2010

Women and Math Performance: the Effects of Stereotype Threat, Math Identity, and Gender Identity

Felicia W. Chu
Seton Hall University

Women and Math Performance: the Effects of Stereotype Threat, Math Identity, and Gender Identity

by

Felicia W. Chu

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Experimental Psychology with a concentration in Behavioral Neuroscience

Department of Psychology

Seton Hall University

May, 2010

Approval Page

Approved By:

Dr. Janine Buckner, Faculty Mentor
and Director of Graduate Studies

Dr. Paige Fisher, Committee Member

Dr. Marianne Lloyd, Committee Member

Dr. Susan Nolan, Committee Member

Dedication

This thesis is dedicated to my parents, Tung and Hui, and my sister, Emily:

Thank you for all your love and support.

Acknowledgments

I would like to sincerely thank Dr. Janine Buckner for all her guidance and support as I have developed and completed my thesis. Without you, this thesis would not have been possible. I am also extremely thankful to my thesis committee — Dr. Paige Fisher, Dr. Marianne Lloyd, and Dr. Susan Nolan — for all their input and encouragement throughout this process. In addition, I am grateful to the members of the Math Department with whom I have worked, for their caring and support during the past two years. You are all amazing people, and I have enjoyed working with you during my graduate assistantship. Thank you also to my friends and fellow graduate students, who have given me advice and reassurance. Lastly, thank you to my family, for believing in me. You have always been there when I've needed you.

Table of Contents

List of Figures

List of Tables

Abstract

Stereotype threat is a phenomenon that emerges when a negative task-relevant stereotype is activated; subsequently, participants may show poorer performance regardless of actual ability. One such stereotype dictates that men are better than women at math. Previous studies (e.g., Lesko & Corpus, 2006; Martens, Johns, Greenberg, & Schimel, 2006; Spencer, Steele, & Quinn, 1999; Quinn & Spencer, 2001) have shown that women underperform on math tests when exposed to such a stereotype threat, but can perform comparably to men, on average, when stereotype threat is nullified. Other work (Thoman, White, Yamawaki, & Koishi, 2008) has revealed differential effects of threat type (innate ability vs. effortful control of performance outcomes), as well as differences in attributions for poor performance, with women being more likely to ascribe poor performance to ability than men (Kiefer & Shih, 2006). The purpose of the present study was to investigate the role of stereotype threat, gender identity, and math identity on math performance. Female participants completed a math test after being exposed to a stereotype condition (no stereotype threat/control, effort stereotype threat, and innate ability stereotype threat). In addition, strength of math identity and gender identity were assessed using questionnaires. Although participants did not differ in number of math problems answered or percentage correct among stereotype conditions, on average, gender identity and math identity influence math performance differently among conditions.

Introduction

According to stereotype threat theories, negative beliefs about a social group may lead to poorer performance from members of that social group, especially if they fear confirming the negative beliefs. Various factors have been implicated in stereotype threat, including individual difference variables (e.g. age, race, gender, and socioeconomic status) as well as threat type (e.g. innate ability stereotype threat and effort stereotype threat). In an early study of stereotype threat, Steele and Aronson (1995) targeted the stereotype that African Americans are not as intelligent as whites; they found that African American students underperformed on a difficult verbal test in comparison to white students when they were told the verbal test was diagnostic of their intellectual abilities. In contrast, African American students' performance improved drastically if they were told the verbal test was not reflective of their abilities. Consequently, researchers and theorists posited that performance on relevant tasks may suffer due to membership in negatively-stereotyped groups. The potential implications of stereotype threat revealed in this study led to additional research on stereotype threat; since the Steele and Aronson study, one of the most specifically targeted negative stereotypes examined relates to math performance in women.

Studies involving stereotype threat and math have focused on aspects of social identity, such as gender identity or ethnic identity. Performance suffers when a person's social identity is negatively stereotyped, or their performance is being compared to that of a person whose social identity is positively stereotyped. Previous studies have shown that aspects of social identity can influence performance on a math test. For example, when Caucasian men were told their performance on a math task was being compared to that of Asian men, their performance worsened, on average (Aronson, Lustina, Good, & Keough, 1998). This pattern was attributed to the stereotyped expectation that Asians excel in math; upon believing that they were being compared to Asian men, Caucasian men may experience stereotype threat and show poorer performance. Similarly, Smith and White (2002) found that when Caucasian men in an advanced mathematics course were exposed to a stereotype that Asians were better at math than Caucasians, their performance on a math test suffered; however, for the group of Caucasian men in the nullification condition, in which participants were told that the test did not show differences among races, this deficit did not occur. Thus, despite Caucasian men being capable in mathematics (as evidenced by their membership in an advanced mathematics course), stereotype threat may still hinder performance on relevant tasks. However, Caucasian men are not typically a direct target of stereotype threat, whereas women are; consequently, more studies have focused upon the effects of stereotype threat on female math performance than on male performance.

In a set of experiments, Spencer, Steele, and Quinn (1999) investigated the role of stereotype threat in math performance by comparing men and women's math test scores. In their first experiment, male and female participants matched for high math abilities (SAT scores at 85^{th} percentile or greater, and a "B" or better in at least one semester (but less than a year) of Calculus) completed either an easy or difficult math test. Spencer et al. (1999) found that women scored equally to men on the easy math test, but worse than men on the difficult math test, on average. In the second experiment, participants were again chosen on the same criteria; however, in this study, they completed two tests: one in which they were told gender differences had been found, and one in which they were told gender differences had not been found. Spencer et al. found that in this case, women who were told that gender differences had been found (a stereotype threat condition) performed worse than men, whereas women's performance was not significantly different, on average, from that of men when they were told that gender differences had not been found on the math test. In a third experiment, men and women completed a math test in which they were told that there were no gender differences; their performance was compared to that of a control group, in which they were told nothing at all regarding gender differences. Similar to the results of Experiment 2, women who were told that there were no gender differences performed equally to men, on average, and women who were not told anything about gender differences performed worse than men, on average. These results taken together indicate that women's math performance is susceptible to the negative influences of stereotype threat, even if not explicitly stated; consequently, participants may have a "chronic" stereotype that men are better at math than women—a negative belief that can hinder women's math performance specifically.

Consequences of Stereotype Threat: Working Memory Deficits

Many studies posit that a significant consequence of stereotype threat, a reduction in working memory, may result in poor strategy generation (e.g. Beilock & Carr, 2005; Beilock & DeCaro, 2007; Beilock, Rydell, & McConnell, 2007; Schmader & Johns, 2003). Indeed, in a study by Schmader and Johns (2003) investigating working memory under conditions of stereotype threat, participants were asked to solve math equations while memorizing words listed at the end of each equation. In the end, female participants assigned to a stereotype threat condition recalled fewer words than did male counterparts, or females *not* assigned to a stereotype threat condition. In addition, these female participants had lower scores on the math test, leading the researchers to conclude that the reductions in working memory were mediators that led to their poor performance.

In a similar set of studies by Beilock (e.g., Beilock & Carr, 2005; Beilock & DeCaro, 2007; Beilock et al., 2007) individuals with high working memory showed poorer choices in problem solving strategies, particularly under high pressure conditions. Individuals with high working memory tend to select more computationally demanding algorithms to ensure accuracy under low pressure conditions; in contrast, under high pressure conditions, they were more likely to choose simple and less effective methods that impeded their speed and caused their performance to suffer (Beilock & DeCaro, 2007). Thus, the high pressure decreased the availability of working memory available for problem-solving, resulting in a decline in performance and increased processing time. In addition, Quinn and Spencer (2001) found that under high pressure, women exposed to stereotype threat were less able to generate problem-solving strategies on word problem tests. These results were not due to actual mathematical ability, since women and men performed equally well when the word problems were converted into numerical mathematical equations. Thus, under stereotype threat, women may attempt to suppress the negative thoughts, which results in an increase in cognitive load; consequently, in such conditions there are only limited cognitive resources available to search for an appropriate problem-solving strategy, and performance decreases as a result.

The deficits in working memory linked to stereotype threat exposure can also carry over to subsequent tasks that are not necessarily related, but may require some of the same cognitive demands as the task performed while under stereotype threat. For example, Trbovich and Lefevre (2003) showed that increasing one's phonological load (by memorizing consonant-vowel-consonant nonwords) resulted in increased latency in answering mental arithmetic problems. In contrast, a visual load (memorizing patterns of asterisks) resulted in a decline in latency, particularly for similar horizontally-presented problems. Because the spatial task comprised of patterns of asterisks did not increase phonological load, the cognitive resources necessary for mental arithmetic were available, unlike the consonant-vowel-consonant nonwords condition. Drawing from this previous work, Beilock, Rydell, and McConnell (2007) asked female participants to complete a math test requiring high working memory while under stereotype threat. Afterwards, these women performed either a spatial or verbal task. Beilock et al. (2007) hypothesized that the verbal task would require similar phonological demands that solving math problems utilized, and thus, participants were expected to show decreased performance on the verbal task. Indeed, this was the case, as the researchers found that the participants' decreased performance under stereotype threat carried over to the verbal task. However, under a no-stereotype threat condition, participants performed at a high level, with greater accuracy and lower response time.

Gender Identity

Although women perform worse, on average, than men on difficult math tests, they perform as well as men, on average, on easier tests; however, if the difficult test is characterized as insensitive to gender differences, underperformance is eliminated (Spencer et al., 1999). This finding suggests that stereotype threat can be nullified. That is, negative effects of a stereotype threat can be prevented or counteracted ("nullified") when participants are explicitly told that a negative stereotype is not true, thus reducing the negative impact of the stereotype. In addition, when women are told a math test will indicate gender differences, women show poorer performance compared to men, but only if they consider gender to be central to their self-concept (Schmader, 2002). When gender was not considered to be an important aspect of their social identity, women did not show lower performance than their male peers. Thus, gender identity, or the extent to which one believes he or she belongs to a particular gender, may mediate the effect of stereotype threat, as people adopt the expectations of their perceived gender that are deemed socially normal ("gender roles"); for example, based on stereotypical gender beliefs, female students may be more inclined to pursue subjects in the arts, while male students may be more inclined to pursue subjects in the sciences, such as math. Similar to Schmader's (2002) study, women who highly identify with a feminine gender role underperform on math tasks, in comparison to less gender-identified women who show smaller negative effects in math performance (Wout, Danso, Jackson, & Spencer, 2008).

An important caveat to this effect of gender is that attenuation in math performance appears to be dependent on source of threat-- whether the stereotype threat is targeted towards an individual (poor performance would be indicative of one's own weakness) or a group (poor performance would be indicative of a group's weakness). As Wout and colleagues (2008) demonstrated, when women were induced to apply self-threat, gender identification did not moderate test performance; however, when targeted by group-threat, women high in gender identification underperformed compared to women low in gender identification. Furthermore, women who have stronger implicit gender-math stereotypes (and thus, are more likely to associate math as being a masculine subject) are less likely to be helped by nullification, possibly because they chronically activate negative stereotypes (Kiefer & Sekaquaptewa, 2007). In contrast, women with weak implicit gender-math stereotypes appear to benefit from nullification of stereotype threat.

Other Dimensions of Identity

Although stereotype threat has been shown to be influenced by one's social identity, many people have more than one dimension to their social identity; stereotypes involving these dimensions may sometimes conflict for an individual. Shih, Pittinsky, and Ambady (1999) investigated the effects of stereotype threat when some of these different dimensions were activated. In particular, they examined the interaction between gender and ethnic identity. Asians are stereotyped to excel at math (a positive stereotype), while women are stereotyped to be poor at math (a negative stereotype). In their study, Shih et al. (1999) explored how Asian American women would perform on a math test under one of three conditions: female identity salient, Asian identity salient, or neither salient. Two aspects of their findings were striking. First, Shih et al. (1999) discovered that women performed their best (were most accurate in their answers) if their Asian identity was activated, and achieved the worst scores when their female identity was activated. However, perceptions of the test, their performance, or skill did not vary by participant identity categories. That is, no significant differences were found between number of problems answered or guessed, liking of the test, and assessment of difficulty, performance, and one's own mathematical skills between groups.

Furthermore, because no significant differences were found among the participants in terms of their reported quantitative scores on the SAT (scores ranged from 600-800, with a mean of 750.9), the differences between the positive stereotype condition and the negative stereotype condition were not reflective of ability; rather, the different activations in social identity contributed to the variations in performance. Similarly, receiving both a positive stereotype (college students are good at math) and a negative stereotype (women are not good at math) can reduce the negative effects of stereotype threat; if female college students received both stereotypes (i.e., a multiple identity condition), deficits were eliminated, but participants' performance suffered the most for participants who received only the negative stereotype (i.e., the gender identity condition) (Rydell, McConnell, & Beilock, 2009).

Domain Identification and Negative Stereotype Processing

The effects of stereotype threat not only vary depending on dimensions of social identity, but also how closely related a participant is to the subject on which they are tested. In particular, several studies have investigated how perception of math (or the degree to which one views math as an important or central aspect of her or his identity) mediates the influence of stereotype threat on test performance (Krendl, Richeson, Kelley, & Heatherton, 2008; Lesko & Corpus, 2006). Women who considered it important that they excel at math had heightened activation in the ventral anterior cingulate cortex (vACC, a neural region in the brain) when performing math-related tasks (Krendl et al, 2008). Activation of this vACC region is associated with social and emotional processing, particularly negative feedback and clinical depression.

The negative effects of stereotype threat were seen not only in brain activations, but in participants' perceptions as well. In studies examining self-reports gleaned from women who completed a math test under stereotype threat, participants reported a higher number of negative thoughts specifically related to the test and to mathematics in general, in comparison to a no-threat control group (Cadinu, Maass, Rosabianca, & Kisener, 2005). Participants exposed to stereotype threat showed the typical results of underperformance in comparison to participants not exposed to stereotype threat. Furthermore, in a study exploring how high math-identified participants respond to stereotype threat, Lesko and Corpus (2006) gave college students a challenging math test with either an active or nullified stereotype threat. Women performed worse than men under the stereotype threat condition, although performance between genders was equal when stereotype threat was nullified. Moreover, high math-identified women under stereotype threat discounted the validity of the test more than less math-identified women and men. Discounting refers to doubting the validity of the test and claiming the test is an incorrect measure of ability. This difference in discounting was not due to perceived performance, because tests scores and explicit feedback were not given. The negative effects of stereotype threat, as seen through brain activations, perceptions of the test, and the discounting response, may thus lead to poorer performance due to the increased bias women feel against them.

As previously discussed, different types of stereotype threat can influence performance in different ways. To test the differential effects of stereotype threat, Thoman et al. (2008) exposed female participants to either an effort stereotype threat or an innate ability stereotype threat before asking them to take math tests. Women were told that men typically performed better than women due to the fact that men spend more effort on such tasks, or that men were better than women due to innate, biological and genetic differences. When exposed to the effort stereotype threat, women answered fewer problems, but attempted and answered a greater percentage of problems correctly. The difference in ratio of correct problems between the innate ability stereotype threat group and the control group was not statistically significant.

Thoman et al.'s (2008) study mainly focused on the roles of the different kinds of stereotype threat; however, it is possible that other factors interact with stereotype threat type in order to influence performance. For example, as previously discussed, gender identity has been shown to influenced performance, and thus, it would be of interest to determine how different gender identities respond to the different types of threat. Furthermore, different levels of math identity may also result in different responses depending on threat condition, and thus, is another variable that should be studied in conjunction with stereotype threat condition. In order to gain a better understanding of stereotype threat on women's math performance, the present study was therefore designed to investigate the influence of these variables in specific ways.

Statement of the Problem and Purpose

The consequences of negative stereotypes can be detrimental, and thus, it is essential to understand the mechanisms that mediate the effects of stereotype threat. Furthermore, various types of stereotype threat negatively impact students' performance, not the least of which is that stereotype threats may hinder students from performing at their true level of ability. It is important, therefore, to determine the factors that influence stereotype threat in order to mitigate or even eradicate such detrimental consequences.

The purpose of this study was to determine the extent to which gender differences in math performance are affected by stereotype threat. In particular, different kinds of stereotype threat (such as those which activate beliefs that differences in performance are due to innate ability vs. due to exerted effort) were studied. Drawing from previous work of Thoman et al. (2008), effort, innate ability, and control conditions were used to assess responses to different stereotype threats in math test scenarios. However, this study expands the previous research by including an analysis of the influences of gender identity and math identity on math performance. To this end, strength of participants' identification with math (a continuous scale) and perceived gender identity (masculine, feminine, androgynous, and undifferentiated) were assessed before they were exposed to one of three stereotype threat conditions. Thus, in recognition of the profound role that specific aspects of identity have in determining self-efficacy and motivation in math, gender identity and the extent to which individuals identify with math were measured. In this way, the extent to which such individual differences in these variables influence vulnerability to particular stereotype threats related to math performance were assessed.

Hypotheses

Math performance, as dependent variable, was defined in this study via two measures: the number of math problems attempted on a math test, as well as the percentage correct out of those items attempted (a metric of "accuracy"). These data were used to compare different types of performance across types of stereotype threat, along four gender identity (GI) groups and a continuous scale of math identity (MI). An interaction was expected between threat type, GI, and

MI for both the number of test items participants attempted to answer, and for measures of accuracy of their answers. These effects are described below, as three sets of hypotheses based upon threat-type condition.

Effort stereotype threat. Under effort stereotype threat, we expected that math performance would be best for women who were highly math-identified and perceived themselves to be more like traditionally masculine than feminine stereotypes (high in masculine GI, low in feminine GI). These high MI, masculine GI women were expected to attempt more questions and have higher accuracy than high MI, "feminine" GI peers (those low in masculine traits, high in feminine traits). Masculine types who were low in MI were also expected to attempt more questions than low MI feminine types, but due to the poor math identification, they were expected to have lower accuracy than those who have high MI.

Regarding "feminine" gender types, those with high MI were expected to focus more effort upon correct answers than their low MI peers. They were therefore predicted to attempt fewer questions but achieve higher accuracy than feminine GI women with low math identity. But the average difference in accuracy between participants with high and low MI was expected be smaller for "feminine" women in comparison to those who were masculine in gender identity.

Based on the gender identity measure employed in the study, two additional gender types were identified: androgynous and undifferentiated individuals. But math performance was not expected to vary significantly for stereotype conditions acrossh androgynous and undifferentiated gender identities. This was because these gender identity types involve specific combinations of both masculine and feminine characteristics--high scores on both masculinity and femininity for androgynous individuals or low scores on both subscales in the case of undifferentiated individuals.

Innate ability stereotype threat. In this condition, interactions were also predicted. For participants in the innate ability stereotype threat, participants with a masculine gender identity were not expected to differ significantly from masculine women in the effort stereotype threat, in terms of number attempted. However, within this ability threat group, masculine types with high math identity were predicted to have greater accuracy than those with low math, masculine identity. The same pattern (high-math would perform better than low-math women) was expected for those with feminine gender identification. Consequently, in contrast to the effort stereotype condition, the differences between high and low math identity were expected to be similar for the masculine and feminine gender identities on average. Again, androgynous and undifferentiated identities were not examined for either measure of math performance.

Control. Participants in the control group were expected to perform similarly to those in the innate ability stereotype group. Furthermore, aspects of gender identity (GI) were not expected to significantly influence performance in the control condition, or at least, not to the same extent as participants in the innate ability condition. Participants in this condition were not being explicitly threatened, and hence, aspects of their gender identity would have been less likely to be activated. However, math identity was expected to be a significant predictor, though the exact relation to performance was not specified. Because students were not being explicitly threatened, we did not expect them to be as motivated to spend increased time on problems. But because they could have an implicit, chronic awareness of gender-math stereotypes, their performance nonetheless was expected to be negatively impacted, regardless of gender identity. In this case, they would not have performed as well as participants in the effort stereotype condition. This prediction would be consistent with the results of Thoman et al. (2008).

A summary of the results we predicted is provided in Table 1.

Table 1. Predicted Results of the Three-way Interaction.

TT x GI x MI	*MI*	*Effort (E)*	*Innate Ability (IA)*	*Control (C)*
Masculine (M)	*High (H)*	High number attempted, may be slightly less than IA-M-H	High number attempted	High number attempted
		High accuracy	High accuracy	High accuracy
	Low (L)	Fewer attempted than high E-M-H	Higher number attempted than E-M-L	Higher number attempted than E-M-L
		Lower accuracy than high E-M-H	Lower accuracy than E-M-L, IA-M-H	Lower accuracy than E-M-L, C-M-H
Feminine (F)	*High (H)*	High number attempted, but fewer than E-M-H	High number attempted, more than E-F-H	High number attempted, more than E-F-H
		High accuracy	Lower accuracy than E-F-H	Lower accuracy than E-F-H, similar to that of IA-F-H
	Low (L)	Few attempted	More attempted than E-F-L	More attempted than E-F-L
		High accuracy	Lower accuracy than E-F-L	Lower accuracy than E-F-L, similar to that of IA-F-L
Androgynous (A)	*High (H)*	Androgynous identity is not expected to be predictive of number attempted; however, more problems will likely be attempted by those high in math identity than those low in math identity		
		Androgynous identity is not expected to be predictive of accuracy; however, higher math identification will result in greater accuracy than lower math identification.		
	Low (L)	Fewer problems attempted than Androgynous-High Math Identity.		
		Lower accuracy than Androgynous-High Math Identity.		
Undifferentiated (U)	*High (H)*	Undifferentiated identity is not expected to be predictive of number attempted; however, more problems will likely be attempted by those high in math identity than those low in math identity		
		Undifferentiated identity is not expected to be predictive of accuracy; however, higher math identification will result in greater accuracy than lower math identification.		
	Low (L)	Fewer problems attempted than Undifferentiated-High Math Identity.		
		Lower Accuracy than Undifferentiated-High Math Identity.		

*Groups are defined in terms of TT-GI-MI (ex. participants in Effort stereotype condition, feminine gender identity, and low math domain will be abbreviated as E-F-L)

Method

Participants

Participants were recruited through the psychology research pool using the online SONA system and a posting on the psychology research bulletin board. All participants received research credit for their participation in the study. In total, 137 female undergraduate psychology students participated in the study and ranged from 18 to 29 years of age ($M = 19.56$, $SD = 1.87$). The students reported being in years 1-5 of their college careers ($M = 1.99$, $SD = 1.07$).

Materials

Prior to completing a math test, participants were asked to answer a series of questionnaires, including a demographics questionnaire, a measure of domain identification, and a gender identity questionnaire.

Math identity. To measure Math Identity (MI), Smith & White's (2001) Domain Identification Measure (DIM) was used (see Appendix B). The DIM is designed to measure the extent to which an individual self-identifies with a particular domain, for example, math. The DIM had high reliability, particularly for the math domains; Cronbach's α ranged from .64 - .89 for items on the math subscale. Items on the DIM included questions or statements such as "How much is math related to the sense of who you are" or "I have always done well at Math"; participants were asked to rate the extent to which they agreed with the statements using a scale ranging from 1 ("Strongly Disagree" or "Not at all") to 5 ("Strongly Agree" or "Very much").

Gender identity. Spence, Helmreich, and Holahan's (1979) EPAQ (see Appendix C) was used to assess gender identity (GI); this scale is comprised of four scales: positive masculinity (e.g., independent, feels superior, active), negative masculinity (egotistical, greedy, arrogant), positive femininity (helpful, devoted to others, aware of others' feelings), and negative femininity (spineless, whiny, subordinates self to others). The positive scales are used to identify four gender types: masculine (high in masculine traits but low in feminine traits), feminine (high in feminine traits but low in masculine traits, androgynous (high in both masculine and feminine traits), and undifferentiated (low in both masculine and feminine traits). The EPAQ had high reliability, with Cronbach's α for each subscale ranging from .76 to .85 (Helmreich, Spence, & Wilhelm, 1981).

Math test. The math test consisted of 16 problems adapted from selected problems on a practice math SAT downloaded from http://www.collegeboard.com. Following a typical SAT section of the same length, participants were given 20 minutes to complete each math test.

Design

This study consisted of a manipulation of stereotype threat type, along with assessments of Gender Identity (GI) and Math Identity (MI). The levels of stereotype threat condition were **effort** stereotype threat, **innate ability** stereotype threat, and **control** (no threat exposure). Math identity (MI) was a continuous scale, and levels of gender identity (GI) included the four groups of masculine, feminine, androgynous, and undifferentiated.

Procedure

The experimental design was as follows:

Table 2. Experimental Design.

[illegible]	
Questionnaires (Demographics, EPAQ, DIM)	
5- Minute Distractor Task (Hidden Pictures)	
Stereotype Statement	Control Statement
Math Test (20 minutes)	
Questionnaire (Perceptions of Test)	

Participants were tested in groups of 1-10. Each participant was assigned a random four-digit code so that their data could not be personally identified; in addition, each participant was randomly assigned to one of the three stereotype condition groups (effort, innate ability, or control). Participants were not aware of the condition to which they were assigned.

When they arrived for the study, students were each given a booklet of materials in which they were instructed to answer some questionnaires and complete a math test to the best of their ability. Participants first completed the demographics questionnaire (see Appendix A), Smith and White's (2001) DIM, and Spence, Helmreich, and Holahan's (1979) EPAQ. After completing these questionnaires, participants completed a 5-minute hidden pictures task (see Appendix D) in which they were instructed to find some objects hidden in a drawing. The purpose of this task was to prevent any potential priming of gender identity or math identity from influencing performance on the math test.

Depending upon the condition to which they were assigned, participants then read a statement/vignette (see Appendix E) in their instructions in which they were either told that gender differences in math arise due to innate ability, or to the amount of effort that men and women expend on problems, or they read a passage in which no stereotype threat is described whatsoever. These vignettes corresponded to the three threat types manipulated in this study: control, effort stereotype threat, and ability stereotype threat. Participants were then given 20 minutes to complete a 16-item math test (see Appendix F) and asked to show all work for each math problem they attempted in order to prevent guessing. After the math test, participants completed a final questionnaire to assess their perceptions of the math test and to ask their opinions on gender differences in math (see Appendix G). Finally, all participants were debriefed on stereotype threat due to the deception included in the vignettes.

Results

Demographics

Participants' responses on the demographics questionnaire revealed a high degree of variability in math abilities; their SAT scores ranged widely, with 62 participants reporting SAT math scores ranging from 250 to 770 ($M = 565.97, SD = 91.77$). However, 75 participants either reported not remembering their SAT scores, only provided the total amount of all sections, or reported scores that were not possible (were not a multiple of 10). As a result we were not able to utilize SAT scores as a baseline indicator of Math ability.

Participants were also asked to report their strongest and weakest academic subjects from the given choices of Math, Science, English, History, and Foreign languages. Thirty-two participants (23.4%) reported that Math was their strongest subject; the most common subject reported for strongest academic subject was English, with 63 participants (46.0%). In contrast, the most common subject reported for weakest academic subject was Math, with 51 participants reporting such (37.2%).

In addition, participants were asked to report whether or not they were currently taking a math class and the math class they were taking. Thirty-six participants (26.5%) reported taking a math class, and 100 participants (73.5%) reported they were not currently taking a math class (one person did not provide a response). Current math courses included Developmental Math (2.2%), Intermediate Algebra (2.9%), Calculus or Quantitative Methods (5.1%), Statistics (11.4%), other (4.4%), or no math class (73.5%).

Perceptions of the Test

After the test, participants responded to a questionnaire to determine their perceptions of the test; they were asked to report on a scale of 1-5 the extent to which they agreed with a statement (1=strongly disagree, 5= strongly agree). The average ratings (and percentages) for each questionnaire item are provided in Table 3.

Table 3. Reported Perceptions of Math Test.

	Strongly Disagree		*Disagree*		*Neutral*		*Agree*		*Strongly Agree*		*Total*
This test was difficult.	8	5.9%	27	20.0%	55	40.7%	38	28.1%	7	5.2%	135
I did well on this test.	20	14.8%	37	27.4%	51	37.8%	24	17.8%	3	2.2%	135
This test was fair.	1	0.7%	8	6.0%	41	30.6%	66	49.3%	18	13.4%	134
I put a lot of effort into this test.	3	2.3%	22	16.5%	42	31.6%	55	41.4%	11	8.3%	133
I felt I had sufficient knowledge to complete this test.	4	3.0%	24	17.8%	47	34.8%	47	34.8%	13	9.6%	135

Self-Reported Beliefs in Gender Differences in Math

On the final questionnaire, participants were asked to report whether they believed in gender differences in math performance, which direction they believed the gender differences favored, and the potential reasons for gender differences. These results are presented in Table 4. Seventy-six participants (56.3%) reported that gender differences did exist, while fifty-nine participants reported that gender differences did not exist (43.7%). One person did not respond. Participants reported their belief in direction of these gender differences; although 59 of the participants had previously reported that gender differences did not exist, some of these participants did speculate on the reasons for gender differences. Of the 80 participants who responded, 68 participants (85%) believed that men were better than women, and 12 participants (15%) believed that women were better than men. When asked to choose whether Effort, Innate Ability, or Other factor was responsible for these gender differences, 39 of 97 participants (40.2%) believed that gender differences were due to effort, and 42 of 97 (43.3%) believed they were due to innate ability. The remaining 16 of these 97 (16.5%) cited other reasons (which they were asked to specify). Other reasons included an interaction between innate ability and effort, differences in brain functioning (not clearly specified as emanating from innate or experiential differences), or cultural differences.

Table 4. Self-Reported Beliefs in Gender Differences in Math Performance.

Do you believe that there are gender differences in math performance?			
Yes	No	Total	
76 (56.3%)	59 (43.7%)	135	
Who do you believe is better?			
Males	Females	Total	
68 (85%)	12 (15%)	80	
Which do you believe is responsible for these gender differences?			
Effort	Innate Ability	Other	Total
39 (40.2%)	42 (43.3%)	16 (16.5%)	97

Math Identity (MI)

Math Identity (MI) was assessed using Smith and White's (2001) Domain Identification Measure (DIM), with possible scores ranging from 5 - 45. A Cronbach's alpha analysis found a reliability of $\alpha = .719$. Participants' scores on the DIM ranged from 10 - 45 ($M = 28.02$, $SD = 8.7$). The median was 28, while the modal score was 34. There were no significant differences on average between stereotype conditions on identification with math ($F(2,136) = .431$, $p = .651$, $\eta^2 = .006$).

Gender Identity (GI)

Spence, Helmreich, and Holahan's (1979) Extended Personal Attributes Questionnaire (EPAQ) was used to assess gender identity; this questionnaire consisted of four subscales (positive masculinity, negative masculinity, positive femininity, and negative femininity). The possible scores on all four subscales ranged from 8 - 40. For this study, a Cronbach's alpha analysis found a reliability of $\alpha = .751$.

In order to categorize each participant as Masculine, Feminine, Androgynous, or Undifferentiated, the medians for the positive masculine and positive feminine subscales were determined; median splits for each of the positive subscales were created in order to separate groups into high and low masculinity and femininity. Those high on masculinity and low on femininity were considered "Masculine", and those high on femininity and low on masculinity were considered "Feminine." However, if an individual scored high on both masculinity and femininity, they were labeled as "Androgynous;" in contrast, those who scored low on both masculinity and femininity were considered "Undifferentiated." A distribution table of gender identity within stereotype threat condition is presented in Table 5.

Table 5. Gender Identity within Stereotype Conditions.

	Effort		*Innate Ability*		*Control*		*Total*	
Masculine	12	*8.9%*	6	*4.4%*	14	*10.4%*	32	*23.7%*
Feminine	4	*3.0%*	13	*9.6%*	6	*4.4%*	23	*17.0%*
Androgynous	21	*15.6%*	11	*8.2%*	14	*10.4%*	46	*34.1%*
Undifferentiated	7	*5.2%*	16	*11.9%*	11	*8.2%*	34	*25.2%*
Total	44		46		45		135	

Number Attempted

Across all participants, the mean number of problems attempted on the math test was 13.23 (*SD* = 3.02) out of 16 problems. The median number answered was 14, with 54 participants (39.4%) answering all 16, indicating there may have been a ceiling effect. Those in the effort condition had a mean of 12.52 (*SD* = 3.51) attempted, while those in the innate ability and control groups had means of 13.71 (*SD = 2.85*) and 13.44 (*2.55*) attempted, respectively (see Figure 1). An ANCOVA on number of attempts, with stereotype threat condition and gender identity (GI) as between-subjects factors, was run with Math Identity (MI) as a covariate. The three-way interaction between stereotype threat condition, GI, and MI was not run due to insufficient participants in some cells (e.g., there were no participants who had a feminine GI and high MI in the effort condition). The interaction of threat type and GI was not found to be significant ($F(6,135) = 1.594, p = .155, \eta^2 = .073$). Moreover, there were no significant differences, on average, between stereotype conditions ($F(2,136) = 1.878, p = .157, \eta^2 = .028$). The effect size for stereotype conditions was only .028, indicating that only 2.8% of the variability in number of problems attempted is accounted for by stereotype condition.

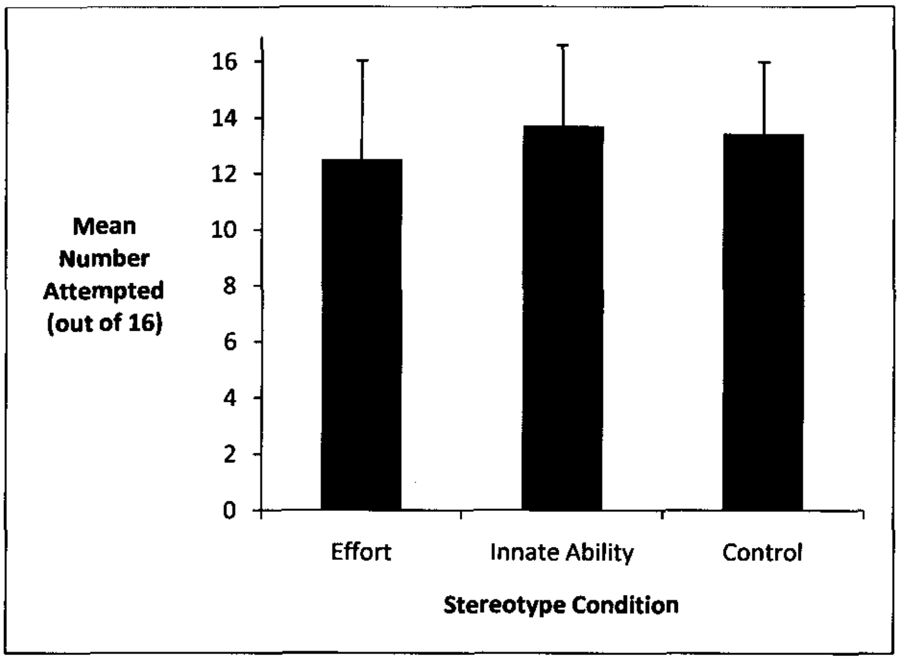


Figure 1. Mean number attempted across stereotype conditions.

In addition, there was no significant difference, on average, among GI groups ($F(3, 134) = .457, p = .713, \eta^2 = .011$). Mean number of items attempted by each GI group are presented in Table 6. There were no significant differences among the stereotype conditions for GI ($F(2, 135) = 1.117, p = .330, \eta^2 = .017$). As math domain identification was measured as a continuous variable, a correlation analysis was run on MI and number attempted; the resulting correlation was not found to be significant ($r(136) = -.009, p = .921$).

Table 6. Mean Number Attempted Among Gender Identity Groups.

	Masculine	*Feminine*	*Androgynous*	*Undifferentiated*
Mean	13.3438	13.7391	12.8667	13.2647
SD	3.04387	2.43492	2.95881	3.36027
N	32	23	45	34

Percentage Correct/Accuracy

Percentage correct was assessed as number of problems correct out of those attempted. Overall, participants varied widely on the tests, scoring within a range of 0% - 93.75% correct (M = 40.01, SD = 20.48). Performance among the stereotype groups varied widely (see Figure 2); those in the effort stereotype group scored a mean of 37.64% (SD = 19.39), those in the innate ability group scored a mean of 41.30% (SD = 20.36), and those in the control group scored a mean of 41.11% (SD = 20.36).

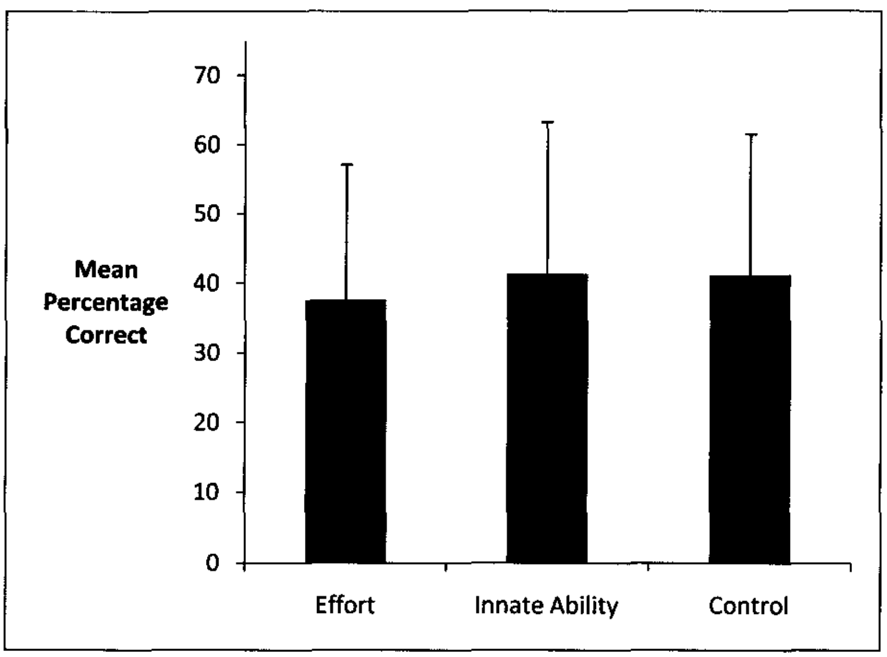


Figure 2. Mean percentage correct across stereotype conditions.

An ANCOVA was conducted on accuracy data, with GI and threat type as between-subjects factors, and MI as a covariate. The interaction of threat type and GI was not found to be significant ($F(6, 134) = .527, p = .787, \eta^2 = .025$). In addition, there was no main effect of threat type ($F(2, 136) = .230, p = .795, \eta^2 = .003$). Gender identity (GI) also did not appear to influence accuracy (see Table 7 for means of percentage correct), on average ($F(3, 134) = 1.264, p = .289, \eta^2 = .029$).

Table 7. Mean Accuracy Across Gender Identity Groups.

	Masculine	*Feminine*	*Androgynous*	*Undifferentiated*
Mean	40.23%	43.21%	36.94%	41.73%
SD	19.24	23.38	19.44	21.25
N	32	23	45	34

Specific analysis of math identity, however, provided some interesting findings. Again, because MI was measured as a continuous variable, a correlation analysis was conducted to examine the relation between MI and accuracy (percentage correct). The correlation was found to be significant ($r(136) = .338, p < .001$). To better understand the effects of math identification, additional correlations were run separately for each stereotype condition. For the effort stereotype group, there was a trend for accuracy to increase as MI grew stronger among participants ($r(46) = .267, p = .073$). For the innate ability and control groups, MI and accuracy were found to be significantly positively correlated (see Table 8 for statistics). Correlations were also computed between MI and accuracy for each GI group and were also found to be statistically significant, except for the androgynous gender identity (see Table 9 for correlations).

Table 8. Correlations of Math Identity and Accuracy Across Stereotype Conditions.

	Effort	*Innate Ability*	*Control*
Pearson Correlation	.267	.363*	.376*
Sig. (2-tailed)	.073	.014	.011
N	46	45	45

Table 9. Correlations of Math Identity and Accuracy Across Gender Identity Conditions.

	Masculine	*Feminine*	*Androgynous*	*Undifferentiated*
Pearson Correlation	.377*	.444*	.184	.571*
Sig. (2-tailed)	.033	.034	.225	.000
N	32	23	45	34

Discussion

Number Attempted

The overall pattern results obtained in this study failed to replicate those of Thoman et al. (2008) and did not support our hypotheses. In general, there did not appear to be differences in number of problems attempted, on average between participants on the basis of stereotype conditions, gender identity groups, or math identity. This was likely due to a ceiling effect. Most participants answered about a 13 items and nearly 40% of all participants answered all 16 questions. Although those in the effort condition had a mean of 12.52 attempted, which was less than that of the innate ability and control groups (13.71 and 13.44, respectively), this difference was not significant. In addition, the effect size of .028 was small, and thus, stereotype condition likely did not contribute to differences in mean number of problems attempted.

In the correlation computed, math identity was not found to be a significantly related to performance as assessed by number attempted. This was surprising, since it is reasonable to expect that people with high degrees of math identification would have strong motivation as well as the ability and desire to perform better than those with low math identification on math tests; therefore they would be expected generally to answer more questions. Again, perhaps the reason that math identity was not found to be related to this measure of math performance was due to the ceiling effect we observed.

There was no significant difference between the gender identity groups for the number of math test items attempted, and the effect size of .011 was trivial. Although it was predicted that those with masculine gender identity would answer the most problems, women with feminine gender identity actually answered the most in our study. However, the differences were not significant, and again, due to the ceiling effect, additional analyses may be needed to assess the effects of gender identity on number attempted.

Because there were only 16 problems on the math test employed in the current study, it is possible that participants felt they needed to attempt all problems; consequently, future studies should include more (and more diverse) problems available for participants to complete. An increased number of problems should be accompanied with the instructions that participants are not expected to complete all problems; this way participants may not feel as if they need to attempt all problems. Furthermore, an increase in number of problems available for participants to complete may also result in a better analysis of the effects of gender identity, math identity, and stereotype condition—if math performance is to be defined as the number of problems attempted. If there is indeed a significant negative correlation between math domain and number of problems attempted, this would suggest that participants are spending more time on problems in order to complete more problems accurately. An explicit measure of motivation or anxiety would also serve to elucidate this relation between these variables.

Percentage Correct

Participants showed wide variation on test performance with respect to percentage correct. In general, the participants performed poorly, as the mean was under 50%. The test consisted of items adapted from a practice SAT section and thus, the participants should have had the knowledge to complete all of the problems. However, many participants indicated they were not currently taking a math class, and it is possible that some knowledge may not have been readily accessible. Most students could not recall their SAT scores, and consequently, it is difficult to gather a baseline of performance.

Because the scores varied so widely, no significant mean differences could be found between the stereotype threat groups. These results differed from those of Thoman et al. (2008). In fact, the mean percentage correct for the effort stereotype group was lower than that of the innate ability and the control groups; however, this was likely due to the large standard deviations found for each stereotype group. The ANCOVA that was conducted indicated that the interaction of threat type and gender identity was not significant, and thus, it appears that the effects of threat type do not depend on gender identity. In addition, there were no main effects of threat type and gender identity.

With respect to math identity however, significant relations were found, indicating that those with higher math identity had increased accuracy, on average. These results are not surprising, as highly-math identified participants are those who likely have the knowledge and desire to perform well on the math test (and consequently do so). Furthermore, math identification appeared to be significantly related to accuracy only for the innate ability and control stereotype groups, but not for the effort stereotype group. Consequently, it seems that under an effort stereotype threat, participants may rely on other aspects of identity, thus reducing the significance of math identity.

In order to better assess the influences of gender identity and stereotype condition, correlation analyses were run on accuracy data. Regarding stereotype threat groups, the innate ability and control groups showed significant positive correlations between accuracy and math identity, indicating that women with higher math identity do perform better than those with lower identification with math in those conditions. However, under effort stereotype threat, there was only a trend towards significance. Under effort stereotype threat, people may subconsciously activate self-percepts such as persistence and become motivated to perform well for the sake of self-confidence; consequently, various aspects of gender identity may also become activated, reducing the influence of math identity but not enough for gender identity to become a significant predictor of math performance. This would be in line with the hypotheses posited at the outset of this study, and are in line with the Thoman et al results.

These results, however, may be skewed by the ceiling effect of number attempted. Since many participants answered the maximum number of questions on the test, differences between the groups on accuracy performance may also have been reduced. Thus, additional studies with an increased number of problems to complete may result in a more accurate view of the influences of threat type, gender identity, and math identity on women's math performance.

Conclusion

Overall, the results did not replicate those of Thoman et al. (2008). This could be due to demographic reasons such as lower mathematical ability and wide variability of the participants in general. Unfortunately, due to patterns of participant response, we were not able to analyze data for this possibility in the current study. The significant effects revealed in this study are small, just as the effects found by Thoman et al. (2008) were also small.

Participants answered less than half of the items correctly on average, even though they attempted most if not all of them; thus, the math test may have been too difficult for them. Furthermore, based upon their questionnaire responses it is evident that many participants knew that they did not perform well (a bit less than half either disagreed or strongly disagreed with the statement "I did well on this test"). Combined with the fact that many participants were not currently taking a math course, it is likely that many of the participants did not feel confident with their knowledge and ability to complete the test well, thus contributing to poor performance and the overall lack of differences between the three stereotype conditions.

Another aspect to be addressed relates to the specific operationalization of variables, such as Math Identification, Gender Identification, and definitions of math performance. The identity variables, in particular, have been widely discussed in the literature but are difficult to measure accurately. Again, in this study, the ranges of scores obtained were not as varied as we had anticipated. These issues limited our analyses and results, and in the future, effort should be made to utilize measures that would provide greater sensitivity to individual differences.

Equally disappointing is the discovery that we may have experienced a ceiling effect of problems attempted in combination with a limited range in math abilities. Indeed, our study does suggest that some aspects of math performance, such as accuracy, may be more prone to identity influence, whereas others, such as number of problems attempted, may not be impacted to the same extent. Perhaps providing a longer test, with items of varied difficulty would eliminate some of these concerns, and would also afford researchers better opportunities to include more performance measures such as response times. These types of tools would allow for more information about participant motivation, hesitancy, and the like.

It is possible that participants had preexisting notions of the stereotype involving women and math performance that negated the stereotype that they were currently under. Further analysis of the descriptive data should enable us to partially explore this possibility. For instance, participant responses to the question about their beliefs about gender differences (and the root causes of these differences) might form the basis of an additional set of analyses exploring the role chronic stereotypes in performance. Furthermore, in our current study, the final questionnaire participants completed revealed that not everyone reported believing in the stereotype condition they were assigned; in fact, some women even reported that they believed that women were better at math than men. As a result, the manipulation (reading a short statement about gender differences in math) may not have been strong enough to influence participants in the intended ways. Measures of chronic attitudes either before or after manipulation of stereotyped beliefs would go a long way to elucidating some of these mentioned problems.

Future research would do well to more carefully assess the manipulation checks and measure chronic or baseline measures of gender stereotypes about math performance. And we are not aware of a single study that has to date explored the interaction of native and primed stereotypes in measures of math performance before, during, and after prime exposure, but this proposed future analysis may shed light on the utility of such a design.

Thoman et al.'s 2008 study indicated that students performed differently on math tests according to the stereotype threat condition in which they were placed, but the present study did not show such clearly-defined differences. Inspection of the data does reveal that participants in different threat conditions may have responded differently. The challenge for subsequent research is to measure how stereotype threat may activate aspects of personality that interact to influence women's math performance.

Future studies may consider a stronger manipulation in order to assess the effects of effort and innate ability stereotype. In addition, the influences of effort and innate ability stereotypes should also be investigated with men to determine if they will respond differently, particularly if they do not fit the cultural stereotype that men are better than women at math. Furthermore, future studies may want to investigate the effects of innate ability and effort stereotype threat at different ages in order to determine how individuals will respond. In particular, children and young adults may be more likely to believe that gender differences are due to effort if they have been told in educational settings that they have the capability and time to improve.

One last set of suggestions emerge in recognizing that as cultural ideas have changed, more women are becoming interested in math and science. As they enter the fields of science, technology, mathematics, and the like, they still undoubtedly face strong stereotypes. Thus, exploring the interactions of gender identity, math identity, and different stereotype conditions across different ages, as well as with men, will provide further insight into combating the negative influence of stereotype threat involving women's math performance.

References

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and 'Choking Under Pressure' in math. *Psychological Science, 16*(2), 101-105. doi:10.1111/j.0956-7976.2005.00789.x

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(6), 983-998. doi:10.1037/0278-7393.33.6.983

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*(2), 256-276. doi:10.1037/0096-3445.136.2.256

Benbow, C.P. & Stanley, J.C. (1980). Sex differences in mathematical ability: fact or artifact? *Science, 210(4475)*, 1262-1264. doi:10.1126/science.7434028

Bonnot, V., & Croizet, J.-C. (2007). Stereotype internalization and women's math performance: The role of interference in working memory. *Journal of Experimental Social Psychology, 43*(6), 857-866. doi:10.1016/j.jesp.2006.10.006

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*(7), 572-578. doi:10.1111/j.0956-7976.2005.01577.x

Cullen, M. J., Waters, S. D., & Sackett, P. R. (2006). Testing stereotype threat theory predictions for math-identified and non-math-identified students by gender. *Human Performance, 19*(4), 421-440. doi:10.1207/s15327043hup1904_6

Helmreich, R.L., Spence, J.T., & Wilhelm, J.A. (1981). A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles, 7*(11), 1097-1108. doi:10.1007/BF00287587

Hyde, J.S. & Mertz, J.E. (2009). Gender, culture, and mathematics performance. *PNAS, 106*(22), 8301-8307. doi:10.1073/pnas.0901265106

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*(3), 175-179. doi:10.1111/j.0956-7976.2005.00799.x

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*(2), 437-444. doi:10.1016/j.jesp.2007.01.007

Kiefer, A., & Shih, M. (2006). Gender differences in persistence and attributions in stereotype relevant contexts. *Sex Roles, 54*(11), 859-868. doi:10.1007/s11199-006-9051-x

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*(5), 825-832. doi:10.1016/j.jesp.2006.08.004

Krendl, A. C., Richeson, J. A., Kelley, W. M., & Heatherton, T. F. (2008). The negative consequences of threat: A functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math. *Psychological Science, 19*(2), 168-175. doi:10.1016/j.jesp.2006.08.004

Lesko, A. C., & Corpus, J. H. (2006). Discounting the Difficult: How high math-identified women respond to stereotype threat. *Sex Roles, 54*(1), 113-125. doi:10.1007/s11199-005-8873-2

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: the effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*(2), 236-243. doi:10.1016/j.jesp.2005.04.010

Miller, H., & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personality and Individual Differences, 37*(3), 591-606. doi: 10.1016/j.paid.2003.09.029

Parsons, J.E. (1983). Attributional processes as mediators of sex differences in achievement. *Journal of Education Equity and Leadership, 3*(1), 19-27. Retrieved from http://www.rcgd.isr.umich.edu/garp/articles/ecclesparsons83a.pdf

Pinker, S. (2005). Sex ed. *New Republic, 232*(5), 15-17.

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*(1), 55-71. doi:10.1111/0022-4537.00201

Rosenthal, H. E. S., Crisp, R. J., & Suen, M.-W. (2007). Improving performance expectancies in stereotypic domains: task relevance and the reduction of stereotype threat. European *Journal of Social Psychology, 37*(3), 586-597. doi:10.1002/ejsp.379

Rydell, R. 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*(5), 949-966. doi:10.1037/a0014846

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194-201. doi:10.1006/jesp.2001.1500

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440-452. doi:10.1037/0022-3514.85.3.440

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: identity salience and shifts in quantitative performance. *Psychological Science, 10*(1), 80-83. doi:10.1111/1467-9280.00111

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*(6), 1040-1057. doi:10.1177/00131640121971635

Spelke, E.S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: a critical review. *American Psychologist, 60*(9), 950-958. doi:10.1037/0003-066X.60.9.950

Spence, J.T., Helmreich, R.L., & Holahan, C.K. (1979). Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors. *Journal of Personality and Social Psychology, 37*(10), 1673-1682. doi:10.1037/0022-3514.37.10.1673

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28. doi:10.1006/jesp.1998.1373

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811. doi:10.1037/0022-3514.69.5.797

Steele, J. R., & Ambady, N. (2006). 'Math is Hard!' The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology, 42*(4), 428-436. doi:10.1016/j.jesp.2005.06.003

Taylor Jr., S. (2005). Males, females, and math: the evidence. *National Journal, 37*(9), 585-586.

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*(9), 702-712. doi:10.1007/s11199-008-9390-x

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*(3), 792-799. doi:10.1016/j.jesp.2007.07.005

Appendices

Appendix A: Demographics Questionnaire

Please fill in or answer each question below. Your data will remain confidential and will only be identified by your individual participant code.

Information about yourself:

Age: ________

Year at Seton Hall (circle one): 1st 2nd 3rd 4th Other ________

Do you have corrected vision? Y or N

If so, are you wearing your glasses/contacts? Y or N

Are you currently sick with an illness or taking any medication that affects your vision, level of attention, or other cognitive abilities? Y or N

Do you have dyslexia or any other conditions that may affect your ability to read the questions in this study?

Y or N

Ethnicity:

Asian American _______

African American _______

Caucasian _______

Hispanic/Latino _______

American Indian _______

Other _______

Are you a U.S. citizen? If not, what is your country of origin? ____________________

Academics

Please indicate your strongest subject (choose only one):

Math _______ Science _______ English _______ History _______ Foreign Languages _______

Please indicate your weakest subject (choose only one):

Math _______ Science _______ English _______ History _______ Foreign Languages _______

Please indicate your scores on the Scholastic Aptitude Test (SAT) or the American College Test (ACT) or both, if applicable:

SAT:

Math ________

Verbal ________

Writing ________

ACT:

Math ________

Science ________

Reading Comprehension ________

Writing ________

Are you currently taking a math or statistics class? Y or N

If so, which one? ____

Appendix B: Smith and White's Domain Identification Measure

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 nor Agree	Moderately Agree	Strongly Agree

1. ______ I learn things quickly in English classes
2. ______ Mathematics is one of my best subjects
3. ______ English is one of my best subjects
4. ______ I get good grades in English
5. ______ I have always done well in Math
6. ______ I'm hopeless in English classes
7. ______ I get good grades in Math
8. ______ I do badly in tests of Mathematics

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

9. ______ How much do you enjoy math-related subjects?
10. ______ How much do you enjoy English-related subjects?
11. ______ How likely would you be to take a job in a math related field?
12. ______ How much is Math related to the sense of who you are?
13. ______ How important is it to you to be good at Math?
14. ______ How important is it to you to be good at English?

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

1	2	3	4	5
Very poor	Poor	About the same	Better than average	Excellent

15. ______ Compared to other students, how good are you at math?
16. ______ Compared to other students, how good are you at English?

Appendix C: Spence, Helmreich, & Holahan's (1979) Extended Personal Attributes Questionnaire (EPAQ)

Following is a list of word pairs and a scale of 1-5 for each pair. Please examine the pairs and circle the number that describes where you fall on the scale. Please do not skip any pairs.

not at all aggressive				very aggressive
1	2	3	4	5
not at all independent				very independent
1	2	3	4	5
not at all spineless				very spineless
1	2	3	4	5
not at all helpful				very helpful
1	2	3	4	5
not at all dominant				very dominant
1	2	3	4	5
not at all self-confident				very self-confident
1	2	3	4	5
not at all servile				very servile
1	2	3	4	5
not at all warm to others				very warm to others
1	2	3	4	5
not at all competitive				very competitive
1	2	3	4	5
does not cry easily				cries very easily
1	2	3	4	5
not at all gentle				very gentle
1	2	3	4	5
feelings not easily hurt				feelings easily hurt
1	2	3	4	5
not at all emotional				very emotional
1	2	3	4	5

not at all active				very active
1	2	3	4	5
low need for security				high need for security
1	2	3	4	5
does not stand up under pressure				stands up under pressure
1	2	3	4	5
not at all gullible				very gullible
1	2	3	4	5
does not make decisions easily				makes decisions very easily
1	2	3	4	5
not at all arrogant				very arrogant
1	2	3	4	5
not at all devoted to others				very devoted to others
1	2	3	4	5
does not subordinate self to others				subordinates self to others
1	2	3	4	5
gives up very easily				never gives up easily
1	2	3	4	5
not at all boastful				very boastful
1	2	3	4	5
not at all kind				very kind
1	2	3	4	5
not at all whiny				very whiny
1	2	3	4	5
not at all understanding				very understanding
1	2	3	4	5
not at all egotistical				very egotistical
1	2	3	4	5

not at all complaining				very complaining
1	2	3	4	5
not at all greedy				very greedy
1	2	3	4	5
not at all dictatorial				very dictatorial
1	2	3	4	5
not at all fussy				very fussy
1	2	3	4	5
not at all cynical				very cynical
1	2	3	4	5
not at all nagging				very nagging
1	2	3	4	5
does not feel superior				feels very superior
1	2	3	4	5
does not look out only for self				looks out only for self
1	2	3	4	5
not at all aware of others' feelings				very aware of others' feelings
1	2	3	4	5
not at all hostile				very hostile
1	2	3	4	5

Appendix D: Hidden Pictures Task

Find these objects in the picture below.

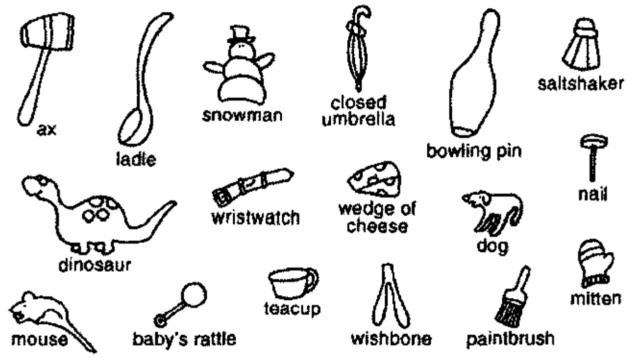


Image from
http://www.highlightskids.com/GamesandGiggles/HiddenPics/HiddenPicsPrintable/h8hiddenArchive.asp

Appendix E: Vignettes

Effort

In previous studies, researchers have found that gender differences exist for certain types of math problems, with men performing better on math tests than women. The gender differences are not due to innate ability or men being born better at math than women (Hyde & Mertz, 2009; Spelke, 2005), but are due to the amount of effort put into finding the solutions (Pajares & Miller, 1994, Parsons, 1983). Consequently, we are interested in determining which types of problems produce greater gender differences.

References

Hyde, J.S. & Mertz, J.E. (2009). Gender, culture, and mathematics performance. *PNAS, 106*(22), 8301-8307.

Pajares, F. & Miller, M.D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: a path analysis. *Journal of Educational Psychology, 86*(2), 193-203.

Parsons, J.E. (1983). Attributional processes as mediators of sex differences in achievement. *Journal of Education Equity and Leadership, 3*(1), 19-27.

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

Innate Ability

In previous studies, researchers have found that gender differences in certain types of math problems exist, with men performing better on math tests than women. Effort is not the cause of these gender differences—these differences are due to innate ability, or men simply being born better at math than women (Benbow & Stanley, 1980; Pinker, 2005; Taylor, 2005). Thus, we are interested in assessing which types of math problems result in significant gender differences.

References

Benbow, C.P. & Stanley, J.C. (1980). Sex differences in mathematical ability: fact or artifact? *Science, 210*(4475), 1262-1264.

Pinker, S. (2005). Sex ed. *New Republic, 232*(5), 15-17.

Taylor Jr., S. (2005). Males, females, and math: the evidence. *National Journal, 37*(9), 585-586.

Control

The purpose of this experiment is to investigate how people solve math problems.

Appendix F: Math Test

Please complete the following test to the best of your ability. Show all work. Items without work will not be scored.

1. If the function f is defined by $f(x) = 4x + 5$, then $2f(x) + 5 =$
   (A) $4x + 10$
   (B) $4x + 15$
   (C) $6x + 10$
   (D) $8x + 10$
   (E) $8x + 15$

2. On Monday, Joe walked 3 miles in 45 minutes. If he walked for 1 hour and 15 minutes at this rate on Wednesday, how far did he walk on Wednesday?
   (A) 3.5 miles
   (B) 4 miles
   (C) 4.5 miles
   (D) 5 miles
   (E) 6 miles

3. In a certain store, the regular price of a television is $600. How much money is saved by buying this television at 20% off the regular price rather than buying it on sale at 10% off the regular price with an additional discount at 10% off the sale price?
   (A) $6
   (B) $10
   (C) $54
   (D) $60
   (E) The discounts are the same.

4. In a vote for favorite flavor of ice cream, 720 votes were cast for 2 flavors, chocolate and vanilla. If chocolate won by a ratio of 8 to 4, what was the number of votes cast for vanilla?
   (A) 90
   (B) 180
   (C) 240
   (D) 360
   (E) 480

5. Line ℓ intersects the x-axis at $x = 5$ and the y-axis at $y = -4$. If line m passes through the origin and is perpendicular to line ℓ, what is the slope of line m?

(A) $-\frac{5}{4}$

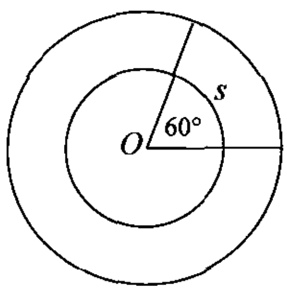
(B) $-\frac{4}{5}$

(C) $\frac{4}{5}$

(D) $\frac{5}{4}$

(E) 1

6. Point O is the center of both circles in the figure below. If the circumference of the large circle is 18π, and the radius of the small circle is half of the radius of the large circle, what is the length of the indicated arc s?



Note: Figure not to scale

(A) 0.75π

(B) 1.5π

(C) 2π

(D) 2.25π

(E) 3π

7. The height of a right circular cylinder is 7 and the diameter of its base is 8. What is the distance from the center of one base to a point on the circumference of the other base?

(A) 15

(B) $\sqrt{113}$ (approximately 10.63)

(C) $\sqrt{76.25}$ (approximately 8.73)

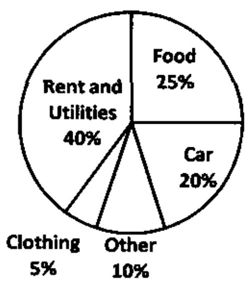
(D) $\sqrt{65}$ (approximately 8.06)

(E) 7.5

8. The average (arithmetic mean) of a and b is 7, and the average of c , d , and e is 16. What is the average of a, b, c, d, and e?
(A) 10.6
(B) 11
(C) 11.5
(D) 12
(E) 12.4

9. If $7 + x$ is 3 less than 7, what is the value of $3x$?
(A) – 33
(B) – 9
(C) – 3
(D) 9
(E) 33

10. The circle graph below shows how Marie's monthly expenses are divided. If Marie spends \$840 per month for rent and utilities, how much does she spend per month for food?



(A) \$210
(B) \$336
(C) \$360
(D) \$420
(E) \$525

11. Eight more than three times a number is equal to four less than five times the number. What is the number?
(A) -3
(B) -0.5
(C) 2
(D) 3
(E) 6

12. If 3 times a number is equal to $\frac{4}{5}$, what is the number?

(A) $\frac{4}{15}$

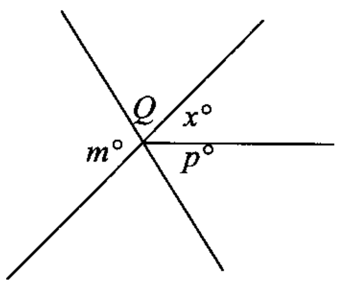
(B) $\frac{5}{12}$

(C) $\frac{4}{3}$

(D) $\frac{12}{5}$

(E) $\frac{15}{4}$

13. In the figure below, lines ℓ and k intersect at point Q. If $m = 123$, and $p = 67$, what is the value of x?



Note: Figure not drawn to

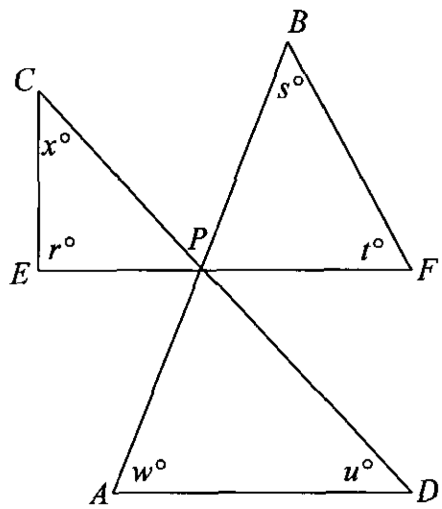
(A) 46
(B) 56
(C) 64
(D) 65
(E) 66

14. In the system of equations below, what is the value of $x + y$?

$$x + y + 5z = 780$$
$$x + y + 3z = 560$$

(A) 110
(B) 220
(C) 230
(D) 450
(E) 460

15. In the figure below, $\overline{AB}$, $\overline{CD}$, and $\overline{EF}$ intersect at P. If $r = 90$, $s = 75$, $t = 55$, $u = 65$, and $w = 45$, what is the value of x?



Note: Figure not drawn to scale.

(A) 30
(B) 40
(C) 60
(D) 70
(E) It cannot be determined from the information given.

16. For a party, Mrs. Jones bought two six-packs of lemon-lime soda, four four-packs of root beer, and three 12-packs of cola. How many total cans of soda did she buy?

(A) 58
(B) 60
(C) 64
(D) 66
(E) 72

Appendix G: Post-Test Questionnaire

On a scale of 1 to 5, please rate the degree to which you agree with the following statements.

	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
This test was difficult					
I did well on this test.					
This test was fair.					
I put a lot of effort into this test.					
I felt I had sufficient knowledge to complete this test.					

While taking the test, did you feel anxious at any time? Y or N

Do you believe that there are gender differences in math performance? Y or N

If so, who do you believe is better?

_______ Males _______ Females

Which do you believe is responsible for these gender differences?

_______ Effort _______ Innate Ability _______ Other (please specify below)