

Effects of stereotype threat on undergraduate women's math performance: Participant pool vs. classroom situations

LINETTE M. McJUNKIN

Psychology and Research in Education, School of Education, University of Kansas <lmcjunker@ku.edu>

This experiment investigated the effect stereotype threat has on women's math performance, utilizing participant pool and classroom testing. The participant sample utilized college students enrolled in undergraduate psychology classes, and the classroom sample utilized students enrolled in an upper-level psychology course. I introduced stereotype threat, alleviated stereotype threat, or did not present stereotype threat, and assessed math scores accordingly. These findings support current research regarding the detrimental effects stereotype threat has on women's math performance. Because there was no difference noted in testing situations, this research suggests that stereotype threat is present in mock testing situations that carry no motivational factors as well as in real-life testing situations.

Keywords: stereotype threat, math performance, gender differences.

The statement "girls just aren't good at math" has been used to explain the gender gap in mathematics for decades. In elementary and middle schools, girls' and boys' scores are relatively equal on standardized tests. However, beginning in junior high school and continuing through adulthood, the gender gap widens between boys' and girls' scores. The National Center for Educational Statistics (NCES) reports that from 1990 to 2003, both girls and boys in the 4th and 8th grades had increases in mean math scores, with boys outperforming girls in 4th grade by one to three points, and 8th grade boys scoring one to two points above their female classmates on the mathematics assessments (NCES, 2003). A study in 2000 found that 15-year-old boys in the United States scored seven points higher than 15-year-old girls in the United States in mathematical literacy (NCES, 2000). Looking at college-bound seniors, the evidence is remarkably similar to the previous age groups. From 1991 to 2001, boys' scores increased from 504 to 507 on the math section of the SAT-I (NCES, 2002) and from 20.6 to 21.4 on the ACT (NCES, 2001). Meanwhile, from 1991 to 2001, girls who took the SAT-I had score increases from 496 to 502 (NCES, 2002), and on the ACT girls' scores went from 19.4 to 20.2 (NCES, 2001). The question remains apparent, if both genders have shown parallel increases in math scores, why do boys still outperform girls on mathematical tests? Research on stereotype threat has offered insight into what causes this constant difference in scores.

Researchers have presented environmental, societal, and educational factors as contributors to the difference between men and women in math scores. Researchers have examined many influences involved in performance gaps including biological models, brain differences, genetic factors, evolutionary processes, and hormonal influences (Keller, 2002). Keller also cited learned helplessness, anxiety, expectations, and values as psychosocial factors as elements,

which may be affecting performance. Research has also extended to examining the role socioeconomic components play in academic performance. Croizet and Claire (1998) examined the repeated demonstration of the relationship between socioeconomic status and intellectual ability, including the dramatic differences in scholastic achievement between low and high socioeconomic individuals. However, recent psychological research on gender differences in mathematical performance has seen an important development in stereotype threat theory.

Stereotype threat occurs when a negative stereotype about one's group becomes self-relevant, typically as an interpretation for something one is doing, an experience one is having, or for the situation one is in, which has relevance to one's self definition (Steele, 1997). Stereotype threat sets up a mutually reinforcing system, the fear of confirming the stereotype leads to behavior that confirms it; moreover, individuals do not have to believe the stereotype to be true for it to influence their behavior (Hyde & Kling, 2001). Any failure on the individual's part could support the stereotype. Members of any stereotyped group are susceptible to anxiety about being stereotyped, which in turn triggers an internal inferiority doubt, causing a decrease in performance, which is at least in part, if not fully, influenced by stereotype threat (Steele). Stereotype threat has been used as an explanation of poorer performance in research involving women's math scores, for the under performance of students from low socioeconomic backgrounds (Croizet & Claire, 1998), and for the continually low standardized test scores of African American youth (Steele). Stereotype threat has also emerged as a possible cause of the inequalities women face upon entering majors and careers dominated by men, such as science, math and engineering (Steele, James, & Barnett, 2002).

Stereotype threat research focusing on gender differences in mathematics attempts to understand the math performance of women, not in terms of abilities, but in terms of the interaction between the individual and a threatening situation posed by societal stereotypes (Spencer, Steele, & Quinn, 1999). Spencer et al. (1999) hypothesized that the apprehension caused by the risk of being judged by a negative stereotype is what disrupts women's math performance and causes the decrease in scores. Spencer and colleagues examined the effect of stereotype threat on women by presenting a difficult math test in three studies. Using a highly selected sample of undergraduate students with a strong background in mathematics in the first of three studies, Spencer et al. attempted to replicate the pattern of gender differences in scores found in the literature. In their second study, they attempted to manipulate stereotype threat by describing their test as either one that has produced gender differences or as one that has not produced gender differences. In their third study, they attempted to replicate their second study using a less selective sample of undergraduate students removing the requirement of a strong background in mathematics. In all three studies, women in the sample group who took the test described as not having gender differences, therefore alleviating stereotype threat, performed better than the men.

Extending stereotype threat research to include high school students, the presentation of a blatant stereotype threat negatively affected girl's performance on difficult math tests in Keller's (2002) study. Keller utilized first year high school students to examine the effects of inducing stereotype threat by describing the test in the experimental group as one that had produced gender differences, claiming boys had done better than girls on the test in the past. Keller found that this blatant statement of performance expectation was enough to cause significantly lower scores from the girls within that experimental group. Simply making the participant's gender salient has had positive effects on women's performance abilities. Cadinu, Maass, Frigerio, Impagliazzo and Latinotti's (2003) study showed women's scores changed accordingly when informed that women perform better, worse, or equal to men on logical-mathematical tests. McIntyre, Paulson, and Lord (2003) manipulated stereotype threat by first telling their participants that women make better participants than men in psychological research and then having them complete a difficult math test. These authors demonstrated the effects of alleviating stereotype threat in their second study sample by having participants read about the success of individual women in male dominated fields such as science, mathematics and engineering (Cadinu et al., 2003). In both studies, McIntyre et al. (2003) found that making gender salient had positive effects on the female participants' scores.

Because of the importance placed on classroom performance, I was interested in extending research to examine how the

presentation and removal of stereotype threat would affect the math scores of women from a volunteer participant pool and a separate classroom environment. In the classroom portion of the experiment, I attempted to mirror actual classroom testing environments by examining the effect of stereotype threat in an upper level psychology course. The classroom participants took the math test under the assumption it was for a grade, and their performance would be affecting their course grade, as a result allowing me to create an environment mirroring actual classroom testing. In line with current research, I expected that the introduction of an obvious stereotype threat would decrease women's performance, and removal of the threat would alleviate any influence on behavior, resulting in significantly higher test scores. Specifically, women within the stereotype threat alleviated conditions would perform better on a math test than women in the stereotype threat induced condition.

METHOD

Participants

Twenty-one women and five men volunteered from the undergraduate psychology participant pool, with a mean age of 20.10 years ($SD = 5.06$) for the women, and 20.8 years ($SD = 4.6$) for the men. Thirty women and 10 men volunteered from an upper level psychology course, with a mean age of 21.43 years ($SD = 2.90$) for the women and 21.7 years ($SD = 1.16$) for the men.

Materials

Test packets in this experiment were identical for the participant pool and the classroom sample. Packets included a personal information sheet to gather gender and age information, an instruction sheet outlining the time limit and directions for completing the test, and a page to introduce stereotype threat ("this test has displayed gender differences, exhibiting lower scores from women than men") or alleviate stereotype threat ("this test has displayed no gender differences; scores between men and women have been relatively equal"), and a 25 item math test comprised of GRE test questions ranging from easy to difficult (Educational Testing Service, 2001). For example, "If $2x + 7 = 12$, then $4x - 7 = ?$ ", "What is the length of a rectangle that has width 10 and perimeter 60?"

Design and Procedure

This study was a $2 \times 2 \times 3$ factorial design, with gender of participant (men, women) serving as an independent variable, testing situation (participant pool, classroom) as an independent variable, stereotype threat information (induced, alleviated, not presented) serving as an independent variable, and the score resulting from the number of questions correctly answered divided by the number of questions possible on the

math test served as the dependent variable. The Institutional Review Board (IRB) approved all procedures. For the participant pool, I distributed and read a standard informed consent form to the participants, and obtained each participant's signature. I then randomly distributed test packets to the participants so that an equal number of men and women were in the induced, alleviated, and not presented (which served as the control group) conditions. Participants in the not presented group received the same information and instructions; however, they had no stereotype threat page included in their packet. Participants then completed the personal information sheet and waited for further instructions. I read instructions aloud; outlining the time limit and informing participants how to complete the test, and directed the participants to read items on all pages, ensuring the participants read the stereotype threat induction or alleviation page, which followed the instruction page. I allowed participants 15 minutes to complete the math test, at which time I collected the packets, then distributed and read aloud a debriefing statement.

In the classroom sample, participants were under the assumption that the test was for a grade in their class. The course professor randomly distributed test packets to the participants so that an equal number of men and women were in each condition. Participants in the not presented group received the same information and instructions, but had no stereotype threat page included in the packet they received. The professor asked participants to complete the personal information sheet and wait for further instructions. He read instructions aloud outlining the time limit and informing participants how to complete the test. In addition, the professor directed participants to read items on all pages, ensuring the participants read the stereotype threat induction or alleviation page, which followed the instruction page. The instructor allowed participants 15 minutes to complete the math test, at which time he collected the test packets, and informed the participants about the study and asked them to volunteer their packets. Upon agreement, the professor distributed and read aloud the informed consent form, asking the participants who wanted to volunteer to sign and return the forms to him. The professor distributed the debriefing statement, read it aloud, and answered questions from the participants.

RESULTS

To examine the effects of stereotype threat within the two testing situations, volunteer participant pool and classroom, I compared the performance data from the participants in the two samples. I expected to find stereotype threat to have a negative effect on the participants receiving the stereotype threat induced packets. In addition, because there was a notion

of consequence placed on the classroom participants' performance, I expected to find a difference in scores between the two testing situations. Before analyzing the data, I focused on the assumptions regarding this study. My exploration revealed I met all assumptions.

To test my hypotheses, I ran a 2 (gender) x 2 (testing situation) x 3 (stereotype threat information) factorial analysis of variance (ANOVA) on the total correct on the math test. I have presented the means and standard deviations of the math test scores in Figure 1: no stereotype threat information ($M = 7.25$, $SD = 2.02$, $n = 16$), stereotype threat induced group ($M = 5.94$, $SD = 3.10$, $n = 18$), and stereotype threat alleviated group ($M = 11.41$, $SD = 3.76$, $n = 17$). I evaluated statistical significance with an alpha level of 0.05 and presented results in an ANOVA table (see Table 1). Analysis revealed a significant interaction between gender and stereotype threat information, $F(2, 54) = 3.85$, $p = 0.03$, partial $\eta^2 = 0.13$. To determine where the differences were, I split the file on gender and ran a one-way ANOVA. Results indicated that only the women within the sample had significant differences in scores, therefore, I hand calculated a Tukey for those scores only. Though the focus of the study was on women and those groups were equal, I chose to use the estimated marginal means because of the unequal n s for gender. Keppel and Zedeck (1989) concluded that unweighted analysis is unaffected by the sample size, implying that the unweighted means reflect the potential outcome had the groups been equal initially. Therefore, by using the unweighted means, I am alleviating any potential influence the unequal means might have on the data. Post hoc analysis revealed the women within the stereotype threat alleviated group performed significantly higher than the women within the stereotype threat induced and control groups that received no stereotype threat information ($p < 0.05$). There was no difference noted between the control or induced groups.

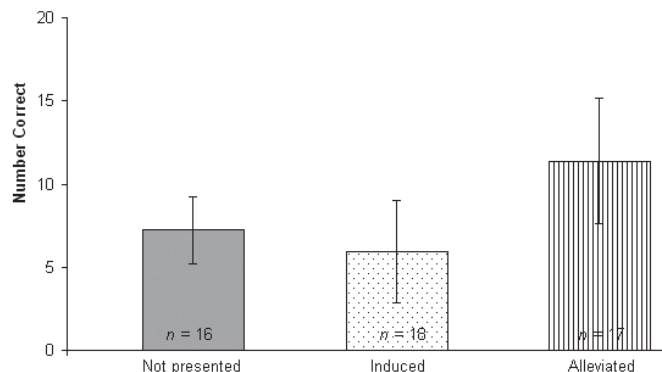


Figure 1. Math scores for women by experimental situation as a function of stereotype threat between subjects. Different superscript letters on column labels indicate significant difference ($p < 0.05$). Error bars depict standard deviations.

Table 1. Analysis of variance for gender, testing situation, and stereotype threat information. Note: Values enclosed in parentheses represent mean square error. S = participants, * $p < 0.05$

Source	df	F	partial n^2	p
Gender (A)	1	0.23	0.004	0.63
Situation (B)	1	0.001	0.00	0.98
Information (C)	2	3.48*	0.11	0.04
A x B	1	1.89	0.03	0.18
A x C	2	3.85*	0.13	0.03
B x C	2	0.32	0.01	0.72
A x B x C	2	3.06	0.10	0.06
S within-group error	54	(8.50)		

DISCUSSION

Stereotype threat is present when a woman takes a math test even if she is not reminded verbally or visually of the stereotype (Oswald & Harvey, 2000-2001), and this threat substantially interferes with a woman's performance, especially on a difficult math test (Spencer et al., 1999). The goal for the present study was to analyze the effect removal of stereotype threat has on women's math test performance in different situations. Similar to the previous research, the removal of stereotype threat had a positive effect on the scores of women from the participant pool and the scores from the classroom sample. These results support the belief that because of the popular social stereotype regarding women's math abilities, stereotype threat is present in diverse situations when a woman is taking a math test. Mere recognition that a negative stereotype applies to oneself cues internalized anxiety (Steele, 1997); as a result, the removal of stereotype threat leads to an increase in performance (Oswald & Harvey).

To extend previous research, I used a classroom sample and was able to analyze the role stereotype threat plays and how performance motivation affects stereotype threat. Performance expectations play a prominent role in research on the cause of stereotype threat (McIntyre et al., 2003). By using classroom presentation, I was able to look at stereotype threat and its interaction within a setting where performance motivation, or motivation to get a good grade on the test, was high. Consistent with Steele (1997), decreasing the interpretive relevance of a stereotype within the performance setting reduces the threat and its detrimental effects. Therefore, though there was not a

notable difference between testing situations, there was a positive effect on the scores by alleviating stereotype threat within the performance domain. This finding is important with regard to knowing what affects the scores of stereotyped groups and how to eliminate their underperformance.

There are clear limitations within this study that I need to address. Lack of sample size and the convenient nature of the sample are two variables that limit the generalization of these results. Because this type of research examines a phenomenon occurring in elementary and secondary education environments, I would have liked to have had a much larger sample and have used participants aside from college students, as they are not representative of the national student population. Another limitation is the status and gender of the experimenter. In the classroom sample, the experimenter was the class professor; however, in the laboratory setting, the experimenter was a student of equal status to the participants. The professor was also a male, whereas the student was female. These are possible confounding variables, and in the future, I would address this concern by ensuring the same experimenter tested all groups, eliminating both issues.

Future research might benefit from exploring the effects of stereotype threat in a mathematical environment, such as using a college math class. Another possible extension could include the effects of alleviation. For example, is alleviation effective with just the person who initiated it, only in an academic environment, and is repeated alleviation needed? In addition, researching the effects of stereotype threat with regard to children and adolescents, and the age when stereotype threat initially affects performance would be important steps in understanding the psychological experiences that cause gender differences in mathematics as well as the underperformance of stereotyped groups within any academic domain.

REFERENCES

- Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology*, *33*, 267-285.
- Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, *24*, 588-598.
- Educational Testing Service. (2001). *GRE: Practicing to take the general test*. (9th ed.)
- Hyde, J. S., & Kling, K. C. (2001). Women, motivation, and achievement. *Psychology of Women Quarterly*, *25*, 364-378.

- Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 47*, 193-198.
- Keppel, G., & Zedeck, S. (1989). *Data analysis for research designs*. New York: W. H. Freeman.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology, 39*, 83-90.
- National Center for Educational Statistics. (2000). *Average reading, mathematics, and science literacy scores \I\ of 15 year olds, by sex: Selected countries, 2000*. Retrieved August 23, 2004, from <http://www.nces.ed.gov/programs/digest/02/tables/dt408.asp>
- National Center for Educational Statistics. (2001). *American College Testing (ACT) score averages, by sex: 1970 to 2001*. Retrieved August 23, 2004, from <http://www.nces.ed.gov/programs/digest/d02/tables/dt137.asp>
- National Center for Educational Statistics. (2002). *Scholastic Assessment Test score average for college-bound high school seniors, by sex: 1966-67 to 2001-02*. Retrieved August 23, 2004, from <http://www.nces.ed.gov/programs/digest/d02/tables/dt134.asp>
- National Center for Educational Statistics. (2003). *Average mathematics scale scores, by gender, grades 4 and 8: 1990-2003*. Retrieved August 23, 2004, from <http://nces.ed.gov/nationsreportcard/mathematics/results2003/scalegender-all.asp>
- Oswald, D. L., & Harvey, R. D. (2000-2001). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology: Developmental, Learning, Personality, Social, 19*, 338-356.
- 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 ability and performance. *American Psychologist, 52*, 613-629.
- Steele, J., James, J. B., & Barnett, R. C. (2002). Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. *Psychology of Women Quarterly, 26*, 46-50.