spencer et al., 1999; Steinberg et al., 2012; Thoman et al., 2008).

Gaps in research examining the role of test instructions on math performance.

The inclusion of control conditions. Research has demonstrated math underperformance under ST conditions, that is when a math test is presented as indicative of math ability. However, the effect of this test instruction cannot be clearly drawn, as much of the existing research on math ST does not include a condition in which a math test is presented as not indicative of math ability (see Croizet et al., 2004 for an exception). In order to draw valid conclusions on the impact of the indicative of math ability instruction, a “not indicative of math ability” condition needs to be included as a part of a study design.

Similarly, the effect of test instruction that compares women’s and men’s math performances cannot be conclusively determined as studies have utilized different control conditions. According to ST theory, a comparable control condition to a “men perform better than women” condition is one in which math performance difference between women and men is not mentioned (i.e., a “no mention of differences” condition; Johnson et al., 2012; Keller, 2002; Rydell et al., 2009, study 1; Schmader, 2002; Spencer et al., 1999, study 2 & 3; Steinberg et al., 2012). This is because it is more reflective of a standard test instruction in a math testing situation. However, as can be seen on table 1, some studies compared the “men perform better than women” condition to different types of control conditions. These control conditions include: (a) one that mentioned, but also nullified differences in math performance (e.g., “the following math test is a collection of questions which have been shown not to produce gender differences in the past;" Keller, 2007, p. 330; Jamieson & Harkins, 2012; Keller & Molix, 2008; O'Brien & Crandall, 2003; Spencer et al., 1999; study 3; Vick et al., 2008), (b) one in which math performance difference is mentioned without a specific direction (i.e., a "mention of differences" condition; Smith & White, 2002), (c) one that mentions the stereotype, but also nullify any differences in math performance (Keller & Dauenheimer, 2003: Smith & White, 2002), and (d) one that suggests equal performance between women and men (i.e., a "mention of no difference" condition; Cadinu et al., 2003; Steinberg et al., 2012). These limitations suggest the need to include a "test not indicative of math ability" condition in order to determine the effect of the "test indicative of math ability" condition on women's math performances. In addition, limitations of research examining the effect of test instruction that compares women's math performances relative to men suggest the need for a more thorough examination of the effects of the "men perform better than women" condition relative to the "no mention of differences" condition.

The effect of instruction suggesting women's superiority in math. Previous research has focused on test instructions that make the negative stereotype in math salient for women. However, the effect of presenting counter-stereotypical information in which women perform better than men (i.e., a "women perform better than men" condition), has been relatively under investigated. One explanation for this gap is that the prevalence of the stereotype of women's math inability may limit the credibility of test instruction that alleges superior math performance by women.

Findings from few studies that investigate the effect of the "women perform better than men" test condition suggest that a reminder of a counter-stereotypic information has the potential to elevate women's performance on a math test and buffer women from the experience of math

ST (Cadinu et al., 2003; Johnson et al., 2012). It has been suggested that this is due to increased self-esteem and self-efficacy (Chalabaev, Stone, Sarrazin, & Croizet, 2008; Fein & Spencer, 1997; Walton & Cohen, 2003) from the counter-stereotypes possibly lowering the accessibility of the negative stereotypes. Lowered accessibility of negative stereotypes can compel women to adopt more flexible and creative cognitive strategies, which boost performance (Goclowska, Crisp, & Labuschagne, 2012).

However, to date, there are only two math ST studies that have examined the effect of the "women perform better than men" condition (Cadinu et al., 2003; Johnson et al., 2012). These studies include different control conditions, which limit conclusions that can be drawn on the effect of the "women perform better than men" condition on women's math performances. Cadinu et al. (2003) found that women in the "women perform better than men" condition, who were informed "on tasks that evaluate logical–mathematical abilities, obtained higher scores than women in the "men perform better than women" conditions and control conditions (i.e., a "mention of no difference" condition) who were told that: 'in the logical–mathematical domain no differences between men and women emerge… men and women show, on average, equivalent scores" (Cadinu et al., 2003, p. 273). On the other hand, a more recent study found that women who were assigned to the "women perform better than men" condition (i.e., those were told that "women are expected to do better than men") performed better than those assigned to the "men perform better than women" conditions, but equally to those assigned to the control condition in which women were "neither women nor men are expected to do better or worse" (Johnson et al., 2012, p. 141).

Studies that included both validity of math test and women-men math performance differences conditions. The majority of past research has only utilized one of the aforementioned types of test instructions to induce ST. These studies present a math test as either indicative/not indicative of math abilities or set up a condition in which women are told of their math performances relative to men on the test. Therefore, there has not been much exploration into scenarios where there can be interactions between these influences, more closely simulating the multiple ST-related factors potentially faced by women in a standard math-testing situation.

To date, three published studies have examined the interactive effects of these two ST-specific instructions. According to ST theory, the “indicative of math ability” condition when combined with instruction that remind women of the stereotype of math ability can further increase the belief that their performances can confirm the stereotype. However, no study has test this hypothesis in a complete factorial design. Bell et al. (2003) and Good et al. (2008) crossed the dimensions of the two ST-specific instructions to determine their interactive effects in an incomplete factorial design. In their studies, all women in the ST conditions were informed that a math test was indicative of math ability, and those who were also informed that the same test was one in which women have performed equally to men (i.e., a “mention of no difference” condition) perform better. Another study compared the effects of a different set of test instructions (Schmader, 2002). In this study, all women in the ST conditions were informed that the math test was indicative of their ability, and those who were also informed that women and men’s scores would be compared performed worse (Schmader, 2002). However, these studies do not include a “test not indicative of math ability” condition, which could aid in determining how each condition alone or in combination can negatively affect women’s math performances. These mixed findings suggest the need to determine the interactive impact of these test instructions in a full factorial design, that is, to examine the effect of the different levels of the comparison of women and men on math performance variable (i.e., women-men math performance differences levels) in the “test indicative of math ability” conditions relative to the “test not indicative of math ability” conditions (see Figure 1).

		Variable 2 Validity of Math test Instruction	
		Level 1 “Test Not Indicative of Math Ability”	Level 2 “Test Indicative of Math Ability”
	Variable 1 “No Mention of Differences”	Condition 1: Stereotype not salient Evaluation not salient	Condition 4: Stereotype not salient Evaluation salient
Condition 1 Women-Men Math Performance Differences Instruction	Level 2 “Men Perform Better than Women”	Condition 2: Stereotype salient Evaluation not salient	Condition 5: Stereotype salient Evaluation salient
	Level 3 “Women Perform Better than Men”	Condition 3: Stereotype salient Evaluation not salient	Condition 6: Stereotype salient Evaluation salient

Figure 1. The “Validity of Math Test” and “Women-Men Math Performance Differences” instructions in a full factorial design

Limitations of studies that include both validity of math test and women-men math performance differences test instructions.

The inclusion of “men perform better than women” conditions. Findings on studies that include both indicative of math ability and comparison of women and men on math performance manipulations are equivocal. Methodological limitations on these studies’ findings do not allow for a clearer understanding of how each factor alone or in combination negatively affect women's math performances. The three existing studies induce ST in different ways (Bell et al., 2003; Good et al., 2008; Schmader, 2002). Schmader's (2002) ST manipulation consisted of describing a math test as indicative of math ability and suggesting that women's and men's scores would be compared to one another (a "mention of difference" condition). However, the ST manipulation utilized by Bell et al. (2003) and Good et al. (2008) consisted of describing a math test as indicative of math ability and suggesting that men perform equally to women (a "mention of no difference" condition). These studies found math underperformance in conditions that remind women of the stereotype, regardless of content. However, no study has yet to examine the role of test instruction that emphasizes the validity of a math test when it is suggested that men performed better than women (i.e., a "men perform better than women" conditions) relative to when it is not suggested (i.e., a "no mention of differences" conditions).

The inclusion of control conditions. Studies that include both test instructions have not included a condition in which a math test is presented as not indicative of math ability (Bell et al., 2003; Good et al., 2008; Schmader, 2002). In order to draw a conclusion on the amount of "threat" that is experienced and its effect on women's math performance, it is necessary to include all possible levels of test instructions in a complete factorial design (i.e., by comparing the "test indicative of math ability" conditions with the "test not indicative of math ability" conditions). Additionally, the three studies compare their comparison of women and men on math performance conditions to one control condition (i.e., a "no mention of differences" condition). As these studies neglect to include a "test not indicative of math ability" condition, they have yet to determine the role of the test instruction that emphasizes the validity of math test in conjunction with women-men math performance differences test instructions that explicitly suggests or does not suggest a difference of math performance between women and men (i.e., the “men perform better than women” conditions relative to the “no mention of differences” conditions).

General gap in research examining the role of ST-specific test instructions.

The inclusion of manipulation checks. While much of ST research focuses on the effect of the experimental induction of ST on performance (Palumbo & Steele-Johnson, 2014), it is essential to determine the extent to which participants remember the content of the manipulation (Appel, Kronberger, & Aronson, 2011; Gibson, Losee, & Vitiello, 2014; Inzlicht & Kang, 2010). In other words, as part of an experimental design, it is necessary to include post-test questions that is designed to assess the effectiveness of the manipulations used to induce ST. However, prior math ST studies examining the independent and interactive role of test instructions have seldom reported the effectiveness of their manipulations (see Jamieson & Harkins, 2009; 2012 for exceptions). Brown and Day (2006) have argued that without such checks, we cannot be sure that ST is responsible for any observed effects.

Individual Difference Factors in Math ST

The preceding section discussed the effect of ST-specific test instructions on women’s math performance. Nevertheless, ST has been shown to be more than a situational predicament elicited “by the mere recognition that a negative stereotype could apply to oneself in a given situation” (Steele, 1997, p. 617). Consistent with Kurt Lewin’s field theory (Lewin, 1943), women’s susceptibility to math ST is influenced by both situational factors (i.e., test instructions) and individual difference factors. In support of studies suggesting the importance of individual difference factors to math ST (e.g., McFarland et al., 2003; Oswald & Harvey, 2000–2001; Schneeberger & Williams, 2003), existing research has identified important individual difference factors that moderate women’s susceptibility to math ST (Steele, 1997). Individual difference factors that have been shown to increase women's susceptibility to math ST is women's strength of identification with math (i.e., the extent to which women care about being good at math; Keller, 2007; Steinberg et al., 2012). This is because the more a woman is identified with a stereotyped group (i.e., math); the more she expects to be perceived as a member of the group, the more motivated she is to disapprove the stereotype when reminded of it. On the other hand, findings from another two studies suggest that a particular group membership, such as a woman's academic major can decrease women's susceptibility to math ST (Crisp et al., 2009; Croizet et al., 2004). The following section focuses on how the two individual difference factors affect women's math performances under math ST.

The role of identification with math. Women's strength of identification with the domain being assessed (i.e., math) is theorized to be an important individual factor that increases women's susceptibility to ST (Steele, 1997). ST theory predicts that women who believe they are good in math and care about doing well in math (i.e., women with strong identification with math) should experience increased pressure to disconfirm the alleged stereotype as it poses a "threat" to their identity as a capable individual. Highly math-identified women were more likely to experience additional pressure and anxiety due to ST, which subsequently interfere with their math performances (Schmader & Beilock, 2012). A woman with low identification with math should not experience additional pressure to disconfirm the stereotype. Therefore, their performance should not be affected in a math ST situation (Aronson et al., 1999; Keller, 2002; Steele, 1997; Steinberg et al., 2012). However, a meta-analysis conducted to determine the moderating impact of ST on women's math performance found that women with a moderate level of identification with math were more susceptible to the ST effect (mean $d = |.52|$, $k = 6$)

than those with strong identification with math (mean *d* = |.29|, *k* = 9) (Nguyen & Ryan, 2008). The following section discusses reasons for this discrepancy.

Gaps in research— the role of identification with math in women's math ST susceptibility.

Equivocal findings. A meta-analysis suggests that women who are moderately math-identified are more susceptible to ST than women who are highly identified with math (Nguyen & Ryan, 2008). Therefore, according to the meta-analysis but inconsistent with ST theory, math-identified women are not more susceptible to ST-specific test instructions than women low in identification with math. One explanation for this discrepancy is that existing studies have chosen to accept the assumption that math-identified women are more vulnerable to ST without experimentally testing the moderating role of identification with math on women's susceptibility to math ST (Nguyen & Ryan, 2008). The following is a review of the various methodological limitations of existing studies.

Selective recruitment. Some studies examining the effect of math ST have chosen to engage in a selective recruitment strategy. Instead of recruiting women with high and low identification with math in order to determine the moderating role of identification with math, existing studies have either sampled women with relatively high SAT math scores (e.g., Cullen, Hardison, & Sackett, 2004; Martens et al., 2006; Marx & Roman, 2002; Quinn & Spencer, 2001; Schmader, 2002; Schmader & Johns, 2003; Shih et al., 1999; Spencer et al., 1999), or women with a particular grade point average (e.g., Aronson, et al., 1999, study 2; Bell et al., 2003; Good et al., 2008). As indicated by the meta-analysis, this sampling limitation may mean that any observed difference in math underperformance in ST math-testing situations can be attributed to ceiling effects and regression to the mean (Nguyen & Ryan, 2008). Some studies subsequently control for SAT scores to account for this limitation (e.g., Gonzales et al., 2002; Inzlicht & Ben-Zeev, 2000, 2003; Keller, 2002, 2007; Lesko & Corpus, 2007; Marx & Roman, 2002; Schmader, 2002; Schmader & Johns, 2003; Vick et al., 2008; Wout, et al., 2008; 2009). In addition, based on the assumption that high performance in performance equals strong identification with math, some studies utilize identification with math as a selection variable, recruiting women on the basis of the importance they place on math and math performances (e.g., Ben-Zeev et al., 2005; Schmader & Beilock, 2012). This selective sampling strategy may have resulted in selection bias: an oversampling of women with supposedly strong identification with math (Nguyen & Ryan, 2008). However, it results in another limitation that is discussed in the following section.

Indirect assessments of identification with math. In addition to a selective recruitment strategy, past studies have neglected to utilize valid and direct assessment of identification with math. Instead, as reported by a recent meta-analysis (Nguyen & Ryan, 2008), many studies utilize an indirect measures of identification with math. Prior ability measures (e.g., SAT math scores; Spencer et al., 1999) are used instead of a validated measure of identification with math because of the assumption that high-performing women must strongly identify with math (Nguyen & Ryan, 2008; Picho et al., 2013). However, strength of identification in math, as measured by Smith and White’s (2001) Identification with math measure is only moderately correlated with prior math performances or scores, suggesting that strong identification with math does not always equal strong math performances (Nguyen & Ryan, 2008). This limitation undercuts the validity of existing studies. In order to determine the moderating role of identification with math, future research needs to refrain from selective recruitment and utilize a direct and valid measure of identification with math.

The impact of both types of test instructions. The various limitations outlined above mean that there are very few studies that have directly investigated whether women's identification with math moderate the effect of ST-specific test instruction. In fact, there are only two studies that have not engaged in selective strategies and utilized indirect identification with math measures (Keller, 2007; Steinberg et al., 2012). Keller (2007, p. 330) found that high school girls with high identification with math performed worse when informed that a math test "had been shown to produce gender differences" (those in the ST condition) relative to those who were informed that a math test "had not been shown to produce gender differences"(those in the control condition). Another study tested the moderating effect of identification with math among a sample of upper-level undergraduate women majoring in engineering, math, or physical science who had completed at least three semesters of calculus classes (Steinberg et al., 2012). Steinberg et al. (2012) found that relative to women with low identification with math, women with strong identification with math in the ST condition (i.e., those who were informed that men had outperformed women) performed worse than those in the control condition (i.e., those who were not informed of performance difference between women and men). These studies tested the moderating role of identification with math, but they do not investigate the influence of both types of test instructions in a full factorial design. More research is needed to determine whether the effect of test instruction that emphasizes the validity of a math test when it is also suggested or not suggested that men had outperformed women in a math test (i.e., the "men perform better than women conditions relative to the "no mention of differences" conditions) varies depending on women's strength of identification with math.

The role of academic major. The large body of research on ST has focused on taking into account the role of individual difference factor, such as identification with math, in increasing women's susceptibility to math ST. However, equally important advances can be made from studying factors that may lower women's susceptibility to math ST effect. According to ST theory, a high level of self-involvement in the relevant domain (i.e., math) can increase women's susceptibility to math ST (Steele 1997). Beyond women's strength of identification with math, ST theory also suggests important group membership differences that may moderate the extent to which women are impacted by math ST. Extending the ST theory, findings from two existing studies suggest how a woman's academic membership (i.e., whether a woman is in a math-intensive major or not in a math-intensive major) can have an important role in moderating the extent to which women experience math underperformance due to math ST (Crisp et al., 2009; Croizet et al., 2004). Based on ST theory, women in math-intensive majors (e.g., math, engineering) should, theoretically, experience increased susceptibility to math ST due to the salience and self-relevance of stereotype of math inability in their everyday educational environment (e.g., in math classes and/or in math-testing situations) relative to women not in math-intensive majors (e.g., history, psychology). In support of this assertion, several research has demonstrated that underperformance due to math ST can occur from simply being a numerical minority in a negatively stereotyped domain (Logel, Iserman, Davies, Quinn, & Spencer, 2009) or when being outnumbered by men when taking a math test (Inzlicht & Ben-Zeev, 2000, 2003; Inzlicht & Good, 2006; Murphy et al., 2007; Sekaquaptewa & Thompson, 2003). Most importantly, one study found math underperformance among female engineering majors who were exposed to ST-specific test instructions (Bell et al., 2003)

In contrast, the two existing studies found that female engineering majors were less susceptible to ST-specific test instruction than psychology majors (Crisp et al., 2009; Croizet et al., 2004). One study found that female engineering majors experienced enhanced math performance, while female psychology majors (considered not a math-intensive) experienced math underperformance in the ST conditions (i.e., when informed, “on this second test we shall be comparing the performance of males to females”; Crisp et al., 2009, p. 175). Similarly, another study found that psychology majors performed worse than science majors did in the ST conditions (i.e., when informed that a test they were about to take was a “valid measure of the general intellectual ability involved in mathematical and logical reasoning”) than when an identical test was presented as a “laboratory exercise related to polygraphic recording of autonomic activity not diagnostic of any ability” (Croizet et al., 2004, p. 725). These findings suggest that being a member of a group traditionally stereotyped to have lower math ability (i.e. women) does not uniformly increase female engineering majors’ susceptibility to math ST. Specifically, being an engineering major, a group stereotyped to have superior math ability, appears to attenuate women’s susceptibility to math ST. These findings suggest the importance of investigating the moderating effects of academic major as it may mitigate the pernicious effect of math ST.

Findings from the two studies are inconsistent with the original conceptualization of ST (Steele, 1997). The authors of the aforementioned studies, Crisp et al. (2009) and Croizet et al. (2004) reason that, although women in engineering and psychology are both aware of negative stereotypes regarding women’s math abilities, women majoring in engineering may have developed strategies to contend with the stereotype of math inability. These strategies may have developed over time in response to the increasing number of math classes and math examinations engineering and science majors have to take (Crisp et al., 2009; Croizet et al., 2004).

Gaps in research—the role of academic major.

Generalizability. Research has only examined the effects of ST-specific test instructions among psychology and science majors (Crisp et al., 2009; Croizet et al., 2004). To date, most math ST research has neglected to take into account the role of academic major or has engaged in selective recruitment to target those who were theorized to be more susceptible to math ST (e.g., Steinberg et al., 2012). Instead, research needs to address this limitation by recruiting women from a variety of academic majors to determine whether being a math-intensive major can decrease women's susceptibility to math ST than being in a not math-intensive major. More specifically, more research is needed to determine whether the effect of the validity of math test instruction in conjunction with a condition that explicitly suggests (or does not suggest) a comparison of math performances between women and men (i.e., the "men perform better than women conditions relative to the "no mention of differences" conditions) varies depending on women's academic major.

The impact of both types of test instructions. Existing studies have not investigated the effects of both types of test instructions. Crisp et al. (2009) examined the effects of a math test that was described as indicative of math ability relative to an identical test that was described as not indicative of math ability. On the other hand, Croizet et al. (2004) investigated the effect of ST that was presented in an indirect way (a "mention of no difference" manipulation). Both studies manipulated math ST instructions in a different way and included different control conditions. Additionally, both studies did not investigate the influence of both types of test instructions in a full factorial design while simultaneously exploring the impact of two individual difference factors (i.e., identification with math and academic major). Further research is needed to determine the role of the test instruction that emphasizes the validity of a math test in conjunction with a condition that explicitly suggests or does not suggest a comparison of math performances between women and men (i.e., the “men perform better than women conditions relative to the “no mention of differences” conditions) among women in math-intensive majors and women not in math-intensive majors.

Current Study

Rationale and Research Questions

The present study was designed to investigate two research questions exploring the situational and individual difference factors that contribute to math ST: 1) what are the main and interactive effects of two types of ST- inducing test instructions on women’s math performances (the Validity of Math Test and Women-Men Math Performance Differences Variable)

2) do two individual difference factors (identification with math and academic major) moderate the impact of these ST-inducing test instructions?

According to ST theory, the most direct way to induce math ST is by including information suggesting women’s poor math performance relative to men within a math test’s instructions (i.e., women-men math performance differences conditions). Indeed, past research on test instructions found that ST-specific test instructions exerted a stronger effect on women’s math performance than test instructions that emphasized the evaluative nature of the test (validity of math test conditions; e.g., Johnson et al., 2012). Therefore, it was hypothesized that women who were informed that men had performed better (i.e., those in the “men perform better than women” conditions) would score lower on the math test than women who were not informed of differences in math performance between women and men (i.e., those in the "no mention of differences" conditions).

This study attempted to extend ST theory by determining the interactive effect of both types of test instructions, whether both test instructions interact to influence women's math performance, simulating the stress in a standard testing situation. Specifically, this study also aimed to determine whether the impact of the Women-Men Math Performance Differences Variable (i.e., the difference in performance in conditions where women were told and not told that men had performed better) would be greater on women's math performance in a condition where the validity of the math test was emphasized than when the test's validity was not mentioned. No past studies that have included both types of test instruction have included a "test not indicative of math ability" conditions in a complete factorial design in order to determine how the test instructions alone or in a combination impact women's math performances. However, based on findings from past studies (e.g., Bell et al., 2003), it was hypothesized that women in the "men perform better than women" conditions would be more negatively impacted when the validity of the math test was emphasized ("test indicative of math ability" conditions) relative to when it was not emphasized ("test not indicative of math ability" conditions).

This study also aimed to determine the moderating effect of individual difference factors, as identified by ST theory (Steele et al., 2002). Specifically, this study aimed to determine whether the impact of the of the Women-Men Math Performance Differences Variable in conditions where the evaluative nature of the test is emphasized would be greater for women high identification with math than for women low in identification with math. Based on ST theory (Steele, 1997; Steele et al., 2002) and findings of the only two studies that have examined the moderating role of identification with math (Keller, 2007; Steinberg et al., 2012), it was hypothesized that the impact of Women-Men Math Performance Differences Variable would be greater on the math performance of women high identification with math in conditions where they were informed of the evaluative nature of the test than on the math performance of women high in math identification with math in conditions where they were not informed of the evaluative nature of the test. In addition, the math performance of women who were low in identification with math should not be negatively impacted by these test instructions.

This study also builds on past studies' findings by testing how a woman's college major may affect her susceptibility to math ST. This is because while research has suggested its role in decreasing women's susceptibility to math ST, no study has determined its moderating effect with the different levels of two types of test instructions included as a part of a study design. Based on predictions from ST theory and findings from past studies (Crisp et al., 2009; Croizet et al., 2004), it was hypothesized that the impact of the Women-Men Math Performance Differences Variable would be greater on the math performance of women in math-intensive majors in conditions in which the evaluative nature of the test was emphasized than on the math performance of women in math-intensive majors in conditions in which the evaluative nature of the test was not emphasized. In addition, the math performances of women not in math-intensive majors should not be negatively impacted by these test instructions. Further, this study examined the interactive impact of the Women-Men Math Performance Differences Variable, Identification with Math, and Major. It was hypothesized that the impact of the Women-Men Math Performance Differences Variable would be greater on the math performance of women in math-identified math-intensive majors than on the math performance of women in math-intensive majors who are low in identification with math. Conversely, the math performances of women not in math-intensive majors should not be negatively impacted by these test instructions.

Finally, given the limitations of past research on the impact of instruction suggesting underperformance by male test takers (i.e., "women better than men" conditions, e.g., Cadinu et al., 2003), exploratory analyses were conducted to determine how its effect on math performance may vary depending on women's majors and strength of identification with math.

Hypotheses

This study specifically aimed to test the following hypotheses:

	Men Perform Better than Women	Women Perform Better than Men	No Mention of Differences in Math Performance	
Test Indicative	μ_{11}	μ_{12}	μ_{13}	$\mu_{1.}$
Test Not Indicative	μ_{21}	μ_{22}	μ_{23}	$\mu_{2.}$
	$\mu_{.1}$	$\mu_{.2}$	$\mu_{.3}$	

H1: This study hypothesized a main effect of the Women-Men Math Performance Differences Variable, such that the women assigned to the "men perform better than women" conditions would score lower on the math test than women in the "no mention of differences" conditions.

$$(\mu_{.1} \neq \mu_{.3})$$

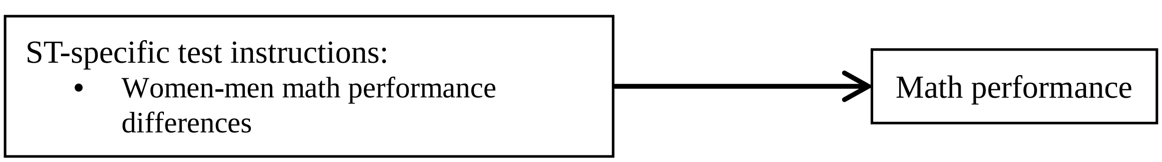

Figure 2. The Effect of Women-Men Math Performance Differences Instruction on Math performance

H2: This study hypothesized a two-way interaction between the Validity of Math Test and the Women-Men Math Performance Differences Variable, such that the impact of the women-men math differences levels (the difference in math scores between women in the “men perform better than women” conditions and “no mention of differences” conditions) would be significantly greater among women in the “test indicative of math ability” conditions than women in the “test not indicative of math ability” conditions.

$$(\mu_{11}-\mu_{13}) \neq (\mu_{21}-\mu_{23})$$

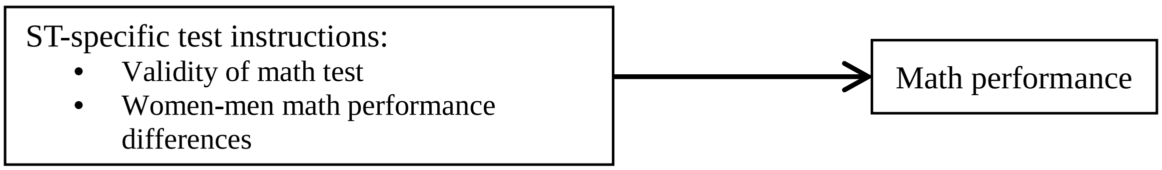

Figure 3. The Interactive Effect of Validity of Math Test and Women-Men Math Differences Instructions on Math performance

H3: This study hypothesized a three-way interaction, such that the difference of the impact of the women-men math performance differences levels between the validity of math test levels (“test indicative of math ability” conditions and the “test not indicative of math ability” conditions) would be greater for women high in identification with math compared to women low in identification with math.

$$([\mu_{11}-\mu_{13}] - [\mu_{21}-\mu_{23}])_{\text{high identification with math}} \neq ([\mu_{11}-\mu_{13}] - [\mu_{21}-\mu_{23}])_{\text{low identification with math}}$$

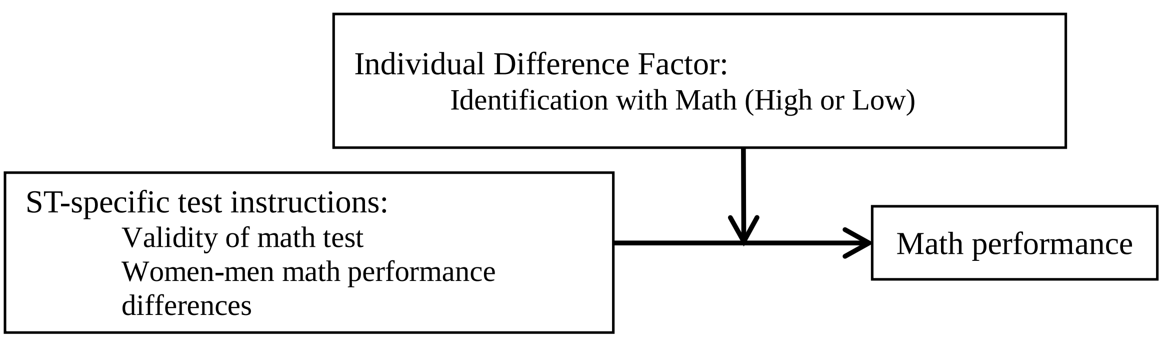

Figure 4. Identification with Math Moderates the Interactive Effect of Validity of Math Test and Women-Men Math Performance Differences Instructions on Math performance

H4: This study hypothesizes a three-way interaction, such that the difference of the impact of the women-men math performance differences levels between the validity of math test levels would be greater for women in math-intensive majors compared to women not in math-intensive majors.

$$([\mu_{11}-\mu_{13}] - [\mu_{21}-\mu_{23}])_{\text{math-intensive majors}} \neq ([\mu_{11}-\mu_{13}] - [\mu_{21}-\mu_{23}])_{\text{not math-intensive majors}}$$

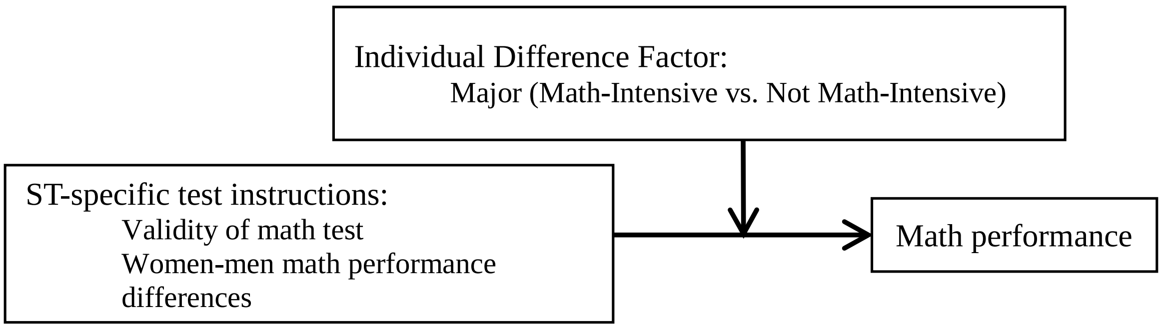

Figure 5. Major Moderates the Interactive Effect of Validity of Math Test and Women-Men Math Performance Differences Instructions on Math performance

H5: This study hypothesized a three-way interaction, such that the impact of the women-men math performance differences levels between women with high identification with math and women with low identification with math would be greater for women in math-intensive majors than for women not in math-intensive majors.

$$([\mu_{.1} \neq \mu_{.3}]_{\text{high identification with math}} - [\mu_{.1} \neq \mu_{.3}]_{\text{low identification with math}})_{\text{math intensive major}}$$

$$\neq$$

$$([\mu_{.1} \neq \mu_{.3}]_{\text{high identification with math}} - [\mu_{.1} \neq \mu_{.3}]_{\text{low identification with math}})_{\text{not math intensive major}}$$

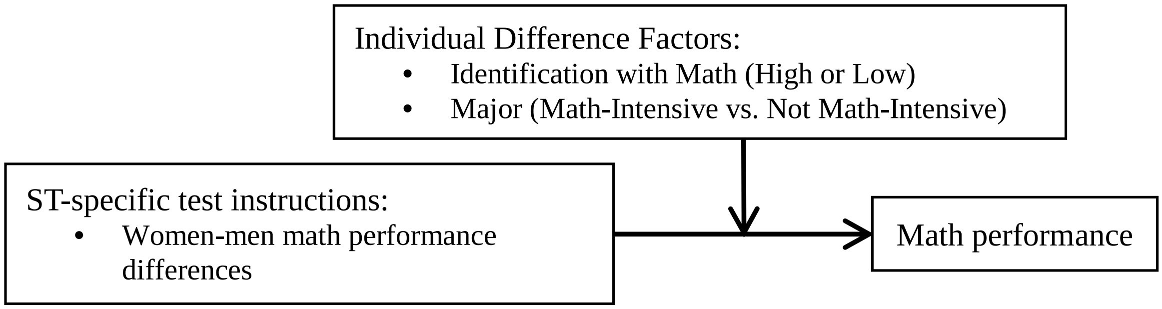

Figure 6. Identification with Math and Major Moderate the Women-Men Math Performance Differences Instruction on Math performance

CHAPTER 3: METHOD

Participants

Participants were undergraduate students recruited from a large Mountain West state university; 231 participants were women in math-intensive majors and 616 were women who were not in math-intensive majors. Consistent with classification models used in prior math ST studies (e.g., Adelman, 1999; Ethington & Wolfle, 1988; Frehill, 1997; Goldman & Hewitt, 1976; Ma, 2011; National Science Foundation, 1994; Paglin & Rufolo, 1990), participants majoring in computer science, economics, engineering, finance, mathematics, physical science, and science/math education were classified as belonging to math-intensive majors. Participants majoring in arts/design, communication studies, history, philosophy, psychology, and sociology, as well as participants reporting their majors as undeclared or undecided, were classified as belonging to majors which were not math-intensive.

Participants ranged in age from 18 to 51 years old, with a mean age of 19.05 (*SD* = 2.19) years. The majority of participants were European American (*n* = 767, 76.5 %). The remaining 23.5 % described themselves as Latina (*n* = 78, 7.4%), African American (*n* = 25, 2.5%), Asian American (*n* = 33, 3.1%), Native American or Pacific Islander (*n* = 10, 1%), and other ethnicities (*n* = 89, 8.5 %). All received psychology research credits for their participation in the study.

Design and Procedure

This study was a 2 (Validity of Math Test: test indicative of math ability, test not indicative of math ability) × 3 (Women-Men Math Performance Differences: men perform better than women, no mention of differences in math performance, women perform better than men) between-participant design. The study's independent variables were two pieces of information embedded within the test instructions. The first piece of information described a math test as an indicator of the student's math ability (indicative of math ability or not indicative of math ability). The second piece of information reported previously observed differences in math performance between women and men (men perform better than women, no mention of differences in math performance, women perform better than men). The levels of these test instructions were fully crossed, to create six experimental conditions (test indicative of math ability and men perform better than women conditions, test indicative of math ability and women perform better than men conditions, test indicative of math ability and no mention of differences in math performance conditions, test not indicative of math ability and men perform better than women conditions, test not indicative of math ability and women perform better than men conditions, test not indicative of math ability and no mention of differences in math performance conditions).The number of questions answered correctly served as this study's dependent variable.

All procedures used in this study were approved by the University Institutional Review Board. Participants completed the study in 2-phases. In Phase 1, they viewed an informed consent document (see appendix A), completed a demographic questionnaire (see appendix B), and the Identification with Math Scale (see appendix C) online. Approximately 1 week (5 to 7 days) later, participants completed the second phase of the study in a computer lab where they were given a math test prefaced with one of the six randomly assigned test instructions. After reading the instructions, participants were verbally asked (see appendix J for the script of instructions) to complete a math test (see appendix K) for 30 minutes and four post-test questions (see appendix L). Finally, participants read an online debriefing statement (see appendix M).

Manipulation of test instructions. All of the study's participants received the following instructions before completing the math test:

> "You are going to work on some math problems and then answer some questions about the problems and about yourself. There are thirty questions or incomplete statements, each followed by five answers. In each case, select the one answer that is the best of the choices offered and then mark it. You will have 30 minutes to complete the math test."

However, additional information was presented before and after these instructions to manipulate the apparent validity of the test as an indicator of ability (adapted from Bell et al., 2003 and Steele & Aronson, 1995), and to suggest a possible differences in performance between women and men (modified from Bell et al., 2003 and Spencer et al., 1999).

The test instructions provided to participants in the *test indicative of math ability* conditions were prefaced with this paragraph:

> "Thank you for participating in this research. We are working in conjunction with the engineering, math, physical science, and behavioral science departments on campus to understand how people persist in engineering, math, physical science, and behavioral science fields. We want to establish profiles of students who show potential for success in these majors. Everything you do today will be at your own pace. It is important that you focus on your test only."

And the test instructions provided to participants in the *test indicative of math ability* conditions were also followed by this paragraph:

> "This test has been shown to be an excellent indicator of one's math aptitude and ability in a large number of settings across a wide spectrum of students, including engineering, math, physical science, and behavioral science majors. This test is especially effective at assessing the limitations in math problem areas of individual engineering, math, physical sciences, and behavioral science students. We ask that you give a strong effort in order to help us in our analysis of your math ability."

These paragraphs were entirely absent in the instruction presented to participants in the *test not indicative of math ability* conditions.

For some participants in the *test indicative of math ability* conditions, the final paragraph embedded an additional sentence. Participants assigned to the *men perform better than women* condition read, “Prior uses of these problems have shown that women have outperformed men on these problems” (see Appendix E), while participants assigned to the wo*men perform better than men* condition read, “Prior uses of these problems have shown that women have outperformed men on these problems” (see Appendix F). No information about performance differences between women and men was inserted into the instructions read by participants in the *no mention of differences in performance* condition (see Appendix D).

The information presented to participants in the *test not indicative of math ability* conditions was substantially shorter. After the instructions, all participants in the *test not indicative of math ability* condition saw the following paragraph:

> “The task you will be working on today was specifically designed to present you with problems varying in their degree of difficulty. We ask that you give a genuine effort and please try hard.”

For some participants in the *test not indicative of math ability* condition, and additional sentence was inserted into the second paragraph, reporting previously observed differences in test performance between women and men. Participants in the *men perform better than women* condition saw, ”Previous results on this test have shown gender differences; that is, men have outperformed women on these problems” (see Appendix H), while participants in the wo*men perform better than men* condition saw, “Previous results on this test have shown that women have outperformed men on these problems” (see Appendix I). Participants in the *no mention of differences in math performance* condition saw no reference to possible differences in math performance between women and men (see Appendix G).

Materials

Pre-test questionnaire. The pre-test questionnaire consisted of questions about age, ethnicity, major, and math classes taken previously.

Identification with Math scale. This 8-item scale measures college students' identification (i.e., personal investment) with math (Smith & White, 2001) and includes items, such as "How important is it to you to be good at Math". Responses are given on a 5-point Likert scale (i.e., 1 = "not at all" to 5 ="very much"). The scale has acceptable levels of internal consistency (Cronbach's α = .93) and test-retest reliability (r = .89) within a period of eight months (Smith & White, 2001). In the current study, the eight items were averaged to form a composite score with higher numbers reflecting stronger identification with math. The scale's internal consistency (Cronbach's α = .90) is comparable to previous studies' (e.g., Keller, 2007).

Math test. The math test used in this study consisted of 30- multiple-choice math questions adapted from the quantitative portion of the Graduate Record Examination's (GRE) practice tests (see appendix M). These questions were used in other ST studies (e.g., Brown & Joseph, 1999; Jamieson & Harkins, 2009; 2012; Johns et al., 2005; Schmader, 2002; Tagler, 2012). Respondents' math performance was represented by the number of math problems answered correctly (α= .70).

Post-test questionnaire. Four questions were completed at the end of the Phase 2 sessions. There were two open-ended questions regarding test difficulty and number of problems they believed they had answered correctly (0-30). In addition, two multiple choice question questions served as a manipulation check. The first question asked participants: "How have women performed relative to men on these problems?" Possible responses for this question were: 1) The test has been shown to be a good indicator of one's math aptitude and ability, 2) The test has been shown to be a poor indicator of one’s math aptitude and ability, 3) No mention of whether the test is a good or poor indicator of one’s math aptitude and ability. The second questions asked participants: “Which of the following is true about the math test you just took?”). Possible responses for this question were: 1) Men have performed better than women, 2) Women and men have performed equally, 3) Women have performed better than men, 4) No mention of how women and men have performed relative to one another.

CHAPTER 4: RESULTS

Sample Demographics by Condition

A Chi Square test was performed on the demographic variables (i.e., major—math-intensive v. non-math intensive and ethnicity) to determine if the sample characteristics differed by condition. The percentage of participants that were assigned to conditions did not differ by major, χ^2 (1, N = 749) = 2.53, p =.77, or by ethnicity, χ^2 (1, N = 730) = 13.43, p = .97.

Manipulation Check

Memory for the validity of math variable. Participants' memory of the content of the test instruction, particularly whether the test they took was described as indicative of their math ability, was first analyzed to access the effectiveness of the ST manipulation. At the end of the study, participants were asked to answer two multiple choice questions concerning what the test instructions had said about the math test they had just completed (i.e., "Which of the following is true about the math test you just took?"). Possible responses were: 1) The test has been shown to be a good indicator of one's math aptitude and ability, 2) The test has been shown to be a poor indicator of one's math aptitude and ability, 3) No mention of whether the test is a good or poor indicator of one's math aptitude and ability. The percentage of women who correctly remembered what the test instructions said is presented in Table 2. A Chi Square test indicated that within each of the Women-Men Math Performance Differences Variable, almost all participants believed that they were in the test indicative of math ability conditions regardless of whether they were or not, χ^2(5, N = 725) = 705.24, p <.001.

Memory for women-men math performance differences variable. Participants were also asked to answer a multiple choice question about what the test instruction said about women's relative performance on the math test (i.e., "How have women performed relative to men on these problems?"). Possible responses were: 1) Men have performed better than women, 2) Women and men have performed equally, 3) Women have performed better than men, 4) No mention of how women and men have performed relative to one another. The percentages of women in each condition who correctly remembered the information provided in the instructions are presented in Table 3. A Chi Square test indicated that there were differences in what participants remembered about how women performed relative to men, $\chi^2(5, N = 733) = 31.61$, p <.001. The analyses suggest that the math ST manipulation may be ineffective. However, examining the influence of the manipulation is still a valid concern.

Main Analyses

Assumptions. Before conducting main analyses, a test of assumption of normality was conducted. According to the Kolmogorov-Smirnov test, the dependent variable, "math correct," was not normally distributed within specific types of test instruction. Assumption of normality was violated for five types of test instructions, except for the one that described the test as indicative of math ability and made no mention of differences in math performance (the control condition), p= .20. Nevertheless, factorial ANOVA was a robust test against violations of normality. A test of the assumption of homogeneity of variance of math performance across test instructions was also conducted. Levene's test of equality of error variances was not significant (p= .48), indicating that the variance in the dependent variable (i.e., math correct) was the same across six conditions and group (i.e., major).

Primary Analyses

Hypotheses regarding the math ST effects via test instructions and the moderating role of Identification with Math and Major on the number of math problems answered correctly were tested in a 2 (Validity of Math Test: test indicative of math ability, test not indicative of math ability) × 3 (Women-Men Math Performance Differences: men perform better than women, no mention of differences in math performance, women perform better than men) × 2 (Major: math-intensive, not math-intensive) × 2 (Identification with Math: high identification with math, low identification with math) factorial analysis of variance. First, as recorded in Table 4, the analysis did not reveal a significant main effect of Women-Men Math Performance Differences Variable. However, follow-up tests were conducted to test hypothesis one: whether women in the "men perform better than women" conditions scored lower on the math test than women in the "no mention of differences" conditions. As recorded in Table 5 and Figure 7, women in the "men perform better than women" conditions did not score significantly lower than women in the "no mention of differences" conditions, $t(710) = -0.57$, 95% *CI* [-1.49, 0.35], $d = -0.11$, inconsistent with hypothesis one.

Second, despite a non-significant interaction effect between the Validity of Math Test and Women-Men Math Performance Differences Variable, follow-up tests were conducted to test hypothesis two: whether the impact of the Women-Men Math Performance Differences Variable (the difference in math scores between women in the "men perform better than women" conditions and "no mention of differences" conditions) were significantly greater among women in the "test indicative of math ability" conditions than women in the "test not indicative of math ability" conditions (see Table 4). Inconsistent with hypothesis two, the impact of the women-men math performance differences instruction on women's math performance was not significantly greater in the "test indicative of math ability" conditions than in the "test not indicative of math ability" conditions, $t(710) = 0.47$, 95% CI [-0.84, 1.78], $d = 0.09$ (see Table 5 and Figure 8).

Third, despite a non-significant 3-way interaction between the Validity of Math Test, Women-Men Math Performance Differences Variable, and Identification with Math, follow-up tests were conducted to test hypothesis three: whether the difference of the impact of the Women-Men Math Performance Differences Variable between the validity of math test condition ("test indicative of math ability" conditions and the "test not indicative of math ability" conditions) will be greater for women high in identification with math compared to women low in identification with math (see Table 5 for descriptive statistics). The analysis revealed a main effect of identification with math, such that women with high identification with math (*Madjusted* = 12.34, *SD* = 5.65) performed better than women with low identification with math (*Madjusted* = 10.54, *SD* = 5.01) (see Table 4). However, follow-up comparison tests did not reveal a significant Validity of Math Test by Women-Men Math Performance Differences by Identification with Math interaction. The analysis indicated that the impact of the Women-Men Math Performance Differences Variable between the validity of math test conditions were not significantly different for women high in identification with math, $t(710) = 0.88$, 95% CI [-0.97, 2.73], $d = 0.17$. Similarly, the impact of ST between the validity of math test conditions were not significantly different for women low in identification with math, $t(710) = 0.05$, 95% CI [-1.80, 1.90], $d = 0.01$. Therefore, inconsistent with hypothesis three, the impact of the Women-Men Math Performance Differences Variable between the validity of math test conditions were not significantly greater for the math performances of women high in identification with math compared to women low in identification with math (see Table 5 and Figure 9).

Fourth, despite a non-significant 3-way interaction between the Validity of Math Test, Women-Men Math Performance Differences Variable and Major, follow-up tests were conducted to test hypothesis four: whether the difference of the impact of the Women-Men Math Performance Differences Variable between the validity of math test conditions ("test indicative of math ability" conditions and the "test not indicative of math ability" conditions) will be greater for women in math-intensive majors compared to women not in math-intensive majors (see Table 5 for descriptive statistics). Follow-up comparison tests did not reveal a significant Women-Men Math Performance Differences by Validity of Math Test by Major interaction. The analysis indicated that the impact of the Women-Men Math Performance Differences Variable between the validity of math test conditions were not significantly different for women in math-intensive majors, $t(710) = -0.08$, 95% CI [-2.61, 2.45], $d = -0.01$. Similarly, the impact of ST between the validity of math test conditions were not significantly different for women not in math-intensive majors, $t(710) = 1.00$, 95% CI [-0.56, 2.56], $d = 0.19$. Therefore, inconsistent with hypothesis four, the impact of the Women-Men Math Performance Differences Variable between the validity of math test conditions were not significantly lower for the math performances of women in math-intensive majors compared to women not in math-intensive majors (see Table 5 and Figure 10).

Finally, despite a non-significant 3-way interaction between the Women-Men Math Performance Differences Variable, Major, and Identification with Math, follow-up tests were conducted to test hypothesis five: whether the difference of the impact of the Women-Men Math Performance Differences Variable will be greater for women in math-intensive majors with high identification with math compared to women in math-intensive majors with low identification with math (see Table 4 for descriptive statistics). Follow-up comparison tests revealed a significant Women-Men Math Differences Variable by Math Identification by Major interaction. Specifically, the impact of ST were significantly greater on the math performances of highly math-identified women in math-intensive majors than women in math-intensive majors who were low in identification with math, $t(710) = -3.11$, 95% CI [-5.84, -0.38], $d = -0.57$. By contrast, the impact of ST were not significantly greater on the math performances of women not in math-intensive majors who were high in identification with math than women not in math-intensive majors who were low in identification with math, $t(710) = -0.54$, 95% CI [-2.11, 1.03], $d = -0.10$ (see Table 5 and Figure 11).

In summary, women who were assigned to the “men perform better than women” conditions did not perform worse than those assigned to the “no mention of differences math performance” conditions, inconsistent with hypothesis one. Inconsistent with hypothesis two, emphasizing the validity of the math test did not contribute to lower score when it is also suggested that men had performed better than women. Inconsistent with hypothesis three, emphasizing the validity of the math test in conjunction with suggesting that men had performed better than women did not contribute to lower score among women high identification with math. Similarly, emphasizing the validity of the math test in conjunction with suggesting that men had performed better than women also did not contribute to lower score among women in math-intensive majors, inconsistent with hypothesis four. However, in support of hypothesis five, suggesting that men had outperformed women on the test negatively impacted the math performance of highly math-identified math-intensive majors.

Additional Analyses

The manipulation check conducted in the preceding section indicated that almost all participants believed that they were in the "test indicative of math ability" conditions (i.e., that the test was a "good indicator of one's math aptitude and ability) regardless of whether they were or not. This suggests that the manipulation may be ineffective. Therefore, the following analyses were conducted independent of the validity of math test condition to determine the effect of the Women-Men Math Performance Differences Variable on women's math performance.

Hypotheses regarding the ST effects via Women-Men Math Performance Differences Variable and the moderating role of Identification with Math and Major on the number of math problems answered correctly were tested in a 3 (Women-Men Math Performance Differences: men perform better than women, no mention of differences in math performance, women perform better than men) x 2 (Major: math-intensive, not math-intensive) x 2 (Identification with Math: high identification with math, low identification with math) factorial analysis of variance. First, as recorded in Table 6, the analysis did not reveal a significant main effect of the Women-Men Math Performance Differences Variable. However, follow-up tests were conducted to test hypothesis one: whether women in the "men perform better than women" condition will score lower on the math test than women in the "no mention of differences" condition. Consistent with the analysis conducted above, but inconsistent with hypothesis one, follow-up analysis revealed that women in the "men perform better than women" condition ($M_{adjusted}$ = 10.80, *SD*= 5.00) did not score significantly lower than women in the "no mention of differences" condition ($M_{adjusted}$ = 11.40, *SD*= 5.40), *t*(722) = -0.60, 95% *CI* [-1.53, 0.33], *d* = -0.11 (see Table 6).

Despite a non-significant 2-way interaction between Women-Men Math Performance Differences Variable and identification with math, follow-up tests were conducted to test hypothesis three: whether the impact of the Women-Men Math Performance Differences Variable (the difference in math scores between women in the "men perform better than women" conditions and "no mention of differences" conditions) will be greater on women high in identification with than women low in identification with math (see Table 4 for descriptive statistics). Consistent with the analysis conducted above, the analysis revealed a main effect of identification with math, such that women with high identification with math (*Madjusted* = 12.35, *SD* = 5.65) performed better than women with low identification with math (*Madjusted* = 10.54, *SD* = 5.01) (see Table 6). Follow-up comparison tests also revealed a significant Women-Men Math Performance Differences by Identification with Math interaction. Consistent with findings from hypothesis three, women high in identification with math in the "men perform better than women" condition scored significantly lower than women high in identification with math in the "no mention of differences" condition. By contrast, women low in identification with math in the "men perform better than women" condition scored equally compared to women low in identification with math in the "no mention of differences" condition, $t(722) = -2.02$, 95% CI [-3.35, -0.69], $d = -0.38$.

Despite a non-significant 2-way interaction between the Women-Men Math Performance Differences Variable and Major, follow-up tests were conducted to test hypothesis four: the impact of the Women-Men Math Performance Differences Variable will be greater on women in math-intensive major than women not in math-intensive majors (see Table 6). Follow-up comparison tests did not reveal a significant Women-Men Math Performance Differences Variable by Major interaction. Inconsistent with findings from hypothesis four, but consistent with findings from the analysis conducted above, women in math-intensive majors in the "men perform better than women" condition did not score significantly better than women in math-intensive major in the “no mention of differences” condition. Women not in math-intensive majors in the “men perform better than women” condition also scored equally compared to women not in math-intensive majors in the “no mention of differences” condition, $t(722) = 0.98$, 95% CI [-0.50, 2.46], $d = 0.19$.

Finally, despite a non-significant 3-way interaction between Women-Men Math Performance Differences Variable, Major, and Identification with Math, follow-up tests were conducted to test hypothesis five: whether the impact of the Women-Men Math Performance Differences Variable between women with high identification with math and women with low identification with math will be greater for women in math-intensive majors compared to women not in math-intensive majors (see Table 6). In support of the hypothesis and consistent with the findings from analysis conducted above, follow-up comparison tests revealed a significant Women-Men Math Performance Differences Variable by Identification with Math by Major interaction. Specifically, women high in identification with math in math-intensive majors in the “men perform better than women” condition scored significantly lower than women in math-intensive majors in the “no mention of differences” condition, $t(722) = -3.25$, 95% CI [-5.47, -1.03], $d = -0.61$. By contrast, women not in math-intensive majors high in identification with math in the “men perform better than women” condition scored equally compared to women not in math-intensive majors low in identification with math in the “no mention of differences” condition, $t(722) = -0.78$, 95% CI [-3.00, 1.44], $d = -0.15$.

Therefore, consistent with the preceding analyses, women who were assigned to the “men perform better than women” condition did not perform worse than those in the “no mention of differences” condition, inconsistent with hypothesis one. However, in support of hypotheses three, women high in identification with math in the “men perform better than women” condition scored significantly lower than those in the “no mention of differences” condition, whereas women low in identification with math’s math scores were not impacted by the test instruction. Consistent with the preceding analysis and findings from hypotheses four, women in math-intensive majors’ math performances were not impacted by the test instruction. Finally, in support of the preceding analysis and findings from hypothesis five, the math performance of women in math-intensive majors who are high in identification with math were negatively impacted by instruction that alleges math underperformance male test takers, while the math performances of math-identified women not in math-intensive majors were not impacted by the instruction.

Exploratory Analyses on the Effect of “Women Perform Better than Men” on Women’s Math Performance

A single variable that coded individuals by their assigned conditions was created (for twelve groups, see Table 7). Then, a one-way ANOVA with *post hoc* comparisons using Tukey HSD was conducted to determine the effects of the “women perform better than men” conditions on number of math problems answered correctly. Two cluster of groups emerged where the collective mean scores of one group were significantly different from the mean scores of another group. As recorded in Table 7, the distinguishing feature of cluster-1 was that it included every group except for the highest scoring group: women in math-intensive major with high identification with math who were in the “no mention of differences” conditions. This suggests that information that alleges math underperformance by male test takers can positively impact the math scores of women low in identification with math.

Cluster-2 consisted of the mean scores of eight groups of women, three of which were not in math-intensive majors. The distinguishing feature of Cluster-2 was that it contained women who were either in the “women perform better than men” conditions, were highly-identified with math, or both. In ascending order, the three women with the highest scores were women in math-intensive major with high identification with math who were in the “women perform better than men” conditions, women not in math-intensive major with high identification with math who were in the “women perform better than men” conditions, and women in math-intensive major with high identification with math who were in the “no mention of differences” conditions. Therefore, suggesting math underperformance by male test takers appeared to enhance the math performances of all women, including women low in identification with math. In addition, while the highest scores were observed in conditions where they were not informed of the performance difference (i.e., those in the “no mention of differences” conditions), the beneficial effects of instruction suggesting male math underperformance were greatest on the math performance of women in math-intensive and not in math-intensive majors with strong identification with math.

CHAPTER 5: DISCUSSION

Despite comparable performance within the classroom, female students generally underperform in high-stakes math examinations relative to male students (Ceci et al., 2009; The College Board, 2011; Willingham & Cole, 1997). ST Theory and research suggest that anxiety produced by the possibility of confirming commonly held negative beliefs about female students' aptitude in mathematics undermines their performance in high-pressure situations (e.g., Johnson et al., 2008; Steele, 1997). Past research has indicated that this phenomenon is exacerbated by instructions which remind test takers of these beliefs (Johnson et al., 2012) or which emphasize the validity of the test's evaluation (Kiefer & Sekaquaptewa, 2007), and that the effects are antagonized by high identification with mathematics (Keller, 2007; Steinberg et al., 2012). This study aimed to replicate and expand upon these findings.

Based on ST theory, this study tested whether women would experience math underperformance when placed in a ST situation that could potentially confirm the prevailing negatives stereotypes about their math ability. Building on past findings (e.g., Bell et al., 2003), this study's first purpose was to evaluate the effects of ST introduced via two types of test instructions upon women's math performances, and to examine the interactive effect between those instructions. First, the test instructions were manipulated so that some emphasized the evaluative nature of the test ("test indicative of math ability"), while others did not ("test not indicative of math ability"). Second, the test instructions were manipulated so that some presented information regarding women's math performance relative to men (women-men math performance differences), either by suggesting underperformance by female test takers ("men perform better than women") or underperformance by male test takers ("women perform better than men”), while some did not mention differences in math performance at all (“no mention of differences in math performance”). The levels of these two factors were fully crossed, to create six experimental conditions.

Further expanding on previous studies suggesting the moderating influences of individual difference factors (e.g., McFarland et al., 2003; Oswald & Harvey, 2000–2001; Schneeberger & Williams, 2003), this study’s second purpose was to determine, whether women’s strength of identification with math, measured directly by the Identification with Math measure (Smith & White, 2001) and indirectly by participants’ chosen major, influences the extent to which women are impacted by these ST-inducing test instructions.

Research Question One: Summary of Findings from the First and Second Hypotheses

This study’s first purpose was to evaluate the effects of ST introduced via two types of test instructions upon women’s math performances, and to examine the interactive effect between those instructions. The first and second hypotheses were designed to answer this question.

First hypothesis. Based on ST theory and past studies, the “men perform better than women” instruction was hypothesized to undermine women’s math performance relative to the “no mention of differences” instruction (e.g., Steinberg et al., 2012). However, inconsistent with past studies’ findings, women in the “men perform better than women” conditions did not score lower than those in the “no mention of differences” conditions. This discrepancy may be due to the content of the “no difference” instructions utilized by the present study and others (e.g., Johnson et al., 2012; Steinberg et al., 2012). For instance, Johnson et al. (2012) explicitly informed women in their study’s control condition that no difference was expected (“neither women nor men are expected to be better or worse”). However, in this study, as in Steinberg and colleagues' study (2012), the lack of difference was not explicitly communicated to women in the "no mention of differences" condition. Participants in the control condition did not receive any information about differences in math performance between women and men, a situation which is more reflective of an instruction presented in an ordinary math testing situation. Although this may have prevented us from supporting our first hypothesis, this approach lends higher external validity to our control conditions. Furthermore, the fact that our hypothesis was not supported may highlight the extent to which the stereotypes of women's math underperformance are embedded in ordinary testing situations. It may be that these young women were so aware of the stereotype when they began the test, or worse, so accepting of the stereotype, that explicitly stating the stereotype-consistent information makes little to no difference in their test-taking experience.

Second hypothesis. Building on ST theory and past research, this study aimed to determine the effects of both types of test instructions, examining a scenario where there can be an interaction between these test instructions, more closely simulating the struggles potentially faced by women in standard math testing situations. No past study had included a "Validity of Math Test" Variable in which a math test was explicitly described as "not indicative of math ability." As such, it had yet to be determined whether a test instruction emphasizing the validity of the math test would impact test performance when the test instructions suggested or did not suggest a comparison of math performances between women and men. Based on past findings (e.g., Bell et al., 2003), it was hypothesized that the difference in performance between women who were assigned to the "men perform better than women" conditions and "no mention of differences" conditions would be greater when the validity of the math test was emphasized. However, this second hypothesis was not supported. This suggests that a test instruction emphasizing the validity of the math test does not necessarily induce additional threat to women's math performances. This is further explored in the "Limitations" and "Findings from Additional Analyses" sections.

Research Question Two: Summary of Findings from the Third, Fourth, and Fifth Hypotheses

The study's second purpose was to increase our understanding of the individual difference factors that account for the relationship between the situational contributors of ST and math underperformance. Specifically, do two individual difference factors (identification with math and academic major) moderate the impact of the aforementioned situational factors (test instructions) on women's math performances? Unlike past studies (Aronson et al., 1999; Keller, 2007; Steinberg et al., 2012), this study's participants were not preselected on the basis of the importance they placed on math (i.e., identification with math) or their past math performances. Therefore, this study was able to determine the moderating effects of identification with math by comparing the effects of the test instructions among women with high and low identification with math.

Third and fourth hypotheses. Our third and fourth hypotheses attempted to explore the interactive effects between the two test instructions and the individual difference factors, identification with math and academic major. Based on past findings condition (e.g., Keller, 2007; Seibt & Forster, 2004) and ST theory (Steele & Aronson, 1995), it was hypothesized that a test instruction emphasizing the validity of the math test would have a greater impact on the math performance of women high in identification with math than women low in identification with math in the "men perform better than women" conditions. Past findings have demonstrated that women who identify strongly with math should become more motivated to disprove the stereotype of math inability in women when reminded of it (e.g., Schmader & Beilock, 2012). Unfortunately, this personal reminder of the stereotype can generate pressure that impairs rather than enhances their math performances (Schmader & Beilock, 2012). On the other hand, based on ST theory and past studies' findings (e.g., Aronson et al., 1999; Keller, 2002; Steinberg et al., 2012; Steele & Aronson, 1995), the performance of women with low identification with math should not be additionally impacted by information regarding women's math performance relative to men. However, inconsistent with hypotheses three, women with high identification with math were not more susceptible to the ST-inducing test instructions. Specifically, the differences in math performance in conditions where women were told and not told that men had performed better was not significantly different for women high or low in identification in math in conditions when the validity of the math test was emphasized than when it was not emphasized.

Based on ST theory and past findings, but unlike the third hypothesis, this study hypothesized that women in math-intensive majors would be less susceptible to the impact of these ST-inducing test instructions than women not in math-intensive majors. Specifically, it was hypothesized that a test instruction emphasizing the validity of the math test would have a lower impact on the math performance of women in math-intensive majors compared to women not in math-intensive majors in the "men perform better than women" conditions (Crisp et al., 2009; Croizet et al., 2004). However, inconsistent with hypotheses four, women in math-intensive majors' were not less susceptible to the influence of the ST-inducing test instructions than women not in math-intensive majors. Specifically, the differences in math performance in conditions where women were told and not told that men had performed better was not significantly different for women in math-intensive and not in math-intensive majors when the validity of the math test was emphasized than when it was not emphasized. Again, the math ST manipulations did not seem to have an effect on the math performances of women with strong identification with math or women in math-intensive majors. The lack of support for hypotheses three and four may be due to several reasons (e.g. the “Validity of Math Test” Variable’s lack of impact), which is revisited later in the “Findings from additional analyses” section.

Fifth hypothesis. This study’s fifth hypothesis tested the interactive effect of the identification with math, major, and the Women-Men Math Performance Differences Variable. Specifically, it was hypothesized that women in math-intensive majors with high identification with math to be more susceptible to test instruction that suggests favorable math performance by men than women in math-intensive majors with low identification with math, whereas the math performances of women not in math-intensive majors should not be affected by the test instruction manipulation. In support of hypothesis five and past findings (Crisp et al., 2009; Croizet et al., 2004; Steinberg et al., 2012; Steele et al., 2002), the differences in math performance in conditions where women were told and not told that men had performed better was significantly different for women in math-intensive with high identification with math, whereas the manipulation did not affect the performance of women not in math-intensive majors. Therefore, consistent with ST theory and past research (Crisp et al., 2009; Croizet et al., 2004; Steinberg et al., 2012; Steele et al., 2002), strong identification with math did significantly increase women’s susceptibility to the “men perform better than women” conditions, but only for women who were actually in math-intensive majors.

Findings from Additional Analyses

The lack of support for hypotheses two, three, and four may be explained by the degree to which the participants accurately recalled the content of the instruction that emphasized the validity of the test. As illustrated in Table 1 and 2, almost all participants in the study seemed to recall the instructions saying that the test was a "good indicator of one's math aptitude and ability" regardless of what condition they were assigned to. This may have to do with the perception that all tests are evaluative, and therefore indicative of ability.

To account for the possibility that the apparent validity of the test had not been successfully manipulated, additional analyses that were independent of the validity manipulation were conducted. In support of the finding for hypothesis one, women in the "men perform better than women" condition did not perform worse. Consistent with the finding above and findings for hypotheses four, women in math-intensive majors in the "men had perform better than women" condition did not score lower than women in math-intensive majors in the "no mention of differences" condition. Additionally, women not in math-intensive majors in the "men had perform better than women" condition scored equally to women not in math-intensive majors in the "no mention of differences" condition. Therefore, women in math-intensive majors were not less susceptible to test instructions that suggest favorable math performances by men than women not in math-intensive majors. However, in partial support of hypotheses three and past findings (e.g., Aronson et al., 1999; Keller, 2002; Steinberg et al., 2012), women with high identification with math in the "men had perform better than women" condition were more likely to perform worse than women with low identification with math in the "no mention of differences" condition. This finding is inconsistent with a meta-analysis that suggests susceptibility to math ST among women who are moderately math-identified relative to women who are highly math-identified (Nguyen & Ryan, 2008). One explanation is that many studies in the meta-analysis were engaged in selection bias that target women with supposedly strong identification with math, which resulted in ceiling effects and regression to the mean. Indeed, current study's finding is consistent with a more recent study's finding (Steinberg et al., 2012) that found greater susceptibility to math ST among women with high identification with math. Additionally, consistent with findings for hypothesis five, these findings suggest that women's self-reported academic major may not be as influential as identification with math when looking at an individual difference factor that may influence women's susceptibility to math ST. For example, approximately 30% of women not in math-intensive majors were undeclared. These individuals could become math-intensive majors later in their academic career, which may influence the validity of the categorization of majors (see "Limitations"). The lack of support for hypothesis four may also be due to the academic major classification criteria used by this study, which is further elaborated in the "Future Research" section.

Findings from Exploratory Analyses

In support of findings from past research examining the effects of the "women perform better than men" conditions on women's math performances (Cadinu et al., 2003; Johnson et al., 2012), findings from this study's exploratory analyses illustrated its beneficial effect on the performance of all women, including women with low identification with math. In particular, its beneficial effect was greater on the math performances of math-identified women, and this was the case for both women in math-intensive and women not in math-intensive majors. Nevertheless, the highest test scores were obtained by math-identified women in math-intensive majors who were not exposed to any ST manipulations supporting findings from the math ST literature (e.g., Steinberg et al., 2012).

Conclusions

In this study, math ST conditions introduced via test instructions did not contribute to women's math underperformance. This was inconsistent with ST theory and past findings (e.g., Johnson et al., 2012; Steele et al., 2002). Consistent with much of past math ST studies, this suggests that the instruction suggesting underperformance by women introduces "threat" into the testing situation without the need for instruction that emphasizes the validity of the math test (e.g., Johnson et al., 2012). This study's methodological limitations may explain this finding, which is revisited later in the "Limitations" section. Inconsistent with past findings (Crisp et al., 2009; Croizet et al., 2004), the present study findings also demonstrates that women in math-intensive majors are not immune to the negative effects of math ST. However, in support of ST theory and past findings (Keller, 2007; Steele et al., 2002; Steinberg et al., 2012), this study found that math ST introduced via test instructions impacts the math performance of women who are most invested in their math performance, those we would expect to be most capable in math: math-intensive majors with high identification with math.

Limitations

A limitation of this study pertains to the content of test instructions that emphasizing the validity of the test's evaluation. As evidenced by a lack of support for hypothesis two, the instruction emphasizing the validity of the test's evaluation is a different and subtler method of math ST inducement than the instruction that presented information regarding women's math performance relative to men. As such, it might have affected aspects of performances that were not assessed by the current study's math test (Stone & McWhinnie, 2008). The instruction emphasizing the validity of the test's evaluation might have affected performance when different math tests that require other types of problem solving strategies were used. In addition, based on an analysis of a post-item question, participants showed poor recall of the instruction emphasizing the validity of the test's evaluation. Therefore, it is a possibility that the instruction might not have successfully conveyed the message that the math test they were completing might not be evaluative of their math abilities. Nevertheless, Brown and Day's (2006) review of the literature indicated that few studies included a post-test questionnaire that assessed the effectiveness of manipulating test instructions (Brown & Day, 2006; see Jamieson & Harkins, 2009; 2012 for exceptions); therefore, it is unclear whether or not this has been an issue in past research. Additionally, although participants did not recall the content of the test instruction after completing the test, those instructions may still have influenced their math performances. Research on priming has established that people do not need to be consciously aware of stimuli to be influenced by it (Bargh, 1996). The content of the test instruction was presented supraliminally, but the processes it influenced may have been non-conscious.

This study's measurement of women's math performance may have limited our ability to detect significant differences in the impact of ST. As evidenced by the low scores received across the conditions, the high difficulty of the test may have produced a floor effect in our measurement of the dependent variable as the low scores restrict variance in the data (across six conditions, scores ranged from 10.24 to 12.36 out of a possible score of 30 points with standard deviations that ranged from 4.94 to 5.79). In particular, this may have contributed to the lack of statistically significant difference between the "men perform better than women" conditions and the "no mention of differences "conditions.

Our categorization of academic major may have had poor discriminant validity, as a measure. Based on past research, academic major should represent the academic environment that the person is immersed in (Logel et al., 2009). However, our definition of not math-intensive major includes participants who have not declared a major and many of these participants may have a math-intensive curriculum and academic environment. Therefore, participants who are classified as being in math-intensive majors are immersed in an academic environment that is focused on mathematics, but, individuals who are classified as being not in a math-intensive major may also exist in that environment. The categorization of major may not adequately discriminate between the academic experiences of the participants, thus reducing the variance between the two groups.

Situational factors in a math-testing environment, independent of the test instruction manipulation might have introduced error variance that obscures the effects of our manipulation. First, we did not control the sex distribution of the room's occupants when administering the test. Second, the characteristics of the experimenter administering the tests were not controlled or otherwise considered. Although the data collection sessions were all led by female experimenters, possible experimenter effects might still have influenced participants' math performance, independent of the test instructions manipulation. Specifically, participants might have responded differently to the test instructions and the math test itself when data collection was led by females of different ethnicities (Armenta, 2010). The experimenters' ethnicities might have also systematically interacted with group size and sex composition to influence women's math performance, independent of the manipulation. As suggested by research regarding the positive effect of in-group role models on women's math performance in math ST situations (e.g., Marx & Roman, 2002; McIntyre et al., 2003), female research experimenters may have represented knowledgeable role models in the math-related fields whose presence refuted the stereotypes in the participants' minds and influence women's math performance independent of the manipulation. Finally, generalization of this study's findings beyond this study may be limited by the use of one type of performance outcome: participants' performance scores on the quantitative section of the Graduate Record Examination, which included a substantial number of problems that require a relatively high-level reading comprehension ability. While this test is commonly used in the math ST research (e.g., Brown & Joseph, 1999; Jamieson & Harkins, 2009; 2012; Johns et al., 2005; Schmader, 2002; Tagler, 2012), the generalizability of this study's findings is limited to the type of math abilities that are assessed by this particular test. Therefore, this study's findings might not represent women's performance in other types of tasks or tests with a different format. For instance, some studies utilized a computerized calculus test where participants were not allowed to work on a math problem once they proceeded to the next problem (Harder, 2000; Steinberg, 2008). In addition, a different pattern of findings may emerge when instead of number of correct, math accuracy scores are chosen as a performance outcome (i.e., number of correct math responses divided by the number of math problems completed; Inzlict & Ben-Zeev, 2000; Johns et al., 2005; Lawrence, Marks, & Jackson, 2010; Schmader & Johns, 2003; Shih et al., 1999). Finally, the ST effects found in this study may not generalize in an actual high-stakes math setting. This study was designed to demonstrate that math ST effects could emerge in a controlled testing environment were women were made aware of the stereotype and its relevance to the math test was explicitly conveyed via test instructions. Therefore, as suggested by Sackett and Ryan (2012), the external generalizability of this study's findings may be dependent on which of this study's experimental manipulation components can be applied in an actual math-testing context.

Future Research

Future research should attempt to determine whether test instruction that describes a math test as indicative of math ability is necessary to induce math ST into a math-testing situation. For instance, future research can include an instruction that more succinctly emphasizes the evaluative nature of the test (e.g., "This test is especially effective at assessing people's engineering limitations in problem areas"; Bell et al., 2003, p. 309). In addition, research can include a control condition in which a math test is characterized as non- evaluative of math ability (e.g., by stating that "We are not interested in your overall score on the test, and, in fact, the problems are in such an early stage of development that we could not say what a particular score would signify"; Bell et al., 2003, p. 309).

Additional research is also needed to determine whether a similar pattern of findings may be found when a different or a more specific classification of academic major is used. While this classification of majors have been used by prior studies (e.g., Adelman, 1999; Ethington & Wolfle, 1988; Frehill, 1997; Goldman & Hewitt, 1976; Ma, 2011; National Science Foundation, 1994; Paglin & Rufolo, 1990), the generalizability of this study's findings may be dependent on this study's particular classification of majors. In addition, the variability within the major clusters used by this study may have obscured the effects of the math ST manipulation and explain a lack of differences in performance among women in math-intensive majors and women not in math-intensive majors. Future research should examine the effects of the math ST manipulation among different and/or more specific clusters of academic major (e.g., all majors within the school of engineering vs. all majors within the school of education or all majors within the school of business vs. all majors within the school of medicine).

Future research should also examine how test instructions that emphasizes the validity of the math test and presents information regarding women's math performance relative to men affect the math performance of men, a non-stereotyped group in math. A meta-analysis has suggested the overall positive effects of instruction that alleges superiority of performance (Walton & Cohen, 2003). However, most ST research has focused on its impact on the performance of stereotyped group members, which means that relatively fewer research has been devoted to examine how instruction that allege in-group members' superiority in math affect positively stereotyped group of individuals (i.e., men). In addition, more research is needed to investigate the moderating role of individual difference factors, such as the role of identification with math on men's math performance in ST-testing situation. This is because like research on the effect of math ST on women's math performance, only a handful of studies have chosen to test the moderating role of identification with math by including men with high and low identification with math (e.g., Smith & Johnson, 2006). Smith and Johnson (2006) found that men low in identification with math "chocked" under pressure when it was suggested that men had outperformed women on a math test, relative to men high in identification with math (Smith & Johnson, 2006). This finding suggests that men who do not personally identify or consider math personally important experienced additional pressure to perform due to the positive performance expectation. However, no research has investigated how instruction that alleges female's superiority in math affects men's math performances. Another potential avenue for research may also include determining other individual difference factors that may moderate the degree to which specific type of ST-inducing test instructions impact performance, such as strength of identification as a woman or as a man.

Implications

These findings suggest the need to educate women who are most motivated to succeed in math-intensive fields about the potential negative effects of math ST. One consequence of underperformance on an important domain is the confirmation of the stereotype. However, a more tangible consequence is an increased likelihood of these women experiencing what Steele, Reisz, Williams, and Kawakami (2002) call “domain disengagement.” These women may psychologically distance themselves from the math-related domain and diminish the importance they place on the domain within their self-concept as a way to protect themselves from future threats. This strategy may result in disidentification (Davies et al., 2002; Major & Schmader, 1998; Major, Spencer, Schmader, Wolfe, & Crocker, 1998), manifested as an increased likelihood of leaving math-intensive fields of study. As women are already under-represented in these fields, field which tends to be more economically prosperous, disengagement and disidentification resulting from ST may sustain and exacerbate a significant source of systemic gender inequality in America. Conversely, findings from the exploratory analysis suggest that women, independent of their strength of identification with math and major, may benefit from test instruction suggesting outgroup members’ (i.e., men’s) potential disadvantage in a math test. Findings regarding the beneficial aspect of the “women perform better than men” conditions may be used to develop intervention programs that can attenuate or reverse the pernicious effects of math ST.

Furthermore, the findings from the exploratory analysis suggest that the least threatening environment for women who have both the motivation and skill to excel as math-intensive majors is one in which no attention is drawn to how women perform relative to men in a math testing situation. However, given the predominance of the stereotype about women’s math ability, it is unlikely that such an environment exist. In addition, past findings suggest that presenting information that indicates a lack of differences in performance may distract high performing women from focusing on the task at hand (Stone & McWhinnie, 2008). While informing women that they perform as well as men can help to nullify the stereotype (Steinberg et al., 2012), it can also convey a mixed message that remind women of the stereotype and can inadvertently activate the ST effects. On the other hand, informing women that they outperform men provides targets of negative stereotypes with a more expansive view of themselves and their group, as well as a positive reappraisal of their math abilities in an otherwise threatening environment. Based on findings from large research literature that supported the general effectiveness of reappraisal for coping with stress (e.g., Ochsner & Gross, 2008), the integrated process of ST model was proposed by Schmader et al. (2008) to explain how reappraisal may reduce ST's negative effect on performance. The model suggests that positive task reappraisal may undercut's the effect of ST by freeing cognitive resources that are otherwise expended to suppress negative emotions (e.g., uncertainty, self-doubts) under ST. Nevertheless, additional research is needed to determine the effect of task reappraisal and the specific circumstances under which the performance of women with the highest motivation and skill is hindered and facilitated in a math-testing situation.

TABLES AND FIGURES

Table 1

Relevant Math ST Studies that have Utilized Women-Men Math Performance Differences

Variable

Women-Men Math Performance Differences Instruction	List of Studies
Mention of Difference	McIntyre et al. (2003)
	Vick et al. (2008)
	Smith and White (2002)
	Rivardo et al. (2011)
Mention of No Differences	Cadinu et al. (2003)
	Steinberg et al. (2012)
Men Perform Better than Women	Cadinu et al. (2003)
	Johnson et al. (2012)
	Keller (2002)
	Kiefer and Sekaquaptewa (2007)
	Rydel et al. (2009)
	Schmader (2002)
	Smith and White (2002)
	Spencer et al. (1999)
	Steinberg et al. (2012)
	Thoman et al. (2008)
No mention of Differences	Johnson et al. (2012)
	Keller (2002)
	Keller and Dauenheimer (2003)
	Rydell et al. (2009, study 1)
	Schmader (2002)
	Smith and White (2002)
	Spencer et al. (1999, study 2 and 3)
	Steinberg et al. (2012)
Mention with Nullification	Keller (2007)
	Keller and Dauenheimer (2003)
	Jamieson and Harkins (2012)
	Keller and Molix (2008)
	O'Brien and Crandall (2003)
	Smith and White (2002)
	Spencer et al. (1999, study 3)
	Vick et al. (2008)

Table 2

Participants' Recall of Test Instructions by Validity of Math Test Variable

		Recalled Validity of Math Test Condition	
		Test Indicative of Math Ability	Test Not Indicative of Math Ability
Assigned Validity of Math Test Condition	Test Indicative of Math Ability	99.4%(342)	0.8%(3)
	Test Not Indicative of Math Ability	99.2%(378)	0.6%(2)

Notes: % refer to percent of correct or incorrect recall by condition, Numbers in parentheses refer to number of participants with correct or incorrect recall by condition

Table 3

Participants' Recall of Test Instructions by Women-Men Math Performance Differences Variable

		Recalled Women-Men Math Performance Differences Conditions		
		Men Perform Better than Women	Women Perform Better than Men	No Mention of Differences
Assigned Women-Men Math Performance Differences Conditions	Men Perform Better than Women	59.7%(145)	20.1%(148)	20.2%(50)
	Women Perform Better than Men	29.5%(70)	54%(129)	16.5%(40)
	No Mention of Differences	28.4%(72)	32.2%(80)	39.4%(99)

Notes: % refer to percent of correct or incorrect recall by condition, Numbers in parentheses refer to number of participants with correct or incorrect recall by condition

Table 4

Analysis of Variance Results for Main Effects and Interaction Effects of Validity of Math Test Condition, Women-Men Math Performance Differences Variable, Major, and Identification with Math

Variable	*df*	*MS*	*F*	*p*	η^2
Main effect of Validity of Math Test (V)	1	35.358	1.276	.259	0.002
Main effect of Women-Men Math Performance Differences (D)	2	66.451	2.399	.092	0.007
Main effect of Major (M)	1	52.413	1.892	.169	0.003
Main effect of Identification with Math (MI)	1	413.404	14.922	.000	0.021
V x D	2	7.364	0.266	.767	0.001
V x M	1	7.299	0.263	.608	0.000
V x MI	1	9.485	0.342	.559	0.000
D x M	2	26.604	0.960	.383	0.003
D x MI	2	39.022	1.409	.245	0.004
M x MI	1	9.404	0.339	.560	0.000
V x D x M	2	14.525	0.524	.592	0.001
V x D x MI	2	8.236	0.297	.743	0.001

V x M x MI	1	9.001	0.325	.569	0.000
D x M x MI	2	50.409	1.820	.163	0.005
V x D x M x MI	2	5.733	0.207	.813	0.001
Within-cells error	710	27.704			

Table 5

Descriptive Statistics for Number Correct as a Function of Conditions

Validity of Math Test Variable	Women-Men Math Performance Differences Variable	Major	Identification with Math	*M*	*SD*	*N*
Test Indicative of Math Ability	Men Perform Better than Women	Math-Intensive	High	11.77	5.39	22
			Low	11.00	4.67	13
			Total	11.49	5.08	35
		Not Math-Intensive	High	12.10	4.88	30
			Low	10.40	4.92	48
			Total	11.05	4.95	78
		Total	High	11.96	5.05	52
			Low	10.52	4.84	61
			Total	11.19	4.97	113
	Women Perform Better than Men	Math-Intensive	High	11.56	4.59	16
			Low	11.92	7.30	13
			Total	11.72	5.84	29
		Not Math-Intensive	High	11.66	5.55	44
			Low	13.44	5.94	46
			Total	12.57	5.79	90
		Total	High	12.95	5.64	62
			Low	11.72	5.92	57
			Total	12.36	5.79	119
	No Mention of Differences	Math-Intensive	High	14.69	5.44	29
			Low	10.50	5.66	8
			Total	13.78	5.68	37
		Not Math-Intensive	High	11.27	3.62	30

Validity of Math Test Variable	Women-Men Math Performance Differences Variable	Major	Identification with Math	*M*	*SD*	*N*
			Low	10.18	4.53	50
			Total	10.59	4.22	80
		Total	High	12.95	4.88	59
			Low	10.22	4.65	58
			Total	11.60	4.94	117
	Total	Math-Intensive	High	13.05	5.36	136
			Low	10.78	5.47	65
			Total	12.32	5.49	201
		Not Math-Intensive	High	11.85	5.74	232
			Low	10.32	4.82	301
			Total	10.99	5.29	533
		Total	High	12.29	5.62	368
			Low	10.40	4.94	366
			Total	11.35	5.37	734
Test Not Indicative of Math Ability	Men Perform Better than Women	Math-Intensive	High	11.47	4.72	17
			Low	9.80	3.77	10
			Total	10.85	4.39	27
		Not Math-Intensive	High	10.18	5.70	51
			Low	10.00	4.76	56
			Total	10.08	5.20	107
		Total	High	10.50	5.47	68
			Low	9.97	4.59	66
			Total	10.24	5.04	134
	Women Perform Better than Men	Math-Intensive	High	13.31	5.07	26
			Low	11.60	7.01	10
			Total	12.83	5.62	36
		Not Math-Intensive	High	12.86	6.20	36

Validity of Math Test Variable	Women-Men Math Performance Differences Variable	Major	Identification with Math	*M*	*SD*	*N*
			Low	10.29	4.59	52
			Total	11.34	5.43	88
		Total	High	13.05	5.72	62
			Low	10.50	5.01	62
			Total	11.77	5.50	124
	No Mention of Differences	Math-Intensive	High	14.00	6.02	26
			Low	9.55	4.11	11
			Total	12.68	5.84	37
		Not Math-Intensive	High	11.49	6.63	39
			Low	9.63	4.61	51
			Total	10.43	5.62	90
		Total	High	12.49	6.46	65
			Low	9.61	4.50	62
			Total	11.09	5.75	127
	Total	Math-Intensive	High	13.12	5.39	69
			Low	10.29	5.05	31
			Total	12.24	5.42	100
		Not Math-Intensive	High	11.35	6.19	126
			Low	9.97	4.64	159
			Total	10.58	5.41	285
		Total	High	11.97	5.97	195
			Low	10.03	4.69	190
			Total	11.01	5.46	385
Total	Men Perform Better than Women	Math-Intensive	High	11.64	5.04	39
			Low	10.48	4.25	23
			Total	11.21	4.76	62
		Not Math-Intensive	High	10.89	5.46	81

Validity of Math Test Variable	Women-Men Math Performance Differences Variable	Major	Identification with Math	*M*	*SD*	*N*
			Low	10.18	4.82	104
			Total	10.49	5.11	185
		Total	High	11.13	5.32	120
			Low	10.24	4.70	127
			Total	10.67	5.02	247
	Women Perform Better than Men	Math-Intensive	High	12.64	4.91	42
			Low	11.78	7.01	23
			Total	12.34	5.70	65
		Not Math-Intensive	High	13.18	6.02	82
			Low	10.92	5.07	96
			Total	11.96	5.63	178
		Total	High	13.00	5.66	124
			Low	11.08	5.48	119
			Total	12.06	5.64	243
	No Mention of Differences	Math-Intensive	High	14.36	5.68	55
			Low	9.95	4.70	19
			Total	13.23	5.75	74
		Not Math-Intensive	High	11.39	5.49	69
			Low	9.90	4.56	101
			Total	10.51	5.00	170
		Total	High	12.71	5.75	124
			Low	9.91	4.56	120
			Total	11.33	5.37	244
	Total	Math-Intensive	High	13.05	5.36	136
			Low	10.78	5.47	65
			Total	12.32	5.49	201
		Not Math-Intensive	High	11.85	5.74	232

Validity of Math Test Variable	Women-Men Math Performance Differences Variable	Major	Identification with Math	*M*	*SD*	*N*
			Low	10.32	4.85	301
			Total	10.99	5.29	533
		Total	High	12.29	5.62	368
			Low	10.40	4.94	366
			Total	11.35	5.37	734

Notes: M = Mean, *SD* = Standard Deviation, *N* = Sample Size

Table 6

Analysis of Variance Results for Main Effects and Interaction Effects of Women-Men Math Performance Differences Variable, Major, and Identification with Math

Condition	*df*	*MS*	*F*	*p*	η^2
Main effect of Women-Men Math Performance Differences (D)	2	78.873	2.868	.057	0.008
Main effect of Major (M)	1	69.852	2.540	.111	0.004
Main effect of Identification with Math (MI)	1	430.341	15.649	.000	0.021
SD x M	2	20.837	0.758	.469	0.002
D x MI	2	45.637	1.660	.191	0.005
M x MI	1	14.145	0.514	.473	0.001
D x M x MI	2	20.837	0.758	.469	0.002
D x M x MI	2	14.145	0.514	.473	0.001
Within-cells error	722	27.499			

Table 7

Average Math Scores as a Function of Major, Women-Men Math Performance Differences Variable, and Identification with Math

		Subset	
	N	1	2
Not Math-Intensive No Mention of Differences Low Identification with Math	101	9.90	
Math-Intensive No Mention of Differences Low Identification with Math	19	9.95	
Not Math-Intensive Men Perform Better than Women Low Identification with Math	104	10.18	
Math-Intensive Men Perform Better than Women High Identification with Math	23	10.48	
Not Math-Intensive Men Perform Better than Women High Identification with Math	81	10.89	10.89
Not Math-Intensive Women Perform Better than Men Low Identification with Math	96	10.92	10.92
Math-Intensive No Mention of Differences High Identification with Math	69	11.39	11.39
Math-Intensive Men Perform Better than Women High Identification with Math	39	11.64	11.64
Math-Intensive Women Perform Better than Men Low Identification with Math	23	10.78	10.78

Math-Intensive Women Perform Better than Men High Identification with Math	42	12.64	12.64
Not Math-Intensive Women Perform Better than Men High Identification with Math	82	13.18	13.18
Math-Intensive No Mention of Differences High Identification with Math	55		14.36

Notes: M = Mean, *N* = Sample Size

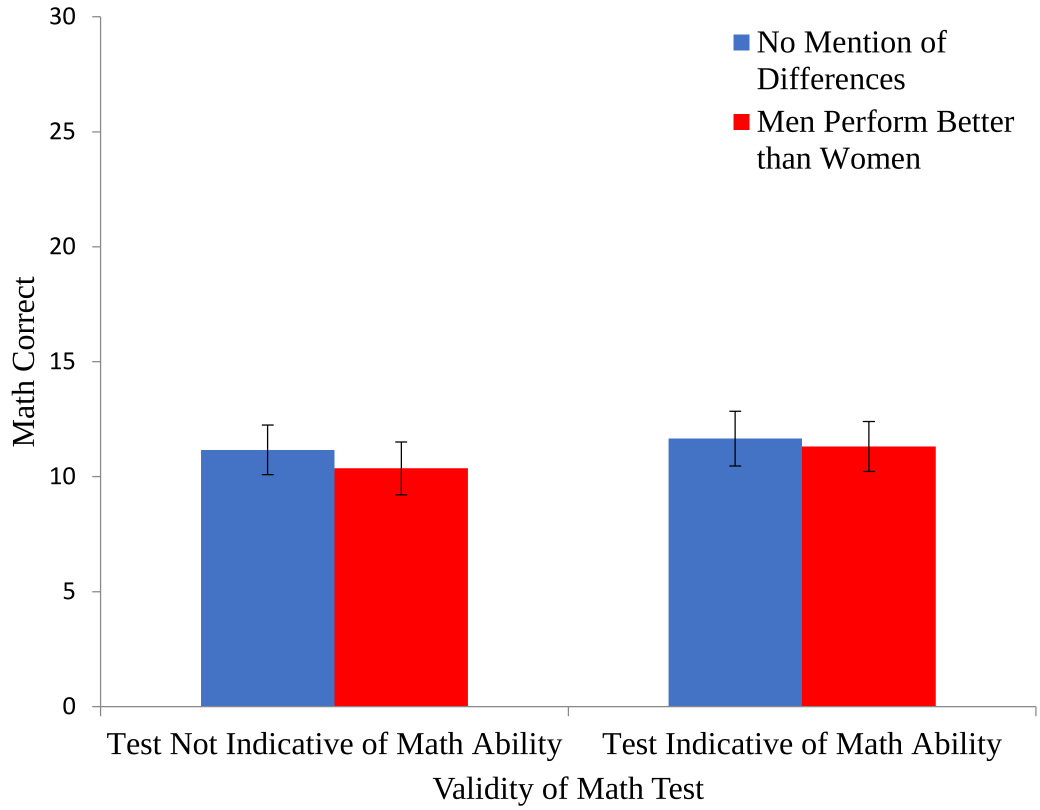

Figure 7. Average Math Scores as a Function of Women-Men Math Performance Differences Variable

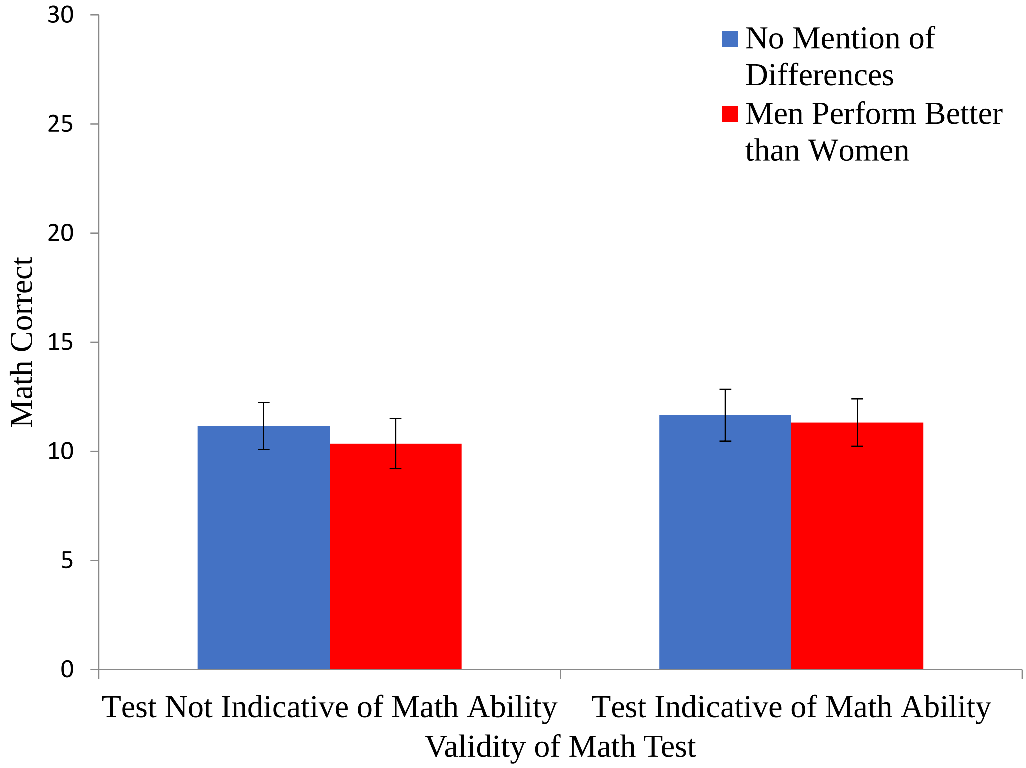

Figure 8. Average Math Scores as a Function of Validity of Math Test and Women-Men Math Performance Differences Variables

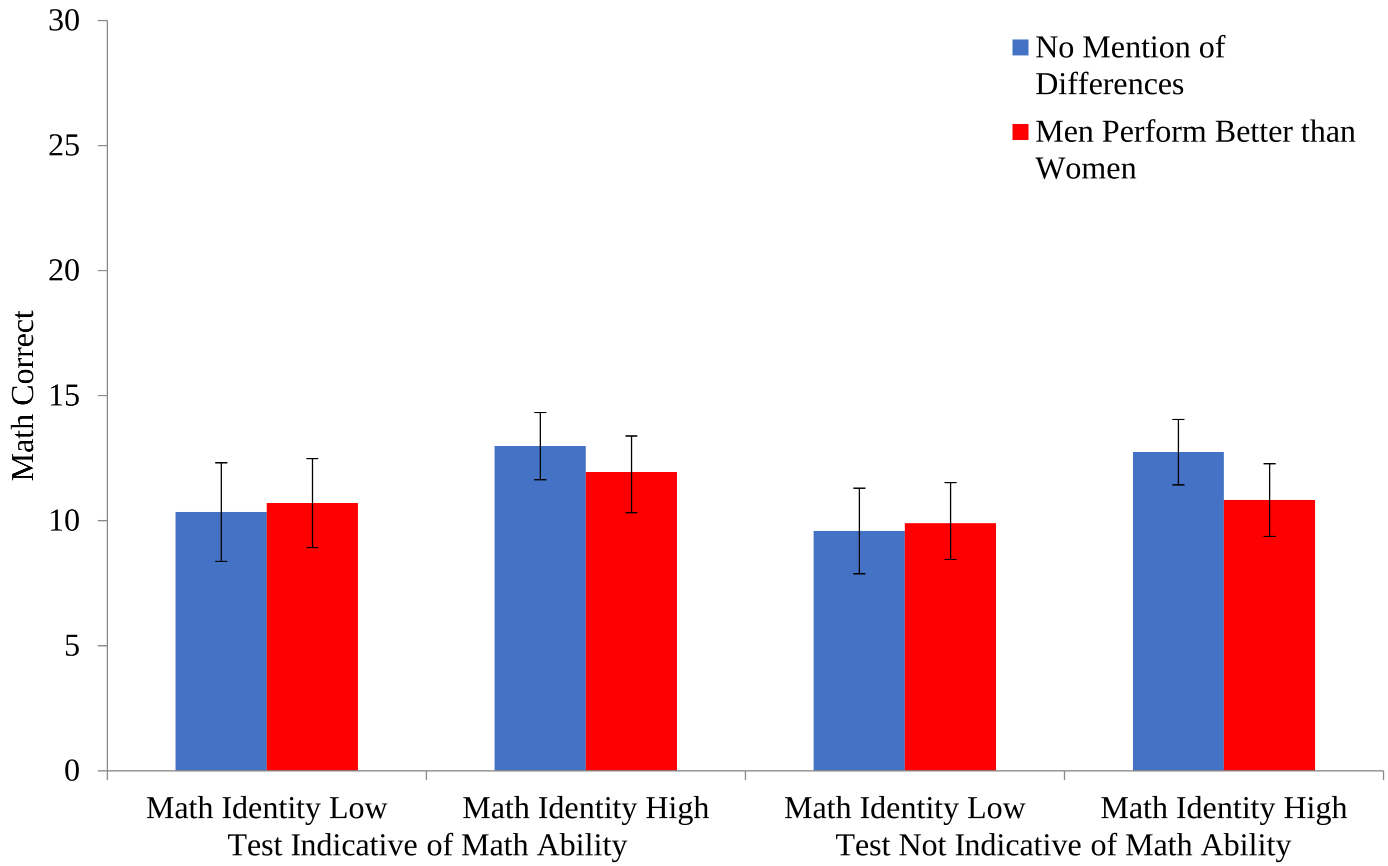

Figure 9. Average Math Scores as a Function of Validity of Math Test Condition, Women-Men Math Performance Differences Variable, and Identification with Math

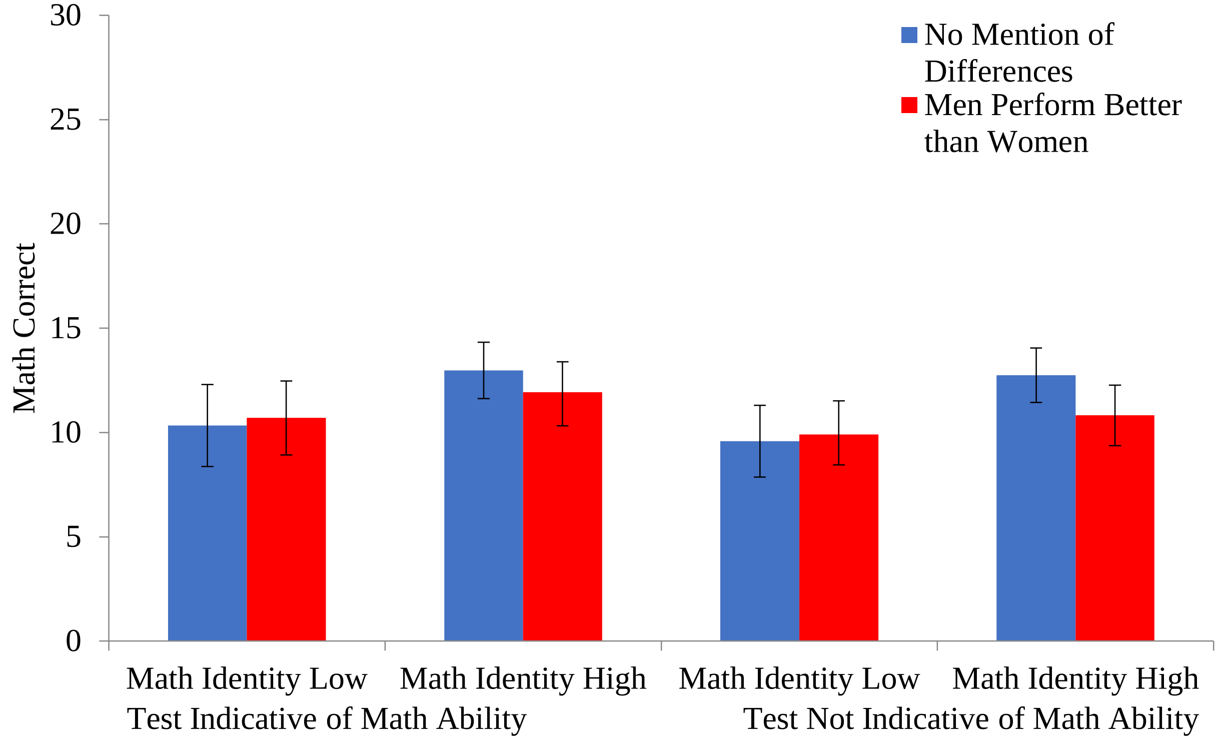

Figure 10. Average Math Scores as a Function of Validity of Math Test Condition, Women-Men Math Performance Differences Variable, and Major

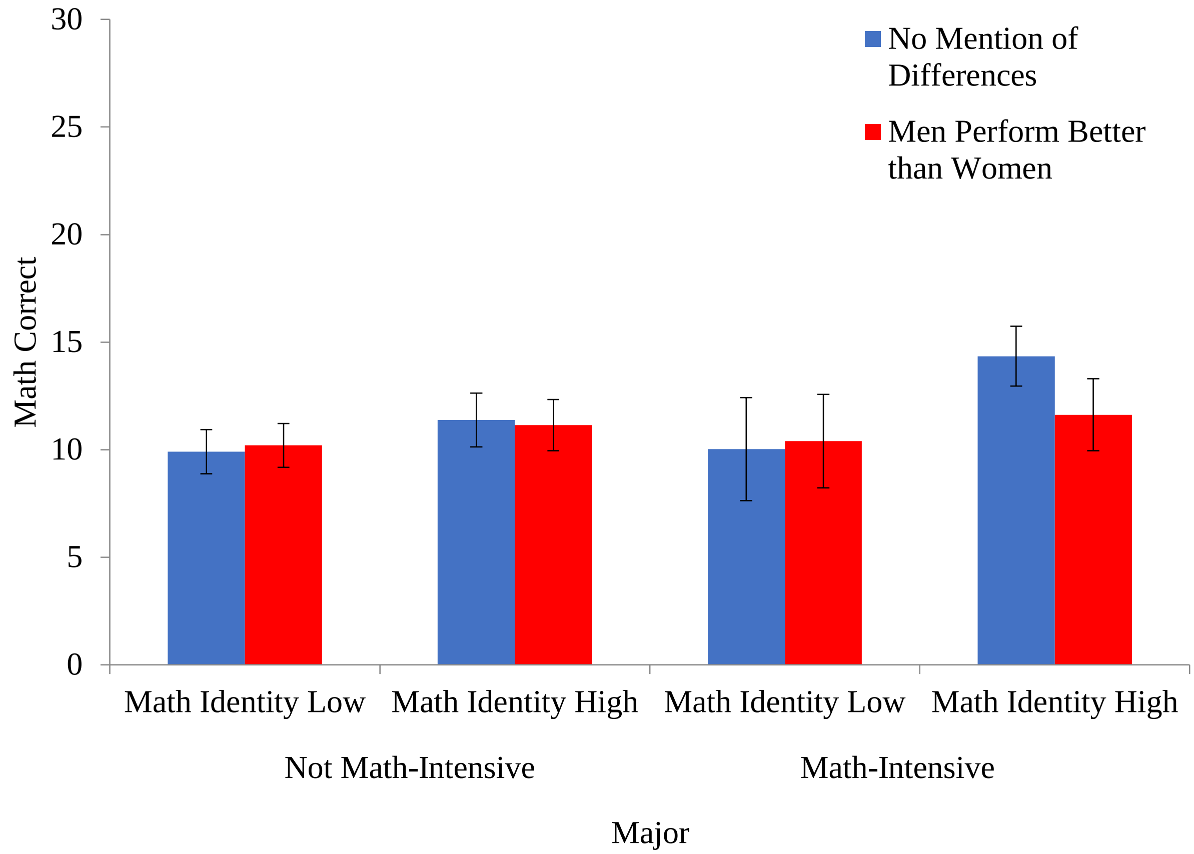

Figure 11. Average Math Scores as a Function of Women-Men Math Performance Differences Variable, Major, and Identification with Math

REFERENCES

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education.

Ambady, N., Shih, M., Stephanie, A., & Pittinsky, T.L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*, 385–390. doi:10.1111/1467-9280.00371

Appel, M., Kronberger, N., & Aronson, J. (2011). Stereotype threat impairs ability building: Effects on test preparation among women in science and technology. *European Journal of Social Psychology, 41*, 904–913. doi: 10.1002/ejsp.835

Aronson, J., Quinn, D. M., & Spencer, S. J. (1999). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 83-103). New York: Academic Press.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype priming on action. *Journal of Personality and Social Psychology*, *71*, 230-244. doi: 10.1037/0022-3514.71.2.230

Beilock, S. L., McConnell, A. R., & Rydell, R. J. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136,* 256–276. doi: 10.1037/0096-3445.136.2.256

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science, 17,* 339-343. doi: 10.1111/j.1467-8721.2008.00602

Bell, A. E., Spencer, S. J., Iserman, E., & Logel, C. E. R. (2003, October). Stereotype threat and women's performance in engineering. *Journal of Engineering Education*, 307-312. doi:10.1037/0096-3445.136.2.256

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*, 174–181. doi:10.1016/j.jesp. 2003.11.007.

Bentler, P. M. (1999). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246. doi: 10.1037/0033-2909.107.2.238

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology, 40,* 247-255. doi:10.1016/S0022-1031(03)00099-4

Brodish, A. B., & Devine, P. G. (2009). The role of performance–avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology, 45,* 180-185. doi:10.1016/j.jesp.2008.08.005

Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on raven's advanced progressive matrices. *Journal of Applied Psychology, 91*, 979-985. doi: 10.1037/0021-9010.91.4.979

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology, 39*, 626–633. doi:10.1016/S0022-1031(03)00039-8

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33,* 267–285. doi: 10.1002/ejsp.145

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 16*, 572-578. doi: 10.1111/j.0956-7976.2005.01577.x

Campbell, S. M., & Collaer, M.L. (2009). Stereotype threat an gender differences in performance on a novel visuospatial task. *Psychology of Women Quarterly, 33,* 437-444. doi: 10.1111/j.1471-6402.2009.01521.x

Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*, 853–859. doi:10.1016/j.jesp.2009.03.003

Ceci, S. J., Williams, W.M., & Barnett, S.M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135*, 218-261. doi: 10.1037/a0014412

Chalabaev, A., Stone, J., Sarrazin, P., & Croizet, J. (2008). Investigating physiological and self-reported mediators of stereotype lift effects on a motor task. *Basic & Applied Social Psychology*, *30*, 18–26. doi: 10.1080/01973530701665256

Crisp, R. J., Bache, L M., & Maitner, A. T. (2009). Dynamics of social comparison in counter-stereotype domains: Stereotype boost, not stereotype threat for women in engineering majors. *Social Influence, 4,* 171-184. doi: 10.1080/15534510802607953

Croizet, J., & 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*, 588-594. doi: 10.1177/0146167298246003

Croizet, J., Despres, G., Gauzins, M., Huguet, P., Leyens, J., & Meot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30,* 721-731. doi: 10.1177/0146167204263961

Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology, 89,* 220-230. doi: 10.1037/0021-9010.89.2.220

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28,* 1615-1628. doi: 10.1177/014616702237644

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23. doi: 10.1111/1467-8721.00160

Ethington, C. A., & Wolfle, L. M. (1988).Women's selection of quantitative undergraduate fields of study. *American Educational Research Journal, 25*, 157–175. doi: 10.2307/1163078

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University

Frehill, L. M. (1997). Education and occupational sex segregation: The decision to major in engineering. *Sociological Quarterly, 38*, 225–249. doi: 10.1111/j.1533-8525.1997.tb00475.x

Gibson, C. E., Losee, J., & Vitiello, C. (2014). A replication attempt of stereotype susceptibility (Shih, Pittinsky, & Ambady, 1999). Identity salience and shifts in quantitative performance. *Social Psychology, 45,* 194-198. doi: 10.1027/1864-9335/a000184

Goclowska, M.A., Crisp, R.J., & Labuschagne, K. (2012). Can counter-stereotypes boost flexible thinking? *Group Processes and Intergroup Relations, 16*, 217-231. doi: 10.1177/1368430212445076

Goldman, R. D., & Hewitt, B. N. (1976). The scholastic aptitude tests "explain" why college men major in science more often than college women. *Journal of Counseling Psychology*, *23,* 50–54. doi: 10.1037/0022-0167.23.1.50

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin, 28*, 659–670. doi: 10.1177/0146167202288010

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*, 17-28. doi:10.1016/j.appdev.2007.10.004.

Harder, J. A. (2000). *The effect of private versus public evaluation on stereotype threat for women in mathematics*. (Doctoral Dissertation, University of Texas, Austin, 1999). Dissertation Abstracts International, Section A: Humanities and Social Sciences, 609 (9-A), 3270.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. doi: 10.1080/10705519909540118

Hyde, J. S., & Kling, K. C. (2001).Women, motivation, and achievement. *Psychology of Women Quarterly, 25,* 364–378. doi: 10.1111/1471-6402.00035

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,* 365–371. doi:10.1111/1467-9280.00272

Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95,* 796–805. doi: 10.1037/0022-0663.95.4.796

Inzlicht, M. & Good, C. (2006). How environments threaten academic performance, self-knowledge, and sense of belonging. In S. Levin & C. van Laar (Eds.), *Stigma and Group Inequality: Social Psychological Approaches* (pp. 129-150). Mahwah, NJ: Erlbaum.

Inzlicht , M., & Kang, S. K. (2010) . Stereotype threat spillover: How coping with threats to social identity affects aggression, eating, decision making, and attention. *Journal of Personality and Social Psychology, 99,* 467-481. doi: 10.1037/a0018951

Inzlicht, M., & Schmader, T. (Eds.). (2012). *Stereotype threat: Theory, process, and application*. Oxford: Oxford University Press.

Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin, 35,* 1301-1314.

Jamieson, J. P., & Harkins, S. G. (2012). Distinguishing between the effects of stereotype priming and stereotype threat on math performance. *Group Processes and Intergroup Relations, 15,* 291-304. doi: 10.1177/1368430211417833

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General, 137,* 691-705. doi: 10.1037/a0013834

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16,* 175–179. doi:10.1111/j.0956-7976.2005.00799.x

Johns, M., & Schmader, T. (2010). Meta-cognitive regulation as a reaction to the uncertainty of stereotype threat. In R.M. Arkin, K.C. Oleson, & P.J. Carroll (Eds.). *The uncertain self: A handbook of perspectives from social and personality psychology* (pp. 176-192). Lawrence Erlbaum Associates.

Johnson, H. J., Bernard-Brak, L., Saxon, T. F., & Johnson, M. K. (2012). An experimental study of the effects of stereotype threat and stereotype lift on men and women's performance in mathematics. *The Journal of Experimental Education, 80,* 137–149. doi: 10.1080/00220973.2011.567312

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 47,* 193–198. doi: 10.1023/A:1021003307511

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology, 77,* 323-338. doi:10.1348/000709906X113662

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29,* 371–381. doi:10.1177/0146167202250218.

Keller, J. & Molix, L. (2008). When women can't do math: The interplay of self-construal, group identification, and stereotypic performance standards. *Journal of Experimental Social Psychology, 44,* 437–444. doi:10.1016/j.jesp.2007.01.007

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology, 43,* 825-832. doi: 10.1016/j.jesp.2006.08.004

Lawrence, J. S., Marks, B. T., & Jackson, J. S. (2010). Domain identification predicts black students' underperformance on moderately-difficult tests. *Motivation and Emotion, 34,* 105-109. doi: 10.1007/s11031-010-9159-8

Lewin K. (1943). Defining the "Field at a Given Time." *Psychological Review.* 50: 292-310. Republished in *Resolving Social Conflicts & Field Theory in Social Science,* Washington, D.C.: American Psychological Association, 1997.

Leyens, J.-P., Yzerbyt, V., & Schadron, G. (1994). *Stereotypes and social cognition.* London: Sage.

Logel, C., Iserman, E. C., Davies, P. G., Quinn, D. M., & Spencer, S. J. (2009). The perils of double consciousness: The role of thought suppression in stereotype threat. *Journal of Experimental Social Psychology, 45,* 299–312. doi:10.1016/j.jesp.2008.07.016

Ma, Y. (2011). College major choice, occupational structure and demographic patterning by gender, race and nativity. *The Social Science Journal 48,* 112–129. doi:10.1016/j.soscij.2010.05.004

Martens, A., Johns, M., Greenberg, F., & Schimel, J. (2006) Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42,* 236–243. doi:10.1016/j.jesp.2005.04.010

Major, B., & Schmader, T. (1998).Coping with stigma through psychological disengagement. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective*, (pp. 219-214). San Diego, CA: Academic Press

Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin, 24*, 34-50. doi: 10.1177/0146167298241003

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting female students math test performance. *Personality and Social Psychology Bulletin, 28,* 1185–1197. doi: 10.1177/01461672022812004

McFarland, L. A., Kemp, C. F., Viera, L., Jr., & Odin, E. P. (2003, April). *Stereotype threat and male–female differences in test performance.* Paper presented at the 18th annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype through salience of group achievement. *Journal of Experimental Social Psychology, 39,* 83–90.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18,* 879-885. doi:10.1111/j.1467-9280.2007.01995.x

National Science Foundation (1994). *Women, minorities, and persons with disabilities in science and engineering, 1994.* Arlington, VA: NSF 94-333.

Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93,* 1314-1334. doi: 10.1037/a0012702

Nussbaum, D., A., & Steele, C. M. (2007). Situational disengagement and persistence in the face of adversity. *Journal of Experimental Social Psychology, 43,* 127-134. doi:10.1016/j.jesp.2005.12.007

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29,* 782–789. doi:10.1177/0146167203029006010

Ochsner, K., & Gross, J. (2008). Cognitive emotion regulation: Insights from social cognitive and affective neuroscience. *Current Direction in Psychological Science, 17,* 153-158. doi: 2048/10.1111/j.1467-8721.2008.00566.x

Osborne, J. W. (2006). Gender, stereotype threat and anxiety: Psychophysiological and cognitive evidence. *Journal of Research in Educational Psychology, 8,* 109–138.

Osborne, J. W. (2007). Linking stereotype threat and anxiety. *Educational Psychology, 27,* 135–154. doi: 10.1080/01443410601069929

Oswald, D. L., & Harvey, R. D. (2000 –2001). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology: Developmental, Learning, Personality, Social, 19,* 338–356. doi: 10.1007/s12144-000-1025-5

Paglin, M., & Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male–female earnings differences. *Journal of Labor Economics, 8,* 123–144.

Palumbo, M. V., & Steele-Johnson, D. (2014). Do test perceptions influence test performance? Exploring stereotype threat theory. *North American Journal of Psychology, 16,* 1-12.

Picho, K., Rodriquez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of Social Psychology, 153,* 299-333. doi: 10.1080/00224545.2012.737380

Quinn, D. M., & Spencer, S. J. (2001). The Interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57,* 55-71. doi:10.1111/0022-4537.00201

Renninger, K.A. (2000). How might the development of individual interest contribute to the conceptualization of intrinsic motivation? In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 375–407). New York: Academic Press.

Rivardo, M. G., Rhodes, M. E., Camaione, T. C., & Legg, J. M. (2011). Stereotype threat leads to reduction in number of math problems women attempt. *North American Journal of Psychology, 13,* 5-16.

Rydell, B. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology, 96,* 949–966. doi:10.1037/a0014846.

Sackett, P. R., & Ryan, A. M. (2012). Concerns about generalizing stereotype threat research findings to operational high-stakes testing. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 249-263). New York: Oxford University Press.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38,* 194–201. doi:10.1006/jesp.2001.1500.

Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science, 19*, 14-18. doi: 10.1177/0963721409359292

Schmader, T., & Beilock, S. (2012). An integration of processes that underlie stereotype threat. In M. Inzlicht and T. Schmader (Eds.), *Stereotype threat: Theory, Process, and Application*. Oxford University Press.

Schmader, T., Forbes, C.E., Zhang, S., & Mendes, W.B. (2009). A meta-cognitive perspective on cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin, 35,* 584-596. doi: 10.1177/0146167208330450

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85,* 440-452. doi:10.1037/0022-3514.85.3.440

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115,* 336-356. doi: 10.1037/0033-295X.115.2.336

Schneeberger, N. A., & Williams, K. (2003, April). *Why women "can't" do math: The role of cognitive load in stereotype threat research.* Paper presented at the 18th meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.

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,* 38-56. doi: 10.1037/0022-3514.87.1.38

Sekaquaptewa, D., & Thompson, M. (2002). The differential effects of solo status on members of high and low status groups. *Personality and Social Psychology Bulletin, 28,* 694–707. doi: 10.1177/0146167202288013

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology, 39,* 68–74. doi: 10.1016/S0022-1031(02)00508-5

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10,* 80-83. doi:10.1111/1467-9280.00111.

Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review, 16,* 177–206. doi: 1040-726X/04/0900-0177/0

Smith, C. S., & Hung, L. C. (2008). Stereotype threat: Effects on education. *Social Psychological Education, 11,* 243-257. doi: 10.1007/s11218-008-9053-3

Smith, J. L., & Johnson, C. S. (2006). A stereotype boost or choking under pressure? Positive gender stereotypes and men who are low in domain identification. *Basic and Applied Social Psychology, 28,* 51‑63. doi: 10.1207/s15324834basp2801_5

Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology, 99,* 99-114. doi: 10.1037/0022-0663.99.1.99

Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61,* 1040–1057. doi: 10.1177/00131640121971635

Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated and nullified stereotypes on mathematical performance: It's not just a female's issue. *Sex Roles, 47,* 179–191. doi: 10.1023/A:1021051223441

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28. doi:10.1006/jesp.1998.1373

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629. doi:10.1037/0003-066X.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology, 69*, 797-811. doi:10.1037/0022-3514.69.5.797

Steele, J., James, J. B., & Barnett, R. C. (2002). Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. *Psychology of Women Quarterly, 26*, 46–50. doi:10.1111/1471-6402.00042

Steele, J. R., Reisz, L., Williams, A., & Kawakami, K. (2007). Women in mathematics: Examining the hidden barriers that gender stereotypes can impose. In R. Burke & M. Mattis (Eds.), *Women and minorities in science, technology, engineering, and mathematics: Upping the numbers* (pp. 159-183). London: Edward Elgar.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with images of one's group: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), A*dvances in Experimental Social Psychology* (pp. 379–440). San Diego: Academic Press.

Steiger, J. H., & Lind, J. M. (1980, June). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Steinberg, J. R. (2008). *Stereotype threat in persistent women.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. 3304888).

Steinberg, J. R., Okun, M.A., & Aiken, L.S. (2012). Calculus GPA and math identification as moderators of stereotype threat in highly persistent women. *Basic and Applied Social Psychology, 34,* 534-543. doi: 10.1080/01973533.2012.727319

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology, 77,* 1213–1227. doi: 10.1037/0022-3514.77.6.1213

Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology, 44*, 445-452. doi:10.1016/j.jesp.2007.02.006

Tagler, M. J. (2012). Choking under the pressure of a positive stereotype: Gender identification and self-consciousness moderate men's math test performance. *The Journal of Social Psychology, 152,* 401-416. doi: 10.1080/00224545.2011.615353

The College Board. (2011). *College bound seniors national profile reports of SAT program test takers.*

Thoman, D. B., White, P. H., Yamawaki, N., & Koishi, H. (2008). Variations of gender-math stereotype content affect women's vulnerability to stereotype threat. *Sex Roles, 58,* 702-712. doi:10.1007/s11199-008-9390-x

Vick, S.B., Seery, M. D., Blascovich, J., & Weisbuch, M. (2008). The effect of gender stereotype activation on challenge and threat motivational states. *Journal of Experimental Social Psychology*, *44,* 624-630. doi:10.1016/j.jesp.2007.02.007

Walton, G., & Cohen, G. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456–467. doi:10.1016/S0022-1031(03)00019-2

Willingham, W.W., & Cole, N. S. (1997). *Gender and fair assessment.* Mahwah: Lawrence Erlbaum.

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology, 44,* 792-799. doi :10.1016/j.jesp.2007.07.005

Wout, D., Shih, M., Jackson, J. S., & Sellers, R. M. (2009). Targets as perceivers: How people determine when they will be negatively stereotyped. *Journal of Personality and Social Psychology, 96,* 349-362. doi: 10.1037/a0012880

APPENDIX A

Information about the Study

PURPOSE	This study is designed to explore whether women are susceptible to stereotype threat under specific conditions.
PROCEDURES	Participation in this research should take approximately one and a half hours over the course of two data collections. If you should agree to participate, we will ask you to provide some information about your experience with math and math classes. We will also ask you to take a math test and reflect on your thoughts and feelings while taking the test. In this study, you will be asked to answer some questions about your thoughts, feelings, attitudes, and behaviors. Additionally, we will ask you to complete a brief demographic questionnaire, which will gather information such as age, sex, and ethnicity. This information will not be used, in any way, to determine your identity. Instead, it will be used to describe the composition of our sample when reporting the data. Should you agree to participate, you will earn 1.5 research credits.
RISKS	Risks are minimal. Participation is voluntary and can be terminated at for any reason at any time, without consequence. It is not possible to identify all potential risks in research procedures, but the researcher has taken responsible safeguards to minimize any potential, but unknown risks.
BENEFITS	There are no known benefits to participating in this research. You may learn more about how psychological research is conducted and learn more about how stereotype threat may affect you. However, the primary gains are in terms of enhanced knowledge about the scientific understanding of stereotype threat theory.
CONFIDENTIALITY	All information is confidential, and your survey will be identified by a code number. All information will be kept in locked file cabinets, and will be accessed by project staff for research purposes only. You will be asked to complete a brief demographic questionnaire for overall sample description.
LIABILITY	The Colorado Government Act determines and may limit Colorado State University's legal responsibility if an injury happens because of this study. Claims against the University must be filed within 180 days of the injury. Questions about participants´ rights may be directed to ***Janell Barker at (970) 491-1655.*** This consent form was approved by the CSU Institutional Review Board for the protection of human subjects in research on June 24, 2009.
PARTICIPATION	Your participation in this research is voluntary. If you decide to participate in the study, you may withdraw your consent and stop participating at any time without penalty or loss of benefits to which you are otherwise entitled.
GETTING STARTED	Click the button below that says START SURVEY. By beginning the survey, you are giving you consent to participate.

Researcher Contact Information
Silvia Sara Canetto, PhD, (970-491-5415), Department of Psychology, C-012 Clark Bldg.

CO-INVESTIGATOR AND PHONE NUMBER FOR QUESTIONS/PROBLEMS:
Deviyanti (Devi), M.A. (970-491-4150), Department of Psychology, C-015 Clark Bldg.

By proceeding to the next page, you agree to voluntarily participate in the study.

Please do not begin this survey unless you have set aside 1/2 hour (30 minutes) to complete it. You will not be able to stop and come back to the survey at a later time. There are 4 pages of survey to complete.

Enter your unique ID. Your unique ID should be the last two digits of the year you were born, then the last two digits of your cell phone number, then the last two digits of your Social Security Number.

For Example:
If you were born in 19**85**, your cell phone number is 555-12**34**, and your Social Security Number is 111-22-33**33**, then your unique ID is: **853433**

Your unique ID is required in order to take attendance and to link surveys completed now with surveys you complete later. Your answers to the survey will not be associated with you personally.

Enter your unique ID:

Start Survey

APPENDIX B

DEMOGRAPHIC FORM

Everything will be kept confidential. Your study ID number (not student ID number) will link your responses.

1. Sex: ________________________

2. Age: ________________________

3. What is your race or ethnicity? Please circle all that apply.
    a. European-American
    b. African-American
    c. Latina(o)/Hispanic
    d. Native American or Pacific Islander
    e. Asian-American
    f. Other (please specify): ___________________________________

4. Are you an International Student? (Circle) Yes No

5. If you are an international student, please write where you are from?
    a. I am not an international student
    b. I am from _______________________________________

6. In what year and semester of college are you?
    a. First year, first semester
    b. First year, second semester
    c. Second year, first semester
    d. Second year, second semester
    e. Third year, first semester
    f. Third year, second semester
    g. Fourth year, first semester
    h. Fourth year, second semester
    i. Other (please specify): _______________________________________

7. What is your major? __________________________

8. How confident are you in your major? (circle)

Very confident	Somewhat confident	Not very confident	Not confident at all

9. What would be your 2nd choice for a major? ___________________________

10. Current overall GPA (write N/A if this is your first semester in college): ____________

11. Current major GPA (write N/A if this is your first semester or you have no major): _____________

12. SAT Math score: _____________ 9b. SAT Verbal score: ___________

13. ACT Math score: ____________ 10b. ACT Total score: ___________

14. How many years of math did you take in high school?
    a. 0
    b. 1
    c. 2
    d. 3
    e. 4
    f. Other (please specify): __________

12. When did you complete your last math class (how many months or years ago)? __________________

13. Please list any math classes you have taken in high school or college (e.g., Calculus, Pre-calculus, Algebra, etc.). Please indicate the class name, when you took it, and the grade you received. If you are currently in a math class, please write "currently" on the second line.

Class name	When took it (high school or college)	Grade
________________________	________________________	__________
________________________	________________________	__________

APPENDIX C

IDENTIFICATION WITH MATH MEASURE

Instructions: *Using the following scale, please indicate the number that best describes how much you agree with each of the statements below.*

1	**2**	**3**	**4**	**5**
Strongly disagree	Moderately disagree	Neither disagree or agree	Moderately agree	Strongly agree

1. ______ Mathematics is one of my best subjects

2. ______ I have always done well in Math

3. ______ I get good grades in Math

4. ______ I do badly in tests of Mathematics

Instructions: *Please indicate the number that best describes you for each of the statements below using the following scale:*

1	**2**	**3**	**4**	**5**
Not at all		Somewhat		Very much

5. ______ How much do you enjoy math-related subjects?

6. ______ How likely would you be to take a job in a math related field?

7. ______ How much is Math to the sense of who you are?

8. ______ How important is it to you to be good at Math?

9. Compared to other students, how good are you at math?

    1. Very Poor

    2. Poor

3. About the same

4. Better than average

5. Excellent

Items taken from Smith & White (2001) Domain Identification Measure

APPENDIX D

MANIPULATION

1. Test indicative of math ability and No mention of performance differences

condition

Thank you for participating in this research. We are working in conjunction with the engineering, math, physical science, and behavioral science departments on campus to understand how people persist in engineering, math, physical science, and behavioral science fields. We want to establish profiles of students who show potential for success in these majors. Everything you do today will be at your own pace. It is important that you focus on your test only.

You are going to work on some math problems and then answer some questions about the problems and about yourself. There are thirty questions or incomplete statements, each followed by five answers. In each case, select the one answer that is the best of the choices offered and then mark it. You will have 30 minutes to complete the math test.

This test has been shown to be an excellent indicator of one's math aptitude and ability in a large number of settings across a wide spectrum of students, including engineering, math, physical science, and behavioral science majors. This test is especially effective at assessing the limitations in math problem areas of individual engineering, math, physical sciences, and behavioral science students. We ask that you give a strong effort in order to help us in our analysis of your math ability.

APPENDIX E

MANIPULATION

2. Test indicative of math ability and Women perform better than women condition

Thank you for participating in this research. We are working in conjunction with the engineering, math, physical science, and behavioral science departments on campus to understand how people persist in engineering, math, physical science, and behavioral science fields. We want to establish profiles of students who show potential for success in these majors. Everything you do today will be at your own pace. It is important that you focus on your test only.

You are going to work on some math problems and then answer some questions about the problems and about yourself. There are thirty questions or incomplete statements, each followed by five answers. In each case, select the one answer that is the best of the choices offered and then mark it. You will have 30 minutes to complete the math test.

This test has been shown to be an excellent indicator of one's math aptitude and ability in a large number of settings across a wide spectrum of students, including engineering, math, physical science, and behavioral science majors. This test is especially effective at assessing the limitations in math problem areas of individual engineering, math, physical sciences, and behavioral science students. Prior uses of these problems have shown gender differences; that is men have outperformed women on these problems. We ask that you give a strong effort in order to help us in our analysis of your math ability.

APPENDIX F

MANIPULATION

3. Test indicative of math ability and Women perform better than men condition

Thank you for participating in this research. We are working in conjunction with the engineering, math, physical science, and behavioral science departments on campus to understand how people persist in engineering, math, physical science, and behavioral science fields. We want to establish profiles of students who show potential for success in these majors. Everything you do today will be at your own pace. It is important that you focus on your test only.

You are going to work on some math problems and then answer some questions about the problems and about yourself. There are thirty questions or incomplete statements, each followed by five answers. In each case, select the one answer that is the best of the choices offered and then mark it. You will have 30 minutes to complete the math test.

This test has been shown to be an excellent indicator of one's math aptitude and ability in a large number of settings across a wide spectrum of students, including engineering, math, physical science, and behavioral science majors. This test is especially effective at assessing the limitations in math problem areas of individual engineering, math, physical sciences, and behavioral science students. Prior uses of these problems have shown that women have outperformed men on these problems. We ask that you give a strong effort in order to help us in our analysis of your math ability.

APPENDIX G

MANIPULATION

4. Test not indicative of math ability and No mention of differences condition

You are going to work on some math problems and then answer some questions about the problems and about yourself. There are thirty questions or incomplete statements, each followed by five answers. In each case, select the one answer that is the best of the choices offered and then mark it. You will have 30 minutes to complete the math test.

The task you will be working on today was specifically designed to present you with problems varying in their degree of difficulty. We ask that you give a genuine effort and please try hard.

APPENDIX H

MANIPULATION

5. Test not indicative of math ability and Men perform better than women condition

You are going to work on some math problems and then answer some questions about the problems and about yourself. There are thirty questions or incomplete statements, each followed by five answers. In each case, select the one answer that is the best of the choices offered and then mark it. You will have 30 minutes to complete the math test.

The task you will be working on today was specifically designed to present you with problems varying in their degree of difficulty. Previous results on this test have shown gender differences; that is, men have outperformed women on these problems. We ask that you give a genuine effort and please try hard.

APPENDIX I

MANIPULATION

6. Test nNot indicative of math ability and Women perform better than men condition

You are going to work on some math problems and then answer some questions about the problems and about yourself. There are thirty questions or incomplete statements, each followed by five answers. In each case, select the one answer that is the best of the choices offered and then mark it. You will have 30 minutes to complete the math test.

The task you will be working on today was specifically designed to present you with problems varying in their degree of difficulty. Previous results on this test have shown that women have outperformed men on these problems. We ask that you give a genuine effort and please try hard.

APPENDIX J

SCRIPTED VERBAL INSTRUCTIONS

(1) *Have students sign in as they arrive and hand them the manipulation sheet to read.*

(2) *Read these instructions to the group:*

"Thanks for coming today. Welcome to the second part of the 'Identity Interference" study. Please turn off your cell phone. This part takes about an hour to complete. I am going to get you started on the paper and pencil portion of the study and then let you complete the rest of the study online. Please read the instructions that you were given when you signed in now, if you have not already done so.

(3) *Give them a minute or two, then ask if everyone is done. When everyone is done reading, read these instructions to the group:*

I will be handing out a test for you to complete in a minute. Please read each page carefully and start with the paper portion first before moving on to the online measures. I will be instructing you to stop working on the paper version 30 minutes from now and collect the papers. Do not start the online survey until you are instructed to do so. Please note that this study is completely voluntary and your answers will be kept separate from your contact information. You may start now.

(4) *Pass out the problems. Give them a 5-minute warning before you have them stop working at 30 minutes. Have them hold onto the papers while you read the next section. At this time, have them find the URL. (http://psy.psych.colostate.edu/surveys/devi/default2.asp)*

(5) *Read these instructions to the group:*

As you can see on the screen, you will need to enter your unique ID for the online portion of the study for confidentiality purposes. Please be sure you use the same unique ID you used for the first survey. Your unique ID should be the last two digits of the year you were born, then the last two digits of your cell phone number, then the last two digits of your Social Security Number in that order. You created your unique ID when you took the first set of surveys. If you have questions or if you are unable to remember your ID number, please raise your hand and I will look up your number from the list. Once you have signed on to the survey, please wait for further instructions.

(6) *At this point, help anyone who can't remember their password.*

(7) *Once everyone has logged into the survey, read these instructions:*

Before you get started, please transfer you unique ID code onto the papers next to the word ‘Code’ in the upper right hand corner. We will collect the papers at the end of the study.

(8) Give them a moment to write the code on their papers. Then continue:

Now turn you attention to the screen. Please read each page of the survey carefully. Work individually and keep focused on your own computer. When you are finished, exit your screen and wait for further instructions.

(9) *As soon as everyone is done, you can excuse the group. Instruct them as follows:*

Thank you for participating in our study. Please turn in your papers to me as you leave.

APPENDIX K

MATHEMATICS TEST

Instructions: *Please complete the following math test. You will have 30 minutes. Please circle your answers. For each problem, please rate how confident you are in your answer, on a scale of 0 to 4, where 0 represents no confidence at all and 4 represents complete confidence.*

1. When a certain number is divided by 7, the remainder is 0. If the remainder is not 0 when the number is divided by 14, then the remainder must be.

    a. 1
    b. 2
    c. 4
    d. 6
    e. 7

0	1	2	3	4
No confidence at all				Complete Confidence

2. If $4,500 was invested in a bond fund when the price per share was $9 and $3,000 was invested in the fund when the price per share was $10, what was the average (arithmetic mean) price per share purchased?

    a. $9.625
    b. $9.50
    c. $9.40
    d. $9.375
    e. $9.20

0	1	2	3	4
No confidence at all				Complete Confidence

3. Which of the following equations can be used to find the value of x if 7 less than 5x is 5 more than the product of 3 and x?

    a. $5x – 7 = 5 + 3x$
    b. $5x –7 = 5 + (3 + x)$
    c. $7 – 5x = 5 + 3x$
    d. $7 – 5x = (5 + 3)x$
    e. $7 – 5x + 5 = 3x$

0	1	2	3	4
No confidence at all				Complete Confidence

4. Mario bought equal numbers of 2-cent and 3-cent stamps. If the total cost of the stamps was $1.00, what was the total number of stamps bought?

a. 25
b. 34
c. 40
d. 46
e. 50

0	1	2	3	4
No confidence at all				Complete Confidence

5. Jane has exactly 3 times as many Canadian as non-Canadian stamps in her collection. Which of the following CANNOT be the number of stamps in Jane's collection?

a. 96
b. 80
c. 72
d. 68
e. 54

0	1	2	3	4
No confidence at all				Complete Confidence

6. Chris gave Jane x cards. He gave Betty one card more than he gave Jane and he gave Paul two cards fewer than he gave Betty. In terms of x, how many cards did Chris give Betty, Jane, and Paul altogether?

a. 3x + 1
b. 3x
c. 3x – 1
d. x – 1
e. $\frac{x}{3}$

0	1	2	3	4
No confidence at all				Complete Confidence

7. Three individuals contributed $800 each toward the purchase of a computer. If they bought the computer on sale for $1,950 plus 10 percent sales tax, how much money should be refunded to each individual?

a. $65
b. $85
c. $150
d. $195
e. $255

0	1	2	3	4
No confidence at all				Complete Confidence

8. A widow received 1/3 of her husband's estate, and each of her three sons received 1/3 of the balance. If the widow and one of her sons received a total of $60,000 from the estate, what was the amount of the estate?

a. $90,000
b. $96,000
c. $108,000
d. $135,000
e. $180,000

0	1	2	3	4
No confidence at all				Complete Confidence

9. If x can have only the values -3, 0, 2, and y can have only the values -4, 2, and 3, what is the greatest possible value for $2x + y^2$?

a. 13
b. 15
c. 16
d. 20
e. 22

0	1	2	3	4
No confidence at all				Complete Confidence

10. If the cost of a long-distance phone call is c cents for the first minute and $\frac{2}{3}$ c cents for each additional minute, what is the cost, in cents, of a 10-minute call of this type?

a. $\frac{5}{3}$ c
b. 6 c
c. $\frac{20}{3}$ c
d. 7 c
e. $\frac{23}{3}$ c

0	1	2	3	4
No confidence at all				Complete Confidence

11. City Y has installed 30 parking meters at 15 foot intervals along a straight street. What is the number of feet between the first meter and the last meter?

a. 200
b. 420
c. 435
d. 450
e. 465

0	1	2	3	4
No confidence at all				Complete Confidence

12. In a certain apartment building exactly 1/3 of the apartments have two bedrooms and exactly 1/7 of the two-bedroom apartments are front apartments. Which of the following could be the total number of apartments in the building?

a. 42
b. 50
c. 51
d. 56
e. 57

0	1	2	3	4
No confidence at all				Complete Confidence

13. The number of connections C that can be made through a switchboard to which T telephones are connected is given by the formula: C = $\frac{T(T-1)}{2}$. How many more connections are possible with 30 telephones than with 20 telephones?

a. 435
b. 245
c. 190
d. 45
e. 10

0	1	2	3	4
No confidence at all				Complete Confidence

14. If 5 percent of a rectangular lot is covered by a rectangular shed that is 25 feet long and 24 feet wide, what is the area of the lot in square feet?

a. 3,000
b. 5,700
c. 12,000
d. 22,500
e. 30,000

0	1	2	3	4
No confidence at all				Complete Confidence

15. If membership in the Elks Club increases from 120 to 150, what is the percent increase?

a. 15%
b. 25%
c. 30%
d. 40%
e. 80%

0	1	2	3	4
No confidence at all				Complete Confidence

16. The length of a rectangular floor is 16 feet and its width is 12 feet. If each dimension were reduced by x feet to make the ratio of the length to width 3 to 2, what would be the value of x?

a. 0
b. 2
c. 4
d. 6
e. 8

0	1	2	3	4
No confidence at all				Complete Confidence

17. The Acme Rent-A-Car agency charges $10.00 per day and $0.10 per mile to rent a car. The Super Rent-a-Car agency charges $20.00 per day and $0.05 per mile to rent a car. If a car is rented for 1 day, at how many miles would the rental charges of the two agencies be equal?

a. 50
b. 100
c. 150
d. 175
e. 200

0	1	2	3	4
No confidence at all				Complete Confidence

18. A school district has 1,989 computers, which is approximately one computer for every 68.6 students. Of the following, which is the closest approximation, in thousands, of the number of students in the school district?

a. 30
b. 120
c. 140
d. 160
e. 200

0	1	2	3	4
No confidence at all				Complete Confidence

19. A secretary typed 6 letters, each of which had either 1 or 2 pages. If the secretary typed 10 pages in all, how many of the letters had 2 pages?

a. 1
b. 2
c. 3
d. 4
e. 5

0	1	2	3	4
No confidence at all				Complete Confidence

20. The rectangular floor of a warehouse is 300 feet wide and 350 feet long. If the width remains fixed, how many additional feet would have to be added to the length to increase the floor area by 20%?

a. 42
b. 50
c. 65
d. 70
e. 84

0	1	2	3	4
No confidence at all				Complete Confidence

21. If 4x is 9 greater than the sum of x and 3y, then x is how much greater than y?

a. 3
b. 6
c. 9
d. 12
e. 15

0	1	2	3	4
No confidence at all				Complete Confidence

22. Two people were hired to mow a lawn for a total of $45. They completed the job with one person working for 1 hour and 20 minutes and the other working 40 minutes. If they split the $45 in proportion to the amount of time each spent working on the job, how much did the person who worked longer receive?

a. $33.75
b. $30.00
c. $27.50
d. $25.00
e. $22.50

0	1	2	3	4
No confidence at all				Complete Confidence

23. If the sum of two numbers is known, which of the following is NOT sufficient to determine the values of the two numbers?

a. One number is greater than the other.
b. The cube of one number is 8.
c. The product of the two numbers is 8.
d. The difference between the two numbers is 2.
e. One number is half the other.

0	1	2	3	4
No confidence at all				Complete Confidence

24. If a person can save $380 in 5 weeks, in how many weeks, at this same rate can the person save 2.6 times this amount.

a. 13
b. 12.5
c. 11
d. 10.6
e. 8

0	1	2	3	4
No confidence at all				Complete Confidence

25. If the sum of 12, 15, and x is 45, then the product of 5 (x + 2) is

a. 100
b. 92
c. 80
d. 41
e. 25

0	1	2	3	4
No confidence at all				Complete Confidence

26. In a certain shipment 2 percent of the boxes shipped were damaged. If the loss per damaged box was $35 and the total loss due to damage was $700, how many boxes were shipped?

a. 2,000
b. 1,000
c. 200
d. 100
e. 20

0	1	2	3	4
No confidence at all				Complete Confidence

27. If the average (arithmetic mean) of two numbers is 20 and one of the numbers is x, what is the other number in terms of x?

a. 40 – x
b. 40 – 2x
c. 20 + x
d. 20 – x
e. 20 – 2x

0	1	2	3	4
No confidence at all				Complete Confidence

28. A watch gains 7 minutes and 6 seconds every 6 days. If the rate of gain is constant, how much does the watch gain in one day?

a. 1 min 1 sec
b. 1 min 6 sec
c. 1 min 11 sec
d. 1 min 16 sec
e. 1 min 21 sec

0	1	2	3	4
No confidence at all				Complete Confidence

29. A time-study specialist has set the production rate for each worker on a certain job of 22 units every 3 hours. At this rate what is the minimum number of workers that should be put on the job if at least 90 units are to produced per hour?

a. 5
b. 8
c. 12
d. 13
e. 30

0	1	2	3	4
No confidence at all				Complete Confidence

30. The daily rate for a hotel room that sleeps 4 people is $39 for one person and x dollars for each additional person. If 3 people take the room for one day and each pays $21 for the room, what is the value of x?

a. 6
b. 8
c. 12
d. 13
e. 24

0	1	2	3	4
No confidence at all				Complete Confidence

Taken from the quantitative portion of the Graduate Record Examination (GRE)

APPENDIX L

POST-TEST QUESTIONNAIRE

1. How difficult did you find the problems?

2. Out of the problems you answered, how many do you think you answered correctly (0-30)?

3. How have women performed relative to men on these problems?
   A) Men have performed better than women
   B) Women and men have performed equally
   C) Women have performed better than men
   D) No mention of how women and men have performed relative to one another

4. Which of the following is true about the math test you just took?
   A) The test has been shown to be a good indicator of one's math aptitude and ability
   B) The test has been shown to be a poor indicator of one's math aptitude and ability
   C) No mention of whether the test is a good or poor indicator of one's math aptitude and ability

APPENDIX M

DEBRIEFING FORM

Identity Interference in Women and Math

Many women have the abilities, drive, and interests to go into math-related careers (e.g., engineering, computer science). However many of these women do not pursue graduate degrees in math or math-related fields such as physics, engineering, or economics. Some researchers (e.g., Steele, 1992; Aronson & Steele, 1995; Steele, 1997; Spencer, Steele, & Quinn, 1999) argue that women can experience “stereotype threat” in math or math-related settings. This study examines if stereotype threat affects women’s self-confidence, math performance, and other math-related skills under specific conditions.

To gain a better understanding of stereotype threat and its effects on women and other groups, please see the chapter on intelligence in your textbook (Myers, D. G. (2008). Exploring Psychology, 7th Ed. in Modules. New York: Worth Publishers). Module 25 on pages 350-351 describes the role of stereotype threat in aptitude test scores and school performance.

Some of the conditions involved telling participants that the math test was diagnostic of personal ability and that similar tests have shown men to perform better than women or that women perform better than men. These statements were created for the purpose of this study and are not true. This deception was necessary so that participants would be put in a stereotype threat situation. Doing so allowed researchers to see the effects of various factors on stereotype threat and the effect of stereotype threat on subsequent math performance. The information generated by this study may therefore contribute to the scientific understanding of stereotype threat theory. Being aware of stereotype threat theory may prevent stereotypes groups such as women from performing poorly and dropping out of math-related fields. Having participated in this study, you are now armed with the awareness of stereotype threat and may be more likely to persist in your field than had you not participated.

We appreciate your participation in this research project. We hope that you have learned something about the process of psychological research and gained some insight about how stereotype threat might apply to you. If you have any questions about this research, please feel free to contact:

Deviyanti (Devi), M.A.
Doctoral Candidate, Department of Psychology
Colorado State University
1876 Campus Delivery
Fort Collins, CO 80523-1876
970-491-4150
Devi.NLN@colostate.edu

Silvia Sara Canetto, PhD
Professor, Department of Psychology
Colorado State University
1876 Campus Delivery
Fort Collins, CO 80523-1876
970-491-5415
Silvia.Canetto@gmail.com

Note: If you would like to have your data removed from this study, tell the researcher before you leave the session.