

4-9-2013

# Does Teaching Women About Stereotype Threat Reduce Its Effects on Math Performance?

Nicole Fuller

*The College at Brockport*, nicolefuller5@gmail.com

Follow this and additional works at: <http://digitalcommons.brockport.edu/honors>

 Part of the [Psychology Commons](#), and the [Science and Mathematics Education Commons](#)

---

## Repository Citation

Fuller, Nicole, "Does Teaching Women About Stereotype Threat Reduce Its Effects on Math Performance?" (2013). *Senior Honors Theses*. 73.

<http://digitalcommons.brockport.edu/honors/73>

This Honors Thesis is brought to you for free and open access by the Master's Theses and Honors Projects at Digital Commons @Brockport. It has been accepted for inclusion in Senior Honors Theses by an authorized administrator of Digital Commons @Brockport. For more information, please contact [kmyers@brockport.edu](mailto:kmyers@brockport.edu).

Does Teaching Women About Stereotype Threat Reduce Its Effects on Math Performance?

A Senior Honors Thesis

Submitted in Partial Fulfillment of the Requirements  
for Graduation in the College Honors Program

By  
Nicole Fuller  
Psychology Major

The College at Brockport  
April 9, 2013

Thesis Director: Dr. Laurel McNall, Associate Professor, Psychology

*Educational use of this paper is permitted for the purpose of providing future students a model example of an Honors senior thesis project.*

**Abstract**

Stereotype threat occurs when people have anxiety in a situation where they fear may conform to a negative stereotype about their group, which subsequently causes decreased performance. One well-known stereotype states women are not as skilled at math as men. The goal of the current study was to test how to reduce the effects of stereotype threat on women's performance in mathematics. Fifty male and female college students completed a math test in either a control condition or in one of two stereotype threat reducing conditions (self-affirmation or teaching). Surprisingly, the results show there were no differences in test performance between the three conditions. This is contradictory to previous research; however more research is necessary to find ways to reduce stereotype threat.

### Does Teaching Women About Stereotype Threat Reduce Its Effects on Math Performance?

A stereotype is when people characterize a group of people as having certain characteristics just because they belong to that group. Stereotype formation is based on the concept of ingroups and outgroups; an ingroup is the group that people belong to and feel a sense of identity with. Outgroups are the groups that people feel like they do not belong to or have an identity with. According to the outgroup homogeneity effect, it is believed that people in outgroups are collectively more similar than people in their own ingroup. Stereotypes can lead to negative effects when it comes to performance and this is known as stereotype threat (Kassin, Fein & Markus, 2011).

Stereotype threat is when people have anxiety or concern in situations where there is a chance they may conform to a negative stereotype about their group. This concern can cause interference because individuals are so concerned about not conforming to the negative stereotype, they end up performing badly and in return conforming to the stereotype they wanted to avoid in the first place. Thus, their poor performance is due to knowing about the negative stereotype and the worry about conforming to it distracts them from the task at hand (Kassin, Fein & Markus, 2011).

One area where stereotype threat has been studied is in relation to women and math performance. It has been documented that women underperform on tests of math ability when compared to men when they become aware of the stereotype that women are not as good at mathematics as men (Spencer, Steele, and Quinn, 1999). This topic is very important in the field of both psychology and education; in psychology it is important because it shows that people's behavior could be greatly affected solely by awareness of a negative stereotype. Stereotype threat is important in an educational setting because data from College Board shows that females

are performing worse than men on the math portion of the SAT. For 2012 college bound seniors, men performed 33 points higher than women on the math portion of the test, however on the reading portion men only scored five points higher (Total Group Profile Report, 2012). This is not a new phenomenon; every year since 1972, men outscored women on the math portion of the SAT by at least 31 points (Total Group Profile Report, 2012). One reason that this may be occurring is due to the well-known, but false, stereotype that women are not as good at math as men. The current research study is attempting to determine if the effects of stereotype threat can be eliminated through teaching women about stereotype threat.

### **Theoretical Perspectives on Stereotype Threat**

Stereotype threat theory was defined by Steele and Aronson (1998) as a “predicament” where members of a social group “must deal with the possibility of being judged or treated stereotypically, or of doing something that would confirm the stereotype” (p. 401). Stereotype threat theory is important for stigmatized social groups because these groups may underperform in certain testing contexts where a relevant stereotype about the social group is emphasized. Crocker and Major (1989) define a stigmatized social group as “belonging to a social category about which others hold negative attitudes, stereotypes, and beliefs” (p. 609). The main hypothesis for why stereotype threat exists is due to performance interference. Performance interference predicts that in a stereotype threatening context, individuals in stereotyped social groups perform worse on a measure that is being evaluated than if they were taking the same measure in a non-stereotype threatening context.

According to primary researchers in the field, there are some conceptual moderators that may change the effect of stereotype threat. The first was proposed by Steele (1997); he stated that how relevant a stereotype is in the test setting may affect the degree of stereotype threat one

feels. There are three ways that stereotypes can be made relevant in testing situations: blatant, moderately explicit, or indirect and subtle. The blatant style explicitly tells members of a stereotyped group about their group's inferior cognitive ability and/or ability performance. An example of this would be if researchers told participants that men score better than women on the measure they were taking. The moderately explicit style tells participants about differences in cognitive ability and/or performance ability in test directions or the context of testing, but participants have to interpret the direction of the differences. For example, the directions of a test could state that men and women perform differently on the test they are about to take. In this condition, the researchers are explicitly stating that differences exist, but it is up to the participants to individually decide what the differences mean (Nguyen & Ryan, 2008).

Lastly, the indirect and subtle style says that differences in ability are not directly stated to participants but the context of tests, participants' subgroup membership, or test taking experience is manipulated. An example of this would be emphasizing that the test is being used for diagnostic purposes or researchers could stress that the test is being taken for evaluative purposes. According to Bargh (1997), stereotype threat theories predict that explicit cues would have a smaller effect on negative performance than implicit cues. This was shown in a study by Kray, Thompson, and Galinsky (2001); female participants performed better on the task when an explicit cue was given that stated "male and female students have been shown to differ in their performance on this task" (p. 949). This is known as the stereotype reactance effect; Kray et al. stated that explicit cues may have been seen by participants as a limit to their freedom and ability to perform which in turn causes the participants to behave inconsistently with the stereotype.

According to Steele et al. (2002), another conceptual moderator for stereotype threat is domain identification. It is theorized that only the members of a social group who strongly

identify with a domain where there is a negative stereotype are the ones who are the most susceptible to confirming to it. It is only these individuals because due to their strong identification to the domain, part of their self-regard depends on how one performs in the domain (Steele et al., 2002). Therefore, the reverse would also be true; participants who do not have any personal relevance to the domain, or participants who do not care about performing well, are not negatively affected by stereotype threat.

A last conceptual moderator for stereotype threat is test difficulty (Steele & Aronson, 1995; Steele et al., 2002). In order for stereotype threat to work, the test must be challenging. A difficult test uses more mental resources than an easy test; therefore the interference that is created due to the stereotype threat has more of an effect on performance during a difficult test. With an easy test not as many mental resources are used so the interference does not have as much of an effect (Steele & Aronson, 1995; Steele et al., 2002).

The first empirical test of stereotype threat was done by Steele and Aronson (1995) involving the effect of stereotype threat on African Americans and their performance on intellectual tests. Steele and Aronson hypothesized that when African Americans complete a scholastic or intellectual task they know there is a chance they may be judged according to a negative stereotype about their group's intellectual ability. Steele and Aronson went on to hypothesize that African Americans worrying about being judged then causes impairments in intellectual functioning when they take a standardized test. In the first study, African American and white students take a test that was made up of problems from the verbal GRE practice tests. There was a stereotype threat condition where the test was described as being "diagnostic of intellectual ability" (p. 799), a non-stereotype threat condition it was just described as a "laboratory problem solving task that was nondiagnostic of ability" (p. 799), and a nondiagnostic

challenge condition which was developed to hopefully increase participant's motivation. The results of this study showed that the African American participants performed worse than the white participants when the test was described as being diagnostic, but when the test was nondiagnostic the African American's performance improved greatly.

Three more experiments done by Steele and Aronson (1995) provided even greater insight on how African Americans are affected by stereotype threat. The studies showed that not only do African Americans perform worse on diagnostic tests, they also complete fewer items than whites and take more time completing those items. It was also shown that while under stereotype threat, African Americans think more about the stereotypes about their group and even avoided preferring racially stereotypic activities and styles. Lastly it was shown that just having the stereotype cognitively available (writing your race down) before the test is enough to hinder performance on an intellectual test, even if that test is declared nondiagnostic. If African Americans are affected so greatly by the mere presence of a stereotype about their race then studies similar to these can be replicated to see the impact that stereotype threat has on other groups with well-known stereotypes, such as women and math performance.

### **Empirical Findings on Stereotype Threat Involving Women and Math Performance**

The first research studies done on stereotype threat and women in mathematics were done by Spencer et al. (1999). They performed a series of three different research studies on this topic. In the first study, their goal was to test a pattern that was found in the aforementioned literature; that women underperform when compared to men on difficult tests, but perform the same with men on easy tests. The difficult test was from the advanced GRE exam in mathematics and the easier test was taken from the quantitative section of the GRE general exam. Participants conducted the study in mixed gender groups of three to six and they were

randomly assigned to the type of test. The tests were taken by the participants on computers and it recorded the participant's responses and how long they took to do each problem. The results showed that women underperformed when compared to men on the difficult test but did just as well as men on the easy test.

In the second study, the researchers tested the effects of stereotype by giving all the participants a difficult math exam but this time varied whether the gender stereotype was relevant to performance. This was done by changing how the test was represented; in the relevance condition they told participants the test had shown gender differences in the past and in the condition where the stereotype was supposed to be irrelevant participants were told that the test had not shown any gender differences. The difficult test used in study one was divided into two halves. Half the participants were told that the first half showed gender differences and that the second half showed no gender differences and the other half of participants were told the opposite. The results showed that when participants were told that the test had gender differences women under performed in relation to men but when they were told that gender differences did not exist they performed at the same level as men.

The third study wanted to replicate the effect of the second study but with a less highly selected sample, on a test with a wider range of problems, and with a control group with no mention of gender differences. In the previous two studies, the participants were required to have taken at least one semester of calculus and scored above the 85<sup>th</sup> percentile on the SAT; however in this study, the participants who were selected scored between 400 and 650 on the SAT and completed less than a year of calculus. The study wanted to see if women's lower performance expectations could be the cause of their underperformance instead of stereotype threat. The study measured participants' evaluation apprehension, state anxiety, and self-

efficacy. First the participants were read the instructions; the no gender differences condition was the same as it was in the second study and for the control condition the participants were given no information about gender differences on the test. The participants then completed a short questionnaire that had questions on evaluation apprehension, self-efficacy, and the state-trait anxiety index and then they completed the math test. The results showed that women underperformed relative to men in the control condition, but performed the same in the no gender difference condition. It also showed that the mean for women in the control condition was lower than each of the other mean and that none of the other means were different from each other. The results also determined that self-efficacy and evaluation apprehension are not likely to be mediators of the effect of stereotype threat and for anxiety no definite conclusions could be drawn; even though anxiety is not shown as a mediator in this study, it is still possible that anxiety does play a key role in stereotype threat.

This study by Spencer, Steele, and Quinn (1999) led to many other studies being done on women and math; one of these studies was by Cadinu, Maas, Rosabianca and Kiesner (2005). The researchers wanted to see how interfering thoughts were created spontaneously in a stereotype threat condition versus a condition with no stereotype threat. To test this, a group of women completed a math test and they were either in a stereotype threat condition or in a no threat control condition. In the test instructions, the stereotype threat condition participants were told that the test showed differences in the scores of men and women and participants in the control condition were told that there were no differences between men and women. The math test was the same for both conditions and consisted of seven math problems. Before each problem there was a blank page where participants were told to write anything that came to their mind and judges coded the statements.

The results of the study showed that there were a significantly lower number of correct responses in the stereotype threat condition than in the control condition. The performance deficit in the stereotype threat group occurred after the first trial and there was a steady decline in performance. Also, participants in the stereotype threat condition had significantly higher amounts of negative math related thoughts. Therefore, the decrease in performance was due to increasing negative domain specific thinking in the stereotype threat condition. This study shows that stereotype threat led to negative thoughts which then led to a decrease in performance, supporting previous research that showed performance deficits are caused by intrusive thoughts that occur during the task.

Implicit gender-math stereotypes are when people non-consciously associate men with math more than they do women. This may cause women to identify with math less, have less favorable attitudes in math, and even have lower performance in math. A study by Kiefer and Sekaquaptewa (2007) attempted to examine if implicit gender-math stereotypes, implicit gender identification, and implicit math identification all moderated stereotype threat effect in women's math performance. The researchers thought women who had strong implicit gender-math stereotypes would perform better under control conditions but not under stereotype threat conditions. To conduct this experiment, a group of all female participants completed a math test first and this is where the manipulation occurred, the participants were told the test was either diagnostic (stereotype threat condition) or non-diagnostic. After the math test, participants completed the implicit association test (IAT) and explicit stereotype measure.

The results of the IATs show that participants associated men with math more than women, they showed implicit gender identification, and they did not show implicit identification with math. The results of the math test showed women's math scores were lower in the threat

condition than in the reduced threat condition. For women with strong implicit gender-math stereotypes, their belief is held so strongly that reduced threat instructions have no effect on them. But women with weak implicit gender-math stereotypes can be influenced by the instructions that eliminate the threat (non-diagnostic condition) and therefore they can do well on the test. However, in the stereotype threat condition both strong and weak implicit gender-math stereotype women perform at the same low level.

In a study by Martens, Michael, Greenberg, and Schimel (2006), the effect of self-affirmation on women's intellectual performance was studied. Self-affirmation theory states that maintaining self-integrity and self-worth is a very important source of motivation. Stereotype threat may result when a person's sense of self-integrity is threatened because the stereotype is describing the person's inferiority or incompetence. Martens et al (2006) wanted to study if self-affirmation would help women perform better in math when they are under a stereotype threat condition that says women are inferior to men in math. The females were divided into three conditions: non-diagnostic test control condition, a stereotype threat condition, and a stereotype threat/self-affirmation condition. Self-affirmation was manipulated by having participants in that condition rank 11 characteristics or values of personal importance and then write about why item number one was personally important and to describe a time when it had been important to them.

The results of the study showed that women in the stereotype threat condition performed worse when compared to the women in the control condition and men in the stereotype threat condition. Additional analysis revealed that women under stereotype threat who self-affirmed performed better than women in the pure stereotype threat condition. Also, women under stereotype threat who self-affirmed before the test were able to correctly answer as many questions as the women in the control condition.

Johns, Schmader, and Martens (2005) conducted a study testing whether teaching women about stereotype would reduce the negative effects. In this research study they had three conditions: a problem solving condition, math test condition, and teaching intervention condition. In the teaching-intervention condition the instructions to participants were the same as in the math test condition and the instructions said that participants may feel anxiety but it may be the result of the stereotype and does not necessarily have anything to do with their performance on the test.

In the problem solving condition, men and women performed equally. In the math test condition women did not perform as well as men, however in the teaching-intervention condition women performed equally to men in that condition and also performed equally as well as women in the problem solving condition. Lastly, women in the teaching intervention condition performed better than the women in the math test condition. This study suggests that learning about the negative stereotype and being aware that it exists reduces the feelings of anxiety and in return helps to reduce the negative effects of stereotype threat. Anxiety is reduced because the women were taught that their anxiety had to do with their fear of conforming to a negative stereotype and not their ability to perform on the tests.

### **Current Study**

The purpose of the current study is to replicate and extend on previous research on reducing the effects of stereotype threat. Research by Johns, Schmader, and Martens (2005) demonstrated that teaching about stereotype threat reduces its effects when women are taking a math test. Teaching is thought to prevent decreased performance because women can now attribute their anxiety when taking the test to the stereotype threat and not on their ability to perform on the test. Due to teaching the participants about stereotype threat, when they are

taking the test and begin to feel anxious, they may remember what they were taught and that the anxious feelings they have are due to stereotype threat rather than their ability to perform. This research is also attempting to replicate the research by Martens, Michael, Greenberg, and Schimel (2006) which showed that stereotype threat could be reduced in the short term by self-affirming before taking a math test. It is thought that a person's sense of self-integrity is what is threatened when someone is negatively affected by stereotype threat; self-affirming helps to solidify a person's feelings of self-integrity which therefore eliminates the negative effects of stereotype threat in the short term.

*Hypothesis 1:* Participants in the teaching and self-affirmation conditions will perform better on the math test than the participants in the general stereotype threat condition at Time 1.

Even if teaching about stereotype threat and self-affirmation reduce the effects of stereotype threat, a question remains about how long these interventions will last. In their original research, Johns, Schmader, and Martens (2005) made a prediction that teaching about stereotype threat provides an attributional cue that goes beyond the immediate testing situation. It is believed that by teaching participants about stereotype threat, it allows them to apply the knowledge to other situations and not just the initial test. When they are taking any test and begin to feel anxious, they may remember what they were taught and that the anxious feelings are due to stereotype rather than their ability to perform. Research by Martens, Michael, Greenberg, and Schimel (2006) found that inducing self-affirmation in participants reduces stereotype threat only in the immediate situation. However, it was speculated that the feeling of self-affirming from one time point could not be carried over (and thus unable to affect performance) at a later time. One reason self-affirmation may not be an effective intervention

over time is that the feelings that one gets from self-affirming are likely only temporary and will not automatically come back during Time 2 of testing three weeks later.

*Research question 1:* Will participants in the teaching condition perform better on the math test during Time 2 compared to the self-affirmation and general stereotype threat conditions?

## **Methods**

### **Participants**

The study began with 96 participants who completed Time 1 of testing. Participants were both male and female and were recruited from Dr. McNall's psychology classes held at The College at Brockport. Dr. McNall had three classes and each class was a different condition (teaching, self-affirmation, or control). All students were allowed to participate, however the data from participants who were of certain races and ethnic backgrounds was not used because these participants may be a victim of another form of stereotype threat which would interfere with this study. This follows research by Johns, Schmader, and Martens (2005), who found gender differences in math scores only in Caucasians.

In total, four matched sets of data were excluded due to the participants being a race other than Caucasian (one from the control group, one from the self-affirmation condition, and two from the teaching condition). Excluding participants who were a race other than Caucasian was done to be consistent with the research.

The data was also only used if it could be matched up exactly from Time 1 to Time 2. Data was matched up by participants providing their birthdate (month and day) on the answer sheet. In the control group, there was Time 1 data from 14 participants that could not be matched up to Time 2 data, so all these participants were excluded. In the self-affirmation group

there was Time 1 data from 13 participants that could not be matched up to Time 2 data, so this data was also excluded. Lastly, in the teaching condition, there was Time 1 data from 10 participants that could not be matched up to data from Time 2 and this data was excluded. The Time 2 data was also excluded if it was not able to be paired up with data from Time 1, in the control group this included data from five participants, in the self-affirmation group 10 participants, and in the teaching group eight participants. The final analysis included data from 50 participants whose data was matched up between Time 1 and Time 2 of testing. The control condition included 17 participants (6 male, 11 female). Twelve participants (4 male, 8 female) were included in the self-affirmation condition, and 21 participants (1 male, 20 female) were included in the teaching condition.

### **Materials**

A single math test was created for the study (See Appendix A). The math test consisted of 10 multiple choice problems that were randomly selected from two GRE quantitative practice exams in the GRE Math Workbook (Madore, 2009). Taking problems from the GRE practice book ensured the problems were difficult for this population, but also not too difficult that the test would be perceived as impossible. For Time 2 of testing, the test had the same problems, but they were in a different order (See Appendix B). The participants were also given an answer sheet to record their answers and demographic information (See Appendix C). Participants were asked to record their birthday (month and year), sex, and race on their answer sheet; birthday information was used to match data from both days of testing and race identification was used to exclude any participants whose data may interfere with the validity of the study. The participants in the self-affirmation condition were given a worksheet before they took the math test; this

worksheet was intended to have the participants self-affirm before the test during Time 1 of testing (See Appendix D).

### **Design**

The current study was a 2 x 3 design. The first independent variable was gender and the two levels were male and female. The second independent variable was condition of the test and the three levels were control, self-affirmation, and teaching about stereotype threat. The first hypothesis in the study involved a between subjects design because the data from the three conditions of the second independent variable are being compared to see if there are differences in the dependent variable, number correct on the math test. In addition, the scores of the males and the females are being compared in order to verify that the women suffered from stereotype threat and to verify that stereotype threat was reduced in the women in the self-affirmation and teaching about stereotype threat conditions. However, I utilized a repeated measures design for the research question because the participants within each condition are being tested twice and their scores are being compared across the two tests in order to see if either of the stereotype threat reducing strategies eliminates the threat in the long term.

### **Procedure**

As mentioned before, each class was a different condition and the participants were tested on two different occasions, approximately three weeks apart. At Time 1, the two intervention conditions were implemented and all three conditions were required to take a math test. During Time 2 of testing three weeks later, all the participants received the same instructions and math test. There were no interventions of any kind at Time 2 to determine if the effects of the teaching and self-affirmation conditions go beyond the immediate testing situation.

At Time 1, participants in each condition completed a math test and were asked to disclose some demographic information on their answer sheet (See appendices A and C). The participants in the control condition were under stereotype threat in order to see if the intervention conditions really reduce the effect of stereotype threat. In order to create a stereotype threat condition the participants were told, “the problems that you will be working on are a direct measure of your intelligence and you should make a strong and genuine effort on the test.” This way to induce stereotype threat was taken from a previous study done by Martens, Michael, Greenberg, and Schimel (2006). After the participants were given the instructions they took the math test.

The participants in the self-affirmation condition first filled out a form that was used to make the participants self-affirm (See Appendix D). This form was developed from research by Martens, Michael, Greenberg, and Schimel (2006). There was a list of 6 characteristics the participants were asked to rank in order of personal importance. Such characteristics included sense of humor, creativity, physical attractiveness, social skills, etc. After they ranked those characteristics they were asked to write about why their most valued characteristic was personally important and then they needed to write about a time that it was important to them. Completing this form allowed participants to self-affirm which hopefully then allowed them to perform better on the subsequent math test. After they completed the self-affirmation form, the participants in this condition were given the same instructions as in the control condition and then they completed the same math test as the participants in the control condition.

In the third condition, the researcher taught the participants about stereotype threat. The researcher told them, “the problems that you will be working on are a direct measure of your intelligence and you should make a strong and genuine effort on the test. It is also important to

keep in mind that if you are feeling anxious while taking this test, this anxiety could be the result of the negative stereotypes that are known in society and have nothing to do with your actual ability to do well on the test.” The script for teaching the participants about stereotype threat was taken directly from research done by Johns, Schmader, and Martens (2005). After the researcher read the script to the participants, they took the same math test that the other conditions completed. Participants in all three conditions were given 15 minutes to complete the test, once time was up the researcher collected the tests from all the participants and told them the second part of the study would take place in approximately three weeks.

Time 2 took place three weeks after Time 1. During time 2 of testing, all the participants received the same instructions and math test. The instructions for Time 2 of testing were the same as in the control condition during Time 1. The math test took during Time 2 of testing consisted of the same problems as the first math test but, in different order (See Appendix B). The answer sheet during Time 2 of testing was the same as the answer sheet in time 1 of testing (See Appendix C) except participants only needed to record their birthday; no demographic information was necessary. Participants were given 15 minutes to complete the math test and once time was up, the tests were collected, and participants were thanked and debriefed about the research project.

The scores of the math test were analyzed by computing the performance accuracy of the participants at Time 1. Performance accuracy was computed by comparing the number correct over the number attempted. Performance accuracy was then compared across all three experimental conditions.

## **Results**

The descriptive statistics (mean and standard deviation) for the three conditions at Time 1 can be found on Table 1.

A one-way ANOVA was run in order to determine if there was a difference across the three conditions (control, self-affirmation, and teaching) regarding performance accuracy at Time 1. The ANOVA was insignificant,  $F(2,47) = 1.309, p > .05$ . Follow up tests using Tukey's HSD showed that there were no significant differences between any of the three groups.

A two way ANOVA was run to determine if there are significant differences in performance accuracy at Time 1 regarding gender and the control, self-affirmation, and teaching conditions. The results showed that the main effect of gender was not significant  $F(1,44) = 1.597, p > .05$ . The main effect of condition was not significant  $F(1,44) = 1.597, p > .05$ . Also, the interaction of gender and condition was not significant with  $F(2,44) = 2.241, p > .05$ . Thus hypothesis 1 was not supported.

Since the hypothesis for the current study was not supported, it was not necessary to test the research question.

### **Discussion**

The intent of this study was to replicate research done on reducing stereotype threat for women in the context of mathematics. Surprisingly, the hypothesis of the study was not supported by the data as participants in the teaching and self-affirmation conditions did not perform better on a math test than the participants in the general stereotype threat condition at Time 1. This runs contrary to the research by Johns, Schmader, and Martens (2005) and Martens, Michael, Greenberg, and Schimel (2006). Johns, Schmader, and Martens (2005) found that teaching about stereotype threat is enough to decrease the negative effects on math performance for women. In research by Martens, Michael, Greenberg, and Schimel (2006), it was also found

that if participants self-affirm before they take a math test, the effects of stereotype threat are reduced. However, in the current research, there was not a significant difference between the control, teaching, and self-affirmation conditions. Thus, it was not necessary to analyze the data from Time 2 regarding whether this effect would last over time as the data from Time 1 was not significant.

### **Limitations**

Although the results of the current study were not significant, there was a lot that was learned from this study. First, it showed that the design of the study, sample size and confounding variables may have influenced my ability to find statistically significant results. First, this study utilized a quasi-experimental design; random selection was not used and therefore a cause and effect conclusion cannot be made. Sample size was a major problem in this study; the only participants in the study were from Dr. McNall's classes so the maximum sample was 141. However, due to the fact that some students were only present on one day of testing and data that was not able to be matched, the final sample size was only 50 participants with each of the three conditions only having between 12 and 21 participants. Having such a small sample of data is a problem because of the law of large numbers. The law of large numbers states that, the larger the size of the sample, the more likely it is that the sample is representative of the population (Gravetter, & Forzano, 2012). Therefore, the values obtained from the sample in this study may not truly be representative of the greater population.

Another problem relating to the small sample size was the unequal numbers of male and female students. Psychology is a predominantly female major, so many of the classes have many more females than males and this was the case in Dr. McNall's classes. I only had a total of 11 male and 39 female participants in the final sample. In fact, in the teaching condition, only the

data from one male participant was included in the analysis. The unequal numbers of male and female students made it difficult to see if the test really induced stereotype threat on the female students because it did not allow me to see if there was a real difference between the scores of the males and the females. For example, in the teaching condition, comparing the scores of 20 female participants to one male participant does not really give an accurate representation of how the females performed compared to the males.

Due to the nature of my study, participants were tested at different times of the day. Participants had to be tested during their class time due to the design of the study (i.e., 8 AM, 9:30 AM, 2 PM). It is not ideal having the participants take the test at different times because the participants could act differently depending on the time of day. For example, participants who took the test in the morning could have been tired and not really paying attention to the instructions, or they could have been very engaged because it is their first class of the day. Likewise, the participants who took the class in the afternoon (teaching condition) could have been tired because it was their last class or they just could have rushed through the exam so they could go home. Based on anecdotal evidence, most of the participants in the teaching condition were done several minutes before participants in the other conditions. Either way, participants could have been affected differently by the time of day and their scores may reflect those differences.

Another major problem in this study was being able to match up the data from Time 1 and Time 2. The instructions for the study stated that participants should put their birthdate (month and date only) on the answer sheet so the data could be matched. However, when trying to match up the data many participants put the month and year instead of month and day which made matching impossible. Also, some participants forgot to put their birthdate on the answer

sheet completely. Lastly, attrition is an issue as the study was over two class periods three weeks apart. Some participants participated in Time 1, but they were not present for Time 2 which again decreased the amount of data that was able to be matched.

### **Future Research**

Finding ways to reduce the effects of stereotype remains an important goal for researchers. Although at least two methods have shown some success in the short term (teaching and self-affirmation), more research is needed to find a method that works to reduce stereotype threat over time. A method is needed where an individual could learn the technique once and then apply it to new situations they encounter down the road. Teaching about stereotype threat offers one possible intervention but it is still unclear whether it works long term. Therefore, both old and new methods, such as boosting self-esteem or reducing anxiety, need to be tested to determine if they reduce stereotype threat in the long term so women can stop being negatively affected by stereotype threat and perform to their potential.

Although there were many confounding variables in this study, there are things that can be easily changed to assess whether teaching about stereotype threat has staying power. The procedure of the study is fairly sound and with a few minor changes the confounding variables could be reduced. First, the study would need to be run with more participants and an equal amount of male and female participants so the difference in performance between males and females can be determined with more certainty. Also, the study should no longer be a quasi-experimental design; the study should not be completed during a class and instead random assignment should be used to place participants in conditions to reduce confounding variables. The research should be conducted during a separate time that participants from any psychology class could come to for extra credit; this way all participants would be tested at the same time

reducing any effects from time differences. Lastly, a different coding method should be implemented in order to match up the tests from Time 1 and Time 2 of testing or the same method could be used, but the instruction will have to be clearer and the researchers will have to carefully check each participant's data before leaving the experiment to ensure accuracy.

Attrition is always going to be a problem in research, but having more participants to begin with and being able to match up the data from all the participants will make attrition less of a problem in future studies. In future studies, researchers may also want to include a manipulation check that would be given to see if the participants really felt stereotype threat. This could be done by having the participants complete an anxiety measure after the researcher induces stereotype threat but before any intervention is implemented. This would allow the researcher to determine if the stereotype threat reduction strategies actually reduced feelings of stereotype threat within the participants.

### **Practical Implications**

Even though this research study had many limitations, previous studies on stereotype threat and its effects on women's math performance point to some specific recommendations. In an educational setting, teachers should work to implement the research on reducing stereotype threat into their classrooms. However, for this to happen, the teachers themselves are the ones that need to be educated about the effects of stereotype threat and the methods that work to reduce it. Once the teachers are aware of the problem, they can work to educate their female students so they are no longer negatively affected by stereotype threat. A training program for teachers should be created by the researchers in stereotype threat reduction so they can learn about stereotype threat. This program can be sent to school districts around the nation and teachers could have an in-service day where they can learn about stereotype threat; teachers can

also learn about effective ways to reduce stereotype threat among their students. Also, future educators could be taught about stereotype threat and the methods to reduce it in their education classes at college so once they get a job they can implement these methods in their classroom.

At this time, one of the most promising methods to reduce stereotype threat in an educational setting appears to be the teaching method as described by Johns, Schmader, and Martens (2005). Education about stereotype threat in schools should begin early; from the start of formal schooling, teachers should emphasize that there is no difference in math ability between girls and boys. In order to reduce stereotype threat in high school and college, math teachers should address the stereotype and teach a lesson to their students about how the stereotype is not true and what one may feel when they are experiencing stereotype threat just like what was done in the research by Johns, Schmader, and Martens (2005). Since research has not yet shown that teaching about stereotype threat reduces it in the long term, it is important that teachers address stereotype threat more than once. Therefore, teachers should remind their students about stereotype threat at times when they have a chance of falling victim to it, like before a math test or before they take a standardized test.

Although there is still a lot more research that needs to be done on reducing stereotype threat, the research that has already been done by Johns, Schmader, and Martens (2005) can be used effectively in an educational setting to help girls struggling at math. It has also been shown in a study by Aronson and William (2004) that teaching about stereotype threat has helped African Americans improve their performance when under stereotype threat conditions. Therefore, the teaching method to reduce stereotype should continue to be tested on other groups of people who are affected by stereotype threat in order to help researchers determine if it is a universal way to reduce its effects. Although the current study was unable to replicate and

extend on findings from previous research, it provided a framework that can be used by future researchers to study methods for reducing stereotype threat both in the short and long term.

## References

- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer Jr.(Ed.), *Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.
- 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.
- Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*, 608–630.
- Gravetter, F. J., & Forzano, L. (2012). *Research methods for the behavioral sciences* (4th ed.). Belmont, CA: Wadsworth Cengage Learning.
- 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.
- Kassin, S., Fein, S., & Markus, H.R. (2011). Stereotypes, Prejudice, and Discrimination (8th ed.), *Social psychology* (pp. 145-200). Belmont, CA: Wadsworth Pub Co.
- 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.
- Kray, L. J., Thompson, L., & Galinsky, A. D. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology, 80*, 942–958.

- Martens, A., Michael, J., Greenberg, J., & Schimel J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*, 236–243.
- Nguyen, 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.
- 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.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629.
- Steele, C. M., & Aronson, J. (1998). Stereotype threat and the test performance of academically successful African Americans. In C. Jencks & M. Phillips (Eds.), *The Black–White test score gap* (pp. 401–427). Washington, DC: Brookings.
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 379–440). San Diego, CA: Academic Press.
- Total Group Profile Report. (2012). *College Board*. Retrieved January 17, 2013, from [media.collegeboard.com/digitalServices/pdf/research/TotalGroup-2012.pdf](http://media.collegeboard.com/digitalServices/pdf/research/TotalGroup-2012.pdf)

Table 1*Performance Accuracy for Time 1*

	Mean	SD
Control	.4725	.178
Self-Affirmation Condition	.4042	.242
Teaching Condition	.3673	.192

Appendix ATest #1

You will have 10 minutes to complete the following 10 math problems. Your performance on the math test will be assessed by how many you get correct so you should attempt to answer as many questions as possible. Please write your answers on the attached answer sheet and write your birthday (month and day only) in the spot at the top of your answer sheet. Your birthday will be used to compare performance between the first test and the second test three weeks later.

**Please do not write on the test**, scrap paper will be provided for you to do your work on. After you complete the test, please fill out the demographic questionnaire on the bottom of your answer sheet and wait until the researcher announces time is up.

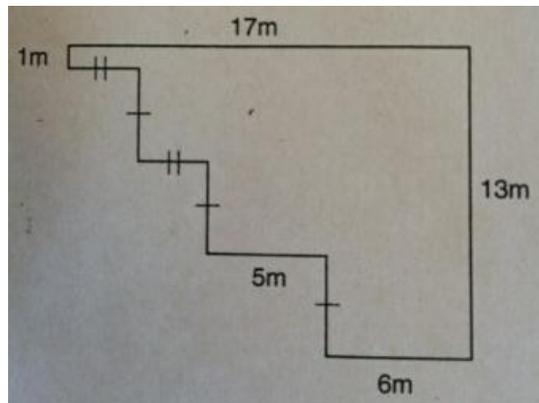
1. The advertised rate for roaming charges is 0.002 cents per second. What is that in dollars per hour?
  - a.) 7.2
  - b.) 2
  - c.) 0.72
  - d.) 0.2
  - e.) 0.072
2. Which of the following can be expressed as the sum of three consecutive integers?
  - a.) 278
  - b.) 49
  - c.) 11
  - d.) -1
  - e.) -6
3. A right triangle has sides 12, 15, and  $x$ . How many possible values are there for  $x$ ?
  - a.) 0
  - b.) 1
  - c.) 2
  - d.) 3
  - e.) Infinite
4. If  $abc$  is a three digit number define  $f(abc) = a \times b - c$ . How many three digit numbers,  $abc$ , are there so that  $f(abc) = 1$ ?
  - a.) 0
  - b.) 5
  - c.) 10
  - d.) 15
  - e.) 25

5. In the follow question, two quantities appear, one in Column A and one in Column B. You must compare them. The correct answer to the question is:

A if the quantity in Column A is greater  
 B if the quantity in Column B is greater  
 C if the two quantities are equal  
 D if it is impossible to determine which quantity is greater

Column A	$u \heartsuit v = (v - u)^u (u - v)^v$	Column B
9 ♥ 11		11 ♥ 9

6. Which of the following values is closest to  $\sqrt{201}$ ?
- 13.3
  - 13.9
  - 14.2
  - 14.8
  - 15.3
7. In 2000 Paul was twice as old as his brother Biko. In 2008 Paul was only 4 years older than his brother. In what year was Biko born?
- 1990
  - 1992
  - 1996
  - 1998
  - 2000
8. Find the perimeter.



- 42 m
- 42 m<sup>2</sup>
- 51 m
- 60 m
- 63 m

9. What is the average of four consecutive odd numbers starting with  $2n + 1$ ?
- a.) 4
  - b.)  $2n + 3$
  - c.)  $2n + 4$
  - d.)  $2n + 5$
  - e.)  $n + 4$
10. Not counting zero, what is the forty-ninth digit in the decimal expansion of  $1/7$ ? Recall  $1/7 = 0.142\dots$
- a.) 1
  - b.) 4
  - c.) 2
  - d.) 8
  - e.) 5

Appendix BTest #2

You will have 10 minutes to complete the following 10 math problems. Your performance on the math test will be assessed by how many you get correct so you should attempt to answer as many questions as possible. Please write your answers on the attached answer sheet and write your birthday (month and day only) in the spot at the top of your answer sheet. Your birthday will be used to compare performance between the first test you took three weeks ago with the test you are taking right now. **Please do not write on the test**, scrap paper will be provided for you to do your work on. After you complete the test, please wait until the researcher announces time is up.

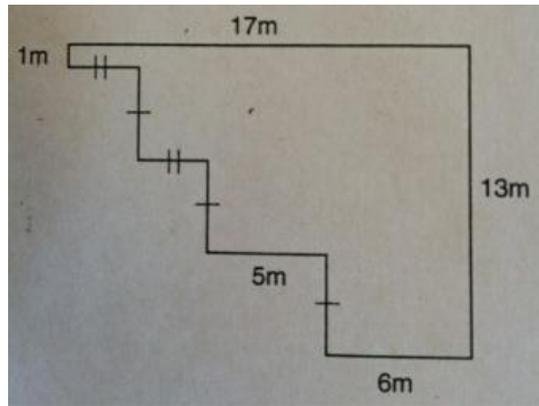
1. In the follow question, two quantities appear, one in Column A and one in Column B. You must compare them. The correct answer to the question is:

- A if the quantity in Column A is greater  
 B if the quantity in Column B is greater  
 C if the two quantities are equal  
 D if it is impossible to determine which quantity is greater

Column A	Column B
$u \heartsuit v = (v - u)^u (u - v)^v$	
9 ♥ 11	11 ♥ 9

2. A right triangle has sides 12, 15, and  $x$ . How many possible values are there for  $x$ ?
- 0
  - 1
  - 2
  - 3
  - Infinite

3. Find the perimeter.



- a.) 42 m  
 b.)  $42 \text{ m}^2$   
 c.) 51 m  
 d.) 60 m  
 e.) 63 m
4. Not counting zero, what is the forty-ninth digit in the decimal expansion of  $1/7$ ? Recall  $1/7 = 0.142857142857\dots$
- a.) 1  
 b.) 4  
 c.) 2  
 d.) 8  
 e.) 5
5. If  $abc$  is a three digit number define  $f(abc) = a \times b - c$ . How many three digit numbers,  $abc$ , are there so that  $f(abc) = 1$ ?
- a.) 0  
 b.) 5  
 c.) 10  
 d.) 15  
 e.) 25
6. The advertised rate for roaming charges is 0.002 cents per second. What is that in dollars per hour?
- a.) 7.2  
 b.) 2  
 c.) 0.72  
 d.) 0.2  
 e.) 0.072

7. In 2000 Paul was twice as old as his brother Biko. In 2008 Paul was only 4 years older than his brother. In what year was Biko born?
- a.) 1990
  - b.) 1992
  - c.) 1996
  - d.) 1998
  - e.) 2000
8. Which of the following values is closest to  $\sqrt{201}$ ?
- a.) 13.3
  - b.) 13.9
  - c.) 14.2
  - d.) 14.8
  - e.) 15.3
9. What is the average of four consecutive odd numbers starting with  $2n + 1$ ?
- a.) 4
  - b.)  $2n + 3$
  - c.)  $2n + 4$
  - d.)  $2n + 5$
  - e.)  $n + 4$
10. Which of the following can be expressed as the sum of three consecutive integers?
- a.) 278
  - b.) 49
  - c.) 11
  - d.) -1
  - e.) -6

Appendix C

Birthday: \_\_\_\_/\_\_\_\_

Answer Sheet: Test #1

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.

Demographic Information:

What is your sex? \_\_\_\_\_

What is your race/ethnicity? \_\_\_\_\_

**Appendix D**

Please rank the following six characteristics and values in order of personal importance:

- \_\_\_\_\_ Sense of Humor
- \_\_\_\_\_ Creativity
- \_\_\_\_\_ Physical attractiveness
- \_\_\_\_\_ Relations with family/friends
- \_\_\_\_\_ Intelligence
- \_\_\_\_\_ Social Skills

Now, please take a couple of minutes and write about why your most valued characteristic is personally important to you and describe a time when it had been particularly important to you.