The Journal of Experimental Education

ISSN: 0022-0973 (Print) 1940-0683 (Online) Journal homepage: www.tandfonline.com/journals/vjxe20

An Experimental Study of the Effects of Stereotype Threat and Stereotype Lift on Men and Women's Performance in Mathematics

Heather J. Johnson, Lucy Barnard-Brak, Terrill F. Saxon & Megan K. Johnson

To cite this article: Heather J. Johnson, Lucy Barnard-Brak, Terrill F. Saxon & Megan K. Johnson (2012) An Experimental Study of the Effects of Stereotype Threat and Stereotype Lift on Men and Women's Performance in Mathematics, The Journal of Experimental Education, 80:2, 137-149, DOI: 10.1080/00220973.2011.567312

To link to this article: https://doi.org/10.1080/00220973.2011.567312

Published online: 10 Feb 2012.

THE JOURNAL OF EXPERIMENTAL EDUCATION, *80*(2), 137–149, 2012

ISSN: 0022-0973 print/1940-0683 online
DOI: 10.1080/00220973.2011.567312

Routledge
Taylor & Francis Group

MOTIVATION AND SOCIAL PROCESSES

An Experimental Study of the Effects of Stereotype Threat and Stereotype Lift on Men and Women's Performance in Mathematics

Heather J. Johnson
Baylor University

Lucy Barnard-Brak
Texas Tech University

Terrill F. Saxon and Megan K. Johnson
Baylor University

In this study the authors examined the differential effects of stereotype threat and lift between genders on math test performance. They asked 3 questions: (a) What is the effect of gender on math test performance?, (b) What is the effect of stereotyping condition (threat, lift, or neither) on math test performance?, and (c) What is the effect of the interaction of gender and stereotyping condition on math test performance? Findings indicated that men performed better on math tests under conditions of stereotype threat than on stereotype lift; women performed better under stereotype lift than on stereotype threat. Practical applications are discussed regarding math test anxiety, social identities, and how teachers might address gender differences regarding stereotype threat and stereotype lift.

Keywords *gender, mathematics, stereotype lift, stereotype threat*

STEREOTYPE THREAT HAS BEEN DEFINED as the threat that others' judgments or stereotypes about a certain group's performance in a given domain (e.g., math) will cause an individual belonging to that group to perform in such a way that it confirms the negative stereotypes held about that group (Steele, 1997). For instance, in the domain of math, women have been stereotyped as performing more poorly than men (Spencer, Steele, & Quinn, 1999; Steele, 1997). Thus,

Address correspondence to Heather J. Johnson, Department of Educational Psychology, Baylor University, One Bear Place #97301, Waco, TX, USA. E-mail: Heather_Hays@Baylor.edu

stereotype threats against women often cause them to perform more poorly in math than men. Stereotype threat occurs when one's ingroup is negatively stereotyped. For instance, gender is a type of group in which women are ingroup members of the female gender and men are ingroup members of the male gender. Similarly, individuals view themselves as members of specific races. African Americans view themselves as ingroup members of the African American race.

Despite the large body of literature examining the effects of stereotype threat on women's math performance, little research has been done to examine the differential effects of stereotype lift among men and women in math performance. *Stereotype lift* is defined as the boosting of performance in a given domain that occurs when an outgroup is negatively stereotyped (Walton & Cohen, 2003). For instance, women view men as an outgroup and African Americans view Caucasians as an outgroup, and each may create negative stereotypes about their respective outgroups (men and Caucasians) to boost their own esteem and performance. Thus, stereotype lift is the opposite of stereotype threat. In the present study, we examined the effects of stereotype threat and stereotype lift on men and women's performance on a math test.

Stereotype Threat

Anyone who cares deeply about the domain tested can be a target of stereotype threat (Aronson et al., 1999), so stereotype threat does not only apply to persons belonging to ethnic minority groups and/or who are female. Although stereotype threat has also been indicated to greatly affect one's performance in several domains (Spencer et al., 1999; Steele & Aronson, 1995), the effect of stereotype threat appears to be most detrimental to an individual's performance when the test of performance is in a relevant domain of high value to the individual and targets the appropriate level of difficulty to assess performance properly (Aronson et al., 1999; Quinn & Spencer, 2001; Spencer et al., 1999; Steele, 1997; Thoman, White, Yamawaki, & Koishi, 2008). This is why women perform more poorly on math tests (relevant domain) than men when stereotype threat exists but only if the math test is the appropriate level of difficulty (Spencer et al., 1999). As such, the focus of this article will be on the stereotype threat against women's math performance.

In addition, when an individual has high expectations for personal performance, the individual also suffers from stereotype threat (Marx, Brown, & Steele, 1999; Steele, 1997; Williams, 2006). One does not need to believe that the stereotype characterizes oneself in order to experience stereotype threat (Steele, 1997). Rather, the threat only needs to be in a relevant domain (e.g., poorer performance on a math test among women and African Americans). Steele (1997) also contended that when an individual perceives a stereotype threat, he or she may seek to prove the stereotype wrong. In doing so, the individual may become overwhelmed and affected by the fear of proving the stereotype correct, thus inhibiting performance.

Stereotype Threat Among Women and Men and its Effects on Math Performance

According to Spencer et al. (1999), stereotype threat is associated with mathematics, specifically affecting men and women in different ways. Math is of some importance to both genders; our study seeks to identify how and in what ways its effects may be beneficial or detrimental to men and women's performances. Because of culturally rooted expectations for women's poor performance in math (Eccles, 1987; Eccles-Parsons et al., 1983), women have often been found to suffer under conditions that magnify their deficiencies, particularly in mathematics (Quinn & Spencer, 2001; Spencer et al., 1999; Thoman et al., 2008). In particular, women perform

worse in mathematics under stereotype threat conditions than in non–stereotype threat conditions (Quinn & Spencer, 2001; Spencer et al., 1999; Walsh, Hickey, & Duffy, 1999). These results suggest that women may be aware of the negative stereotypes about their gender's performance in mathematics. This awareness can lead to a stereotype threat that creates a self-fulfilling prophecy in which women, despite their efforts, succumb to the threat (McIntyre, Paulson, & Lord, 2003; Steele & Aronson, 1995). For example, when told that math exams frequently produce gender differences in performance results, women's performance suffered (Spencer et al., 1999; Walsh et al., 1999). Women appeared to infer that the gender differences in results coincided with the stereotype that women are not as proficient in math (compared with men) and thus applied this stereotype to themselves and acknowledged defeat before performing the task (Spencer et al., 1999; Steele & Aronson, 1995). Women's decreased performance under stereotype threat conditions may erroneously suggest the logical fallacy that women do underperform men; however, in non–stereotype conditions, women appear to have performed the same as men (Spencer et al., 1999).

In contrast with women, men rise to the challenge of stereotype threat by increasing their level of performance (Spencer et al., 1999; Williams, 2006). For instance, men outperformed other men who were not in stereotype threat conditions when told that gender differences were produced by a test (Williams, 2006). Because men expect to do well on math exams (Brown & Josephs, 1999), completing a math exam provides them the opportunity to preserve or increase their math skill status (Josephs, Newman, Brown, & Beer, 2003). Thus, men may not perceive the stereotype threat to be true or accurate. As a result, men may be motivated to prove the stereotype as being incorrect in order to maintain their perceived dominance in math (Wheeler & Petty, 2001; Williams, 2006).

There is a gap in the literature regarding the assessment of the level of stereotype threat felt by men. Although to our knowledge there is no research that has analyzed how men feel when threatened, according to Williams (2006), one can draw conclusions about this relation on the basis of the influence of stereotype threat on performance. Men rise to challenge the threat when told that their performance tends to be worse than women (Niederle and Vesterlund, 2007). Research indicates that, although men and women are overconfident about their relative performance, men tend to be more overconfident than women (Beyer, 1990; Beyer & Bowden, 1997). In addition, when given a choice to engage in a competitive task, men are more likely to participate. However, according to Lundeberg, Fox, and Puncochar (1994), gender differences in confidence to participate often only occur when tasks are in a masculine domain. Depending on the individual's perception of the task, overconfidence may or may not be a factor in performance.

In the present study, we examined the effects of stereotype threat among men and women on math performance in order to test whether our results were in line with the previous literature. Our main focus in the present study, however, was to examine how the effects of stereotype lift act differently than those of stereotype threat among men and women. In the subsequent section, we briefly discuss stereotype lift.

Stereotype Lift

As mentioned, *stereotype lift* causes an individual to perform better than those in the presence of a non–stereotype lift condition (Walton & Cohen, 2003). The presence of stereotype lift has been suggested to enhance behavior and increase the level of performance of the individual being presented with the lift. The negative labeling or stereotyping of the "outgroup" can indirectly

provide an individual with a perceived advantage (Walton & Cohen, 2003, p. 456). In addition, this negative stereotyping of an outgroup member can enhance self-esteem. For example, among individuals whose self-esteem had been threatened, the derogation of a target by negatively stereotyping him or her mediated an increase in self-esteem (Fein & Spencer, 1997). Thus, negatively stereotyping outgroup members leads to increases in performance and self-esteem (Fein & Spencer, 1997; Walton & Cohen, 2003). When cognizant of the weaknesses of outgroups, an individual may approach the task more highly endorsing the assumption that he or she can and will succeed (Chalabaev, Stone, Sarrazin, & Croizet, 2008).

Research Questions

The purpose of the present study was to examine the effects of stereotype threat and lift according to gender on math test performance. To achieve this purpose, we examined three research questions:

1. What is the effect of gender on math test performance controlling for past stereotype threat vulnerability?
2. What is the effect of stereotyping condition of threat, lift, or neither on math test performance controlling for past stereotype threat vulnerability?
3. What is the effect of the interaction of gender and stereotyping condition on math test performance also controlling for past stereotype threat vulnerability?

In performing our analyses, we controlled for stereotype threat vulnerability before being exposed to an experimental condition of threat, lift, or neither.

METHOD

Participants

A sample of 458 volunteered to complete the study by responding to a recruitment e-mail message sent out to a random sample of 3,000 e-mail addresses at a large, private university located in the southwestern United States. Approximately 39% ($n = 178$) of the sample were men and the remaining 61% ($n = 280$) were women. With regard to ethnicity, approximately 5.6% ($n = 26$) identified themselves as Native American, whereas 8.9% ($n = 41$) were Asian American, 9.8% ($n = 45$) were African American or Black, 10.7% ($n = 49$) were Hispanic, and 79.9% ($n = 366$) identified themselves as White. These frequencies and resulting percentages exceed the total sample size as participants were permitted to select all ethnic categories with which they identified. A total of 24 different college majors were represented in the sample. The mean self-reported grade point average was 3.02 ($SD = 0.82$). Age ranged from 18 to 55 years ($M = 22.2$ years, $SD = 6.35$ years).

Measures

In the present study, participants were randomly assigned to one of three stereotyping conditions: (a) no stereotype threat or lift (control condition); (b) stereotype lift, and (c) stereotype threat. After

being randomly assigned to one of these three conditions as represented by three different online surveys, all participants were directed to complete a series of demographic questions including but not limited to gender, ethnicity, major, and classification. On the basis of their responses to the question of gender, students were directed to complete a stereotype threat vulnerability scale specific to their gender known as the Stereotype Vulnerability Scale (SVS). The SVS was developed as part of a funded dissertation project (Spencer, 1993) and measures the degree to which an individual reports feeling threatened by a negative stereotype threat with regard to a specific domain of academic success. Data from the SVS were used to control for an individual's preexisting vulnerability to stereotype threat. The SVS comprises eight items and is rated on a 7-point, Likert-type scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). For our study, we contextualized the SVS for the domain of mathematics. The following are two examples of the same items for men and women, respectively, that were subsequently reverse-coded:

- Men rarely face unfair evaluations in math classes.
- Women rarely face unfair evaluations in math classes.

Previous studies have indicated that the SVS has adequate psychometric properties for the domain of mathematics. Steele, James, and Barnett (2002) reported that the SVS had achieved an acceptable level of internal consistency with $\alpha = .84$ for the scores obtained in the domain of mathematics. Barnard, Burley, Olivarez, and Crooks (2008) examined the psychometric properties of the SVS and revealed a similar level of internal consistency with $\alpha = .82$ for the scores obtained in the domain of mathematics. In the present study, scores obtained appeared to demonstrate a similar, acceptable level of internal consistency with $\alpha = .81$ for the domain of mathematics. With regard to construct validity, Barnard et al. (2008) performed confirmatory factor analyses on the SVS, which revealed acceptable psychometric properties as contextualized to the domain of mathematics. Using data from the present study, we performed an additional confirmatory factor analysis of the SVS. Maximum-likelihood confirmatory factor analysis results indicate evidence to support a one-factor structure with a comparative fit index value of .98, a Tucker-Lewis index value of .96, and a root mean square error of approximation value of .04, $CI_{90\%}$ [.02, .06].

After completing the SVS, participants proceeded to the experimental condition to which they were randomly assigned. Students who were randomly assigned to the no stereotype threat or lift condition received the following prompt:

> For the following math questions, neither women nor men are expected to do better or worse. Please answer the following questions to the best of your ability and note that your responses are anonymous and confidential.

The stereotype lift and stereotype threat conditions were contextualized according to participants' reported gender. Thus, students who self-identified as men and who were randomly assigned to the stereotype lift condition received the following prompt:

> For the following math questions, men are expected to do better than women. Please answer the following questions to the best of your ability and note that your responses are anonymous and confidential.

Students who self-identified as women and who were randomly assigned to the stereotype lift condition received the following prompt:

> For the following math questions, women are expected to do better than men. Please answer the following questions to the best of your ability and note that your responses are anonymous and confidential.

In addition, students who self-identified as men and who were randomly assigned to the stereotype threat condition received the following prompt:

> For the following math questions, men are expected to do worse than women. Please answer the following questions to the best of your ability and note that your responses are anonymous and confidential.

Students who self-identified as women and who were randomly assigned to the stereotype threat condition received the following prompt:

> For the following math questions, women are expected to do worse than men. Please answer the following questions to the best of your ability and note that your responses are anonymous and confidential.

Thus, all prompts were ability-based stereotypes rather than effort-based stereotypes (Thoman et al., 2008). After we provided their respective prompt, we assessed students' math test performance. To measure math test performance, we administered five multiple-choice math questions. We evaluated these math questions for their content validity as being aligned with questions that may be found in the Scholastic Aptitude Test (SAT; College Board, 2009). According to Liu, Schuppan, and Walker (2005), questions from the SAT are reliable assessments in measuring mathematics performance. Scores obtained for these math questions achieved an acceptable level of internal consistency with $\alpha = .90$. These scores were then summed to construct a math test performance composite. A maximum-likelihood CFA was performed to address the sufficiency of the five items to measure math performance given the sample size and the domain. This analysis revealed a one-factor structure with a comparative fit index and Tucker-Lewis index (i.e., the nonnormed fit index) value of 1.00 with a root mean square error of approximation value of 0.01, $CI_{90\%}$ [0.00, 0.03].

Procedure

The entire study was administered online. A recruitment e-mail message was sent to a random sample of 3,000 student e-mail addresses that was previously obtained. We oversampled because there was no tangible incentive for students to participate. From this random sample, approximately 276 of the e-mail addresses were undeliverable because the student's e-mail account was full or the student no longer attended the university. From the remaining 2,724 randomly selected student e-mail addresses, approximately 458 students responded to volunteer to complete the study online. On the basis of these numbers, an estimate of our response rate would be 16.81%. This estimate, however, may be considered conservative given that there are an unknown number of recruitment e-mail messages that were delivered to spam or junk e-mail folders. Of the 458 students who self-selected to participate in our study, 422 students completed the study in its entirety. Approximately 7.8% of the cases were missing completely or in part. Given the low amount of missing data, values for missing data were handled using a listwise method of deletion. All analyses were performed in SPSS (Version 17.0).

Analyses

To examine our research questions and achieve the purpose of the present study, we analyzed data in a two-way analysis of covariance (ANCOVA). To perform this ANCOVA, we evaluated the assumption of the homogeneity of slopes by examining the relation between the dependent variable (e.g., math test performance) and the covariate (e.g., stereotype threat vulnerability in math) across the different levels of the independent variables (e.g., gender and stereotyping condition). This interaction was statistically nonsignificant, $F(2, 421) = 2.79, p = .062$, indicating that the assumption of the homogeneity of slopes may be considered satisfied. The assumption of the homogeneity of variances was also evaluated and indicated that this assumption may be considered tenable also with a statistically nonsignificant test statistic, Levene's $F(5, 422) = .872$, $p = .50$. Values for Cohen's f and d were calculated as the measures of effect size as appropriate. Values of .20, .50, and .80 and larger of Cohen's d may be considered as small, medium, and large, respectively, whereas values of .10, .30, .50 and larger on Cohen's f may be considered as small, medium, and large, respectively, in social and behavioral science research such as the present study (Cohen, 1988). These effect size benchmarks set by Cohen (1988) are appropriate for interpreting our results in this study. It should be noted, however, that these effect sizes are context specific and should be interpreted in terms of the outcome variables examined. We suggest medium to large effect sizes in regard to academic performance, our outcome variable, should be taken into consideration by educational systems.

RESULTS

Research Questions

In answering our first research question as to whether there were differences in math test performance scores according to stereotyping condition, we performed a 2×3 ANCOVA. In controlling for stereotype threat vulnerability, there were no statistically significant differences in math test performance scores according to stereotyping condition, $F(2, 421) = .413$, $p = .662$. In removing the covariate of stereotype threat vulnerability, there still were no statistically significant differences. Figure 1 contains a summary diagram of these results.

To answer our second research question, we considered the main effect of gender as associated with differences in math test performance scores. Again, we controlled for stereotype threat vulnerability in these analyses. Our results indicate that there were statistically significant differences in math test performance scores according to gender, $F(1, 421) = 8.621$, $p = .004$, $d = .299$. This value of Cohen's d indicates a small to medium sized effect (Cohen, 1988). Men ($M = 1.046$, $SD = 0.749$) appeared to significantly outperform women ($M = 0.820$, $SD = 0.758$) on math test performance.

In examining our third and final research question, we examined whether there were statistically significant differences in math test performance scores according to a possible interaction of gender and stereotyping condition. Our results indicated a significant interaction between gender and stereotyping condition on math test performance scores, $F(2, 421) = 3.852$, $p = .022$, $f = .14$. This value of Cohen's f indicates a small effect (Cohen, 1988). Upon finding a significant interaction, we performed simple effects analyses according to the procedures outlined by Green and Salkind (2004).

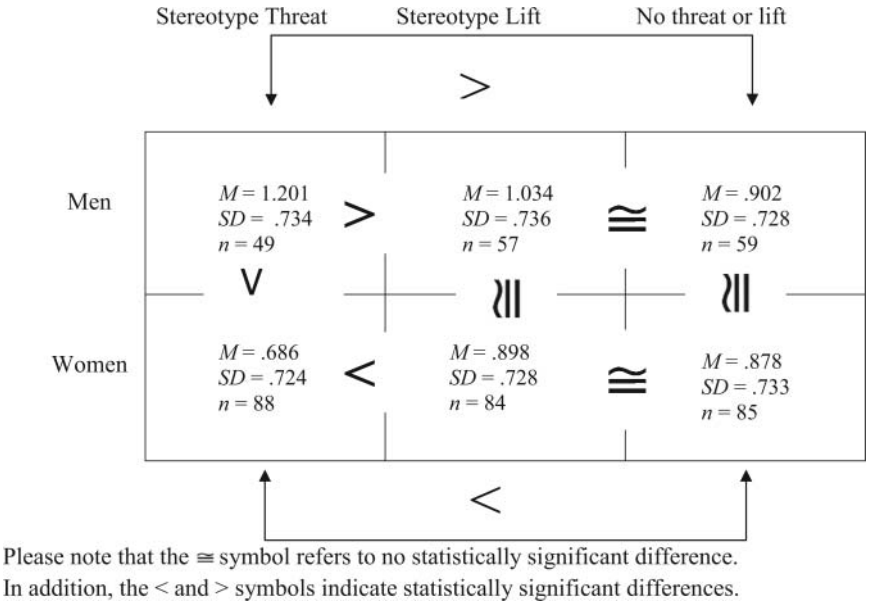

FIGURE 1 Diagram summarizing results, by condition.

Gender Differences

In the no stereotype threat or lift condition, the math test performance scores of men and women did not significantly differ, $F(1, 421) = .043$, $p = .835$. In addition, there were no statistically significant differences in math test performance scores according to gender in the stereotype lift condition, $F(1, 421) = 1.024$, $p = .312$. However, in the stereotype threat condition, there were statistically significant differences in math test performance according to gender, $F(1, 421) = 13.426$, $p < .001$, $d = .706$. This value of Cohen's d indicates a moderate to large effect size (Cohen, 1988). On the basis of this result, it appears that men ($M = 1.201$, $SD = 0.734$) achieved better math test performance than did women ($M = 0.686$, $SD = 0.724$) in the stereotype threat condition.

Among men, there were no statistically significant differences in math test performance according to the no threat or lift condition versus the stereotype lift condition, $F(1, 421) = .857$, $p = .355$ and the stereotype lift condition versus the stereotype threat condition, $F(1, 421) = 1.21$, $p = .272$. However, men achieved significantly better math test performance scores in the stereotype threat condition ($M = 1.201$, $SD = 0.734$) versus the no threat or lift condition ($M = -0.902$, $SD = 0.728$), $F(1, 421) = 4.63$, $p = .032$, $d = .40$. This value of Cohen's d indicates a medium effect size (Cohen, 1988).

Among women, there were no statistically significant differences in math test performance scores between the no threat or lift condition and the stereotype lift condition, $F(1, 421) = 0.04$, $p = .841$. Women achieved the best math test performance scores in the no threat or lift condition ($M = 0.878$, $SD = 0.733$) and in the stereotype lift condition ($M = 0.898$, $SD = 0.728$). Women performed the worst in terms of math test performance in the stereotype threat condition

($M = 0.686$, $SD = 0.724$) in comparison with the stereotype lift condition, $F(1, 421) = 4.52$, $p = .034$, $d = .29$; and the no threat or lift condition, $F(1, 421) = 5.45$, $p = .02$, $d = .26$.

DISCUSSION

This study is the first, to our knowledge, to examine simultaneously the effects of stereotype threat and lift among men and women on math performance. Although men outperform women on math performance overall, stereotype threat and lift have differential effects according to the gender of participants. In line with past research, we found that women performed more poorly on a math test when faced with stereotype threat (Quinn & Spencer, 2001; Spencer, 1993; Spencer et al., 1999; Walsh et al., 1999) whereas men performed better on a math test when faced with stereotype threat (Spencer et al., 1999; Williams, 2006). However, this study demonstrates that those patterns persist even when comparing stereotype threat conditions to stereotype lift conditions. In other words, men perform better on math tests when under stereotype threat than when under no threat or stereotype lift. Women, alternatively, perform better under no threat or stereotype lift than under stereotype threat. Whereas previous research has demonstrated that individuals should show a performance boost after exposure to an outgroup's inferiority (see Walton & Cohen, 2003), the present study demonstrates that men do not benefit from this stereotype lift.

Stereotype Threat, Lift, and the Potential Role of Arousal

Given these results, we can interpret that men rise to the challenge of stereotype threat by performing better, whereas women perform better under stereotype lift. This difference in performance under stereotype threat has been found to be related to the negative connotations that women associate with math (Wigfield & Meece, 1988); namely, women felt vulnerable to the negative stereotype of women performing more poorly in math than did men. In addition, women experience more negative affective reactions to math than men do, and this emotionality is strongly tied in with the affective components of test anxiety (Liebert & Morris, 1967). Stereotype threat among women has been linked with levels of arousal (Ben-Zeev, Fein, & Inzlicht, 2005), and women's decreased performances on math tests is attenuated by allowing them to misattribute their arousal. Future research should assess whether this test anxiety, in turn, may be the component of stereotype threat that is linked to higher levels of arousal in women during math tests. Because these higher levels of arousal lead to decreases in math test performance, one hypothesis is that the stereotype lift condition may have increased performance in women by inducing less arousal than the stereotype threat condition did. If supported, this hypothesis could demonstrate women could be more likely to perform better on a test in a gender-relevant domain (i.e., math) when under these low arousal conditions.

If women performed more poorly under higher levels of arousal (stereotype threat condition), why did men perform the best under high levels of arousal and worse under lower levels of arousal (stereotype lift)? Future research should examine whether or not men approach the task of completing a math exam with the intention of challenging the stereotype threat condition (Brown & Josephs, 1999). Instead of having to overcome a preexisting negative stereotype, men simply seek to uphold the stereotype regarding their gender's superiority in math ability

and performance. In contrast, when men are presented with stereotype lift conditions, they are not especially concerned with their performance because an "opportunity to enhance their status" does not exist (Josephs et al., 2003, p.162). Future research could examine whether the arousal associated with stereotype threat on math exams for men is associated with more of a challenge rather than anxiety (as with women) because a negative stereotype against men is not a commonly held stereotype. As such, men may not feel threatened but rather challenged in the face of stereotype threat. Future studies are needed, however, to determine whether challenging men on a math task produces the same results as presenting them with stereotype threat.

Stereotype Threat, Lift, and Social Identity

Another reason that stereotype lift may have differential effects on men and women is because of the role that social identity might play. The social self has been shown to play a crucial role in stereotype threat (Steele, Spencer, & Aronson, 2002) and in stereotype lift (Marx & Stapel, 2006). In stereotype threat, as discussed, the more strongly that one identifies with the group about which a negative stereotype exists, the more that he or she will be negatively affected by stereotype threat (J. Steele, et al., 2002). However, stereotype lift has been shown to only be affective in nontargets of stereotypes (i.e., men) if their social identities are made salient (Marx & Stapel, 2006). The more direct this social manipulation becomes, the stronger the effects are for stereotype lift. Women, however, require no activation of their social identity because the task is already tied to their social identity strongly enough. Hence, in future experiments, researchers should more strongly activate men's social identity to determine whether stereotype lift will have a stronger effect.

Stereotype Privilege

Our findings directed us to discover a concept we would like to label *stereotype privilege*, defined as the ability to counteract stereotype threat in a certain domain by those individuals who have been consistently perceived having an advantage in this certain domain, such as men in math. On the basis of this definition, our study activated stereotype privilege in men under the stereotype threat condition. Stereotype privilege, in this sense, is not a coping mechanism, but rather an offensive or counteracting measure.

Practical Applications and Suggestions for Future Research

The present study provides a multitude of practical applications to the classroom for ways to increase math test performance among men and women. In line with past research, women in this study performed the worst under stereotype threat. Thus, efforts should be made to reduce the effects of stereotype threat. Because arousal leads to decreases in performance among women when under stereotype threat (Ben-Zeev et al., 2005), teachers should look for ways to either reduce this level of arousal or allow women to attribute it to another source. In addition, when women identify with other social identities (e.g., intelligent student) instead of their gender identity, they perform better on math tests (McGlone & Aronson, 2006). Thus, teachers could implement ways to remind stereotype vulnerable students of their more achieving identities.

What the present study demonstrates and adds to the literature is that using stereotype lift can also decrease the effects of stereotype threat among women. In other words, by lifting the stereotype threat, the detrimental effects on math performance can be removed.

For men, it appears that stereotype threat causes the best performance in math. However, future research needs to be conducted before these conclusions can be fully drawn. First, future research should examine whether challenging men improves math test scores to the same degree as presenting them with stereotype threat. If it does, this may prove to have less mental and social consequences than negatively stereotyping them. In addition, men's social identity needs to be made more salient before fully concluding that stereotype lift is not effective in improving math scores. Future research should also assess feelings of stereotype threat experienced by men and women.

Because attitudes toward mathematics and one's own mathematical ability are correlated with mathematics achievement (Tartre & Fennema, 1995), the findings from our study may have a direct effect on classrooms. On the basis of our results, more attention needs to be given to build confidence in students' abilities in mathematics early in their education. Girls demonstrate less self-confidence in their mathematical ability than do boys quite early (Tiedemann, 2000). It is especially a high priority to reverse the rapidly declining numbers of women in STEM—science, technology, engineering, and math—careers now that the government is investing money to improve opportunities for students to attain high standards in these areas. Many women, however, are not entering these fields because they see those professions as being dominated by men and as not being aligned with their interests (Eccles & Wigfield, 1995). Despite progress in the area of math and science careers for women, women continue to be less likely to enter science, math, and technology professions (National Science Foundation, 2000).

Limitations

One limit of the present study was the sample used. Individuals in this study are fairly homogenous in nature. For example, participants were university students, with an average age of 22 years, and from middle to upper class socioeconomic statuses. In addition, only 19% of the students contacted participated in the study. Thus, future research should examine this phenomenon in more heterogeneous samples.

In this study, participants were not asked to indicate in which subject they are studying. We tried to control for variables that might affect our ability to generalize the results by randomly selecting our sample. However, the population from which we gathered our participants may be cause for concern. We accessed women from a university who represent a high-ability based sample. Consequently, there is a limit to which we can use our results to generalize measured perceptions about math ability across various populations.

Despite these limitations, given that the focus of the present study was with respect to gender, we consider findings from our sample data with respect to gender to be particularly generalizable given the college gender gap in the United States where more women attend institutions of higher education than do men (Marklein, 2010), as is the case of the institution in the present study.

Another limit of the study is that we did not test any attenuators or possible mediators. Are there certain variables that could be attenuating or mediating the effects of the stereotype threat

or lift conditions? Future research should begin addressing these possible mechanisms in order to inform teachers and parents more clearly of ways to improve math performance.

Conclusion

Our study highlights the differences between the ways men and women perform under stereotype threat and lift. Although studies have shown that men perform better under stereotype threat conditions than women (Williams, 2006), our study is the first to demonstrate how men and women perform on math tests under conditions of non–stereotype threat, stereotype threat, and stereotype lift. These new findings add greatly to the currently sparse literature on stereotype lift.

AUTHOR NOTES

Heather J. Johnson is a graduate student in the Department of Educational Psychology at Baylor University. **Lucy Barnard-Brak** is an assistant professor in the Department of Educational Psychology at Texas Tech University. **Terrill F. Saxon** is an associate professor in the Department of Educational Psychology at Baylor University. **Megan K. Johnson** is a fifth-year Ph.D. candidate in the Department of Psychology and Neuroscience at Baylor University. Her research areas include prejudice, stereotyping, cognitive rigidity, and religiosity.

REFERENCES

Aronson, J., & Lustina, M., Good, C., Keough, K., Brown, J. L., & Steele, C. M. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, *35*, 29–46.

Barnard, L., Burley, H., Olivarez, A., & Crooks, S. (2008). Measuring vulnerability to stereotype threat. *Electronic Journal of Research in Educational Psychology*, *6*, 51–64.

Been-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, *41*, 174–181.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, *59*, 960–970.

Beyer, S., & Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, *23*, 157–172.

Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology*, *76*, 246–257.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

College Board. (2009). *SAT*. Retrieved from http://www.collegeboard.com/student/testing/sat/about/SATI.html

Eccles, J. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, *11*, 135–172.

Eccles-Parsons, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco, CA: Freeman.

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

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, *73*, 31–44.

Green, S. B., & Salkind, N. G. (2004). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Josephs, R., Newman, M., Brown, R., & Beer, J. (2003). Status, testosterone, and human intellectual performance: Stereotype threat as status concern. *Psychological Science*, *14*, 158–163.

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, *20*, 975–978.

Liu, J., Schuppan, F., & Walker, M. E. (2005). *Evaluating SAT II: Mathematics IC items in the SAT I population*. Retrieved from http://www.collegeboard.com

Lundeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly confident but wrong: gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, *86*, 114–121.

Marklein, M. B. (2010). *College gender gap remains stable: 57% women*. Retrieved from http://www.usatoday.com/news/education/2010-01-26-genderequity26_ST_N.htm

Marx, D., Brown, J. L., & Steele, C. M. (1999). Allport's legacy and the situational press of stereotypes. *Journal of Social Issues*, *55*, 491–502.

Marx, D., & Stapel, D. A. (2006). Understanding stereotype lift: On the role of social self. *Social Cognition*, *24*, 776–791.

McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, *27*, 486–493.

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 Science Foundation. (2000). *Women, minorities, and persons with disabilities in science and engineering*. Arlington, VA: Author.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, *122*, 1067–1101.

Quinn, D., & Spencer, S. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, *57*, 55–71.

Spencer, S. J. (1993). *The effect of stereotype vulnerability on women's math performance*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Spencer, S. J., Steele, C. M., & Quinn, D. (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. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.

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

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*, 46–51.

Tartre, L. A., & Fennema, E. (1995). Mathematics achievement and gender: A longitudinal study of selected cognitive and affective variables (Grades 6–12). *Educational Studies in Mathematics*, *28*, 199–217.

Thoman, D. B., White, P. H., Yamawaki, N., & Koishi, H. (2008). Variations of gender-math stereotype content affect women's vulnerability to stereotype threat. *Sex Roles*, *58*, 702–712.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, *92*, 144–151.

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles*, *41*, 219–240.

Walton, G., & Cohen, G. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, *127*, 797–826.

Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology*, *80*, 210–216.

Williams, K. (2006). The effects of stereotype threat on test performance of male and female college students. *College Student Journal*, *40*, 679–684.