BASIC AND APPLIED SOCIAL PSYCHOLOGY, 34:534–543, 2012

ISSN: 0197-3533 print/1532-4834 online
DOI: 10.1080/01973533.2012.727319

Psychology Press
Taylor & Francis Group

Calculus GPA and Math Identification as Moderators of Stereotype Threat in Highly Persistent Women

Julia R. Steinberg
University of California, San Francisco

Morris A. Okun and Leona S. Aiken
Arizona State University

The present research tested whether the effect of stereotype threat on calculus performance was moderated by calculus GPA and math identification in advanced undergraduate women majoring in science, technology, and engineering (STEM) fields. Women ($n = 102$) were randomly assigned to one of three conditions—stereotype threat, gender equivalence, or no mention (of gender). Confirming stereotype threat theory, at high levels of calculus GPA and math identification, women performed the worst in the stereotype threat condition, intermediate in the gender equivalence condition, and best in the no mention condition. Strategies to counter the inimical effects of stereotype threat are discussed

Women in science, technology, engineering, and mathematics (STEM) fields are aware that others may judge their behavior, including their performance, in light of the stereotype that women have inferior mathematics skills relative to men (Seymour & Hewitt, 1997). Despite this awareness, some women successfully persist in STEM fields. For instance in 2007, 18.5% of engineering, 21.0% of physics, and 43.9% of mathematics bachelor degrees were awarded to women (National Science Foundation, Division of Sciences Resources Statistics, 2009). Some of these persistent women—particularly those with the greatest skills and strongest identification with these fields—may be susceptible to stereotype threat effects. The possibility of being judged in a situation by a negative stereotype about one's group is termed a stereotype threat situation (Steele, 1997). The present research employs stereotype threat theory (Steele, 1997) to address whether two individual difference characteristics—domain skill level and domain identification—moderate the impact of negative stereotypes on women's math test performance. We focus on the population of upper-level undergraduate women who are successfully majoring in STEM fields because such effects are postulated to be the strongest for those with the most skill, identification, and success in academic fields (Steele, 1997).

According to the originator of stereotype threat theory, Claude Steele (1997),

> stereotype threat should have its greatest effect on the better, more confident students in stereotyped groups, those who have not internalized the group stereotype to the point of doubting their own ability and have thus remained identified with the domain—those who are in the academic vanguard of their group. (p. 617)

Stereotype threat theory contends that stereotype threat increases for those in the stigmatized group with more domain identification and skill level. Although this proposition is generally accepted among stereotype threat scholars and has been demonstrated for some groups in which academic stereotypes exist (e.g., White men compared to Asian men regarding math; Aronson et al., 1999), little research has empirically tested these propositions, particularly for the stereotype regarding women's inferior math abilities relative to men's. To test whether this proposition holds among women, the

Correspondence should be sent to Julia R. Steinberg, Department of Psychiatry, University of California, San Francisco, 3333 California Street, Suite 465, Box 0848, San Francisco, CA 94143-0848. E-mail: julia.steinberg@ucsf.edu

present research investigates whether the impact of stereotype threat on calculus test performance is stronger for women persisting in STEM fields with higher calculus grade point averages (GPAs) and with higher math identification, compared to those with lower calculus GPAs and math identification. We first examine math identification and calculus GPA as individual moderators and then examine whether they jointly influence the relationship between stereotype threat and calculus test performance. This is the first study to examine whether the individual characteristics of domain skill and identification *synergistically* influence the effect of stereotype threat on test performance.

UNTESTED PREDICTIONS FROM STEREOTYPE THREAT THEORY

Steele (1997) proposed that stereotype threat affects those most identified with or self-invested in the academic domain and those who have survived structural barriers in the domain. Specifically, those most identified with the academic domain are expected to perform poorly in the domain when under stereotype threat, whereas those least identified with the domain are not expected to suffer stereotype threat effects because they are not invested in performing well. Individuals who have high math identification suffer in stereotype threat situations because they are invested in the domain, and experience a level of arousal that interferes with cognitive capacities, particularly for difficult tasks like the one used in the current study (Aronson et al., 1999; Jamieson & Harkins, 2007, 2009; Keller, 2007; Schmader & Johns, 2003). Because individuals with low math identification do not experience such interference, they are not expected to perform poorly in stereotype threat situations.

Surprisingly little research has examined whether math identification moderates the relationship between stereotype threat and test performance for the stereotype about women's inferior math skills relative to men's. Even though research has examined the effects of stereotype threat in samples of women who are deemed highly identified with math (e.g., Brown & Pinel, 2003; Carr & Steele, 2009; Inzlicht & Ben-Zeev, 2003; Lesko & Corpus, 2006; Logel, Iserman, Davies, Quinn, & Spencer, 2009; Marx & Roman, 2002; Spencer, Steele, & Quinn, 1999), only one study has examined whether the performance effects of stereotype threat vary as a function of women's math identification (Keller, 2007). Keller (2007) postulated and found that high school girls highly identified with math performed worse on difficult math items when in a threatening as opposed to a nonthreatening condition, whereas those with low identification performed better in a threatening versus nonthreatening condition. This study did not sample women persisting in STEM fields who have overcome substantial obstacles in a male-dominated field. Based on stereotype threat theory, we expected that there would be a significant interaction between stereotype threat and math identification on a difficult math test such that the deleterious effect of stereotype threat on calculus test performance would increase as math identification increases (H1).

It remains to be seen whether women in STEM fields with low math identification exhibit any stereotype threat effects. Among those with low math identification, researchers have found better performance in the threatening relative to the nonthreatening conditions among White men threatened by the stereotype about White versus Asian men's math abilities and among high school girls threatened by the stereotype about women's versus men's math abilities (Aronson et al., 1999; Keller, 2007). These findings suggest that for those with low math identification, threatening conditions are appropriately motivating, whereas nonthreatening conditions are not. Because in the current study, we model math identification as a moderator, we are able to explore whether women in STEM fields with low math identification perform better in a threatening versus nonthreatening condition.

Another characteristic that increases vulnerability to the effects of stereotype threat in academic domains is skill level. According to stereotype threat theory, individuals with greater skill are expected to be more negatively affected by stereotype threat (Steele, 1997), and those with lower skill are not expected to suffer stereotype threat effects, because they don't have the skill level to perform well regardless of the threatening nature of the situation. However, little to no research has examined whether domain skill in scholastic fields interacts with stereotype threat to influence performance. Skill level has not been ignored in the stereotype threat literature and is frequently operationalized by a measure of previous performance in the domain. Instead of modeling skill level as a moderator variable, studies have typically controlled for skill level in order to test whether the effect of stereotype threat is statistically significant beyond that of skill level or to increase the power to detect the effects of stereotype threat on performance (Steele & Aronson, 2004). Many studies examining the effect of the stereotype about women's math abilities have sampled those with relatively high SAT math scores (e.g., Martens, Johns, Greenberg, & Schimel, 2006; Marx & Roman, 2002; Quinn & Spencer, 2001; Schmader, 2002; Schmader & Johns, 2003; Shih, Pittinsky, & Ambady, 1999; Spencer et al., 1999) or have controlled for SAT math scores for these reasons (e.g., Gonzales, Blanton, & Williams, 2002; Hollis-Sawyer & Sawyer, 2008; Inzlicht & Ben-Zeev, 2000, 2003; Keller, 2002, 2007; Lesko & Corpus, 2006; Marx & Roman, 2002; Schmader, 2002; Schmader & Johns, 2003; Steele & Aronson, 1995; Vick, Seery, Blascovich, & Weisbuch, 2008; Wout, Danso, Jackson, & Spencer, 2008; Wout, Shih, Jackson, & Sellers, 2009). As Steele (1997) noted, it is those with greatest skill who are expected to suffer the most from stereotype threat, implying those with less skill in the domain will suffer less or not at all. Here we test this hypothesis explicitly. Although we expect those with higher math skill to suffer stronger stereotype threat effects than those with lower math skill, it remains to be seen whether those with lower skill levels (but sufficient levels to be persisting in the domain) suffer any negative effects of stereotype threat.

For the outcome of performance on a set of calculus problems, we believe previous performance in calculus classes is a more appropriate measure of skill level than SAT math scores because the SAT does not include calculus problems. In the current study, because the task which the participants performed was calculus based, we used prior GPA derived from grades in calculus classes rather than prior SAT math scores. When appropriate, other studies have used prior grades in math courses as the measure of skill (Hollis-Sawyer & Sawyer, 2008; Keller, 2002, 2007). Similar to math identification, the adverse effect of stereotype threat on calculus test performance was expected to increase as calculus GPA increased (H2). Those with low calculus GPAs were expected to perform poorly across stereotype threat conditions.

SYNERGISTIC INTERACTION OF MATH IDENTIFICATION, CALCULUS GPA, AND STEREOTYPE THREAT

Although individuals persisting in the math and science fields are likely to have high levels of math identification and skill in the math domain relative to those not persisting in these fields, there may be some variability in these characteristics. Moreover, "at each level of schooling, it [stereotype threat] affects the vanguard of these groups, those with the skills and self-confidence to have identified with the domain" (Steele, 1997, p. 614). Therefore, even among upper-level undergraduate women majoring in STEM fields, those with the greatest skill and identification should be more strongly affected by stereotype threat situations than those with less skill and identification. Individuals who are able to perform well in a domain and who also are highly invested in a domain may be the most susceptible to the pernicious effects of stereotype threat. In other words, students who are both capable of performing well and highly motivated to perform well have the greatest potential to be undermined by stereotype threat. Although we expect students with high levels of math identification or calculus skill to suffer more stereotype threat effects (H1 and H2), we further hypothesize that calculus skill and math identification will have a synergistic effect on the magnitude of the impact of stereotype threat. More specifically, we predict that women with high levels of both characteristics will be most severely affected by stereotype threat because they have the most to lose in a stereotype threat situation. Consequently, we expect a three-way interaction between stereotype threat, calculus skill, and math identification on math test performance. The strongest inimical effects of stereotype threat were expected to be observed among women with high levels of both individual difference variables, and the effect of stereotype threat on calculus test performance was expected to be attenuated if the level of either individual difference variable was lower (H3).

CREATING A STEREOTYPE THREAT SITUATION FOR WOMEN IN STEM FIELDS

Our manipulation of stereotype threat consisted of presenting information regarding men's and women's performance on the calculus task that the women were about to perform. More specifically, (a) women were told that there were gender differences such that men had performed better than women, (b) women were told that there were no gender differences and men and women had performed equally, or (c) no mention was made of gender. To create a stereotype threat situation, some researchers have directly stated that men have outperformed women (or that women have performed worse than men) on the test (or type of task) that participants are about to perform (e.g., Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Keller, 2002; Smith, Sansone, & White, 2007, Smith & White, 2002; Stangor, Carr, & Kiang, 1998). As in these studies, we reasoned that directly telling women in STEM fields that men had performed better than women on the task in which they were about to perform would induce stereotype threat. To create a nonstereotype threatening or safe condition, several studies have explicitly stated that women and men have performed equally or that there were no gender differences on the task in which participants were about to perform (Cadinu et al., 2003; Good, Aronson, & Harder, 2008; Keller, 2007; Smith et al., 2007; Smith & White, 2002; Spencer et al., 1999; Stangor et al., 1998). As in these studies, we created a nonthreatening condition by explicitly stating both that no gender differences had been found and that men and women had performed equally on the task in which they were about to perform.

For women persisting in STEM fields, it is not clear whether working on calculus problems with no stereotype information would induce an implicit stereotype threat or serve as a safe, nonthreatening condition. Some studies have found that not mentioning gender or men's and women's math performance has served as a nonthreatening condition (Keller, 2002; Rydell, McConnell, & Beilock, 2009; Rydell, Rydell, & Boucher, 2010; Schmader, 2002), whereas others have found that it has (implicitly) served as a threatening condition (Campbell & Collaer, 2009; Good et al., 2008; Smith et al., 2007; Smith & White, 2002; Spencer et al., 1999). We suspected that not mentioning gender for women who are persisting in STEM fields might be a safe condition because these women have defied the stereotype, overcome obstacles in the stereotyped domain, and frequently had their performance evaluate on tasks similar to the one employed herein. To test this, we included a condition in which no mention of gender was made.

By examining different combinations of calculus GPA and math identification level in one model, we expected to glean a more complete picture of which women persisting in STEM fields are affected under stereotype threat conditions and to whom programs addressing stereotype threat effects should be directed. In addition, by investigating which manipulations undermine performance, which enhance performance, and which women are negatively affected, we hoped to gain better understanding of how to help women persisting in mathematics fields maintain high levels of academic achievement.

METHOD

Participants

To be eligible for this study, participants had to be female junior or senior college students majoring in engineering, math, or physical science fields and to have completed all three semesters of calculus required for these concentrations. Recruitment techniques included e-mails, in-class or club announcements, flyers, and referral by professors or friends telling participants about the study. Incentives to participate included a raffle for $150, extra credit, or class credit. One hundred four women participated in this study.

Procedure and Manipulations

The procedure took place in two parts. First, at least 1 week before the laboratory part of the study, participants completed an online questionnaire that included the measure of math identification described next. Then, participants came to a room with individual cubicles and computers to complete the second part. A female experimenter met each participant at the door to the room and waited until all participants arrived. All participants completed the second part with no, one, or two other participants in the room. Each participant was seated at a cubicle that contained a computer and headphones. Participants were told that they would be completing the study individually on the computer and that the computer would signal when the study was completed. Each participant received a piece of scratch paper for the problems and was told to put on her headphones. Then, the experimenter began the MediaLab computer program for the participant(s) and left the room. The computer program included the manipulations (to which subjects were randomly assigned), calculus problems, and some posttest measures.

After providing informed consent, but before exposure to the stereotype threat manipulations, each participant heard and read that the study's investigators were working in conjunction with other departments to "understand how individuals persist in engineering, math, and physical science fields." All participants were also informed that they were going to work on some calculus problems similar to the ones that they had been exposed in their calculus classes and that they would have 30 min to work on 15 problems. Further, participants were told they should indicate their answer by checking the box next to their answer.

Then participants heard and read the stereotype threat message. In the stereotype threat (ST) condition, women were told there were gender differences such that men had performed better than women ($n = 35$) on this task. In the gender equivalence (GE) condition, participants were told that there were no gender differences and that men and women had performed equally ($n = 36$). Finally, in the no mention of gender (NM) condition, participants received no information at all about men's and women's performance ($n = 31$).

Participants then worked on up to 15 calculus problems for up to 30 min. Problems were presented individually on the computer screen. Participants marked an answer and pressed continue to move to the next problem. Participants did not have to mark an answer to continue to the next problem, but they were aware they could not return to a problem once they pressed continue to go to the next problem. After working on the math problems, participants completed a posttest questionnaire that included items assessing grades in their completed calculus college-level courses. Once the posttest questionnaire was completed, participants read a debriefing form and were free to ask the experimenter questions.

Measures

Calculus test. In the present study, a 15-item calculus test used by Good et al. (2008) for a similar sample of students was employed to assess performance. The items were multiple choice with five options per problem and were taken from the GRE Mathematics Subject Exam.

Dependent Variable

Performance measure. Total number of problems answered correctly was the measure of performance.[1]

Moderator Variables

Math identification. Math identification was assessed in the pretest questionnaire with four items adapted from Schmader's (2002) gender identification measure. Women rated their agreement on a 5-point scale (0 to 4) to the following statements: 1. Being good at math is an important part of my self-image, 2. Being good at math is unimportant to my sense of what kind of person I am, 3. Being good at math is an important reflection of who I am, and 4. Being good at math has very little to do with how I feel about myself. Items 2 and 4 were reverse scored. The Cronbach alpha for this scale in this sample was .79.

Calculus GPA. In the posttest questionnaire, participants indicated their grades for each college-level calculus course they had completed. Calculus GPA was computed by averaging their grades in their calculus courses.

RESULTS

Analyses are reported based on the data from 102 women; one woman was excluded for failure to report calculus GPA, the other due to a missing math identity score. The mean calculus GPA was 3.4 ($SD = 0.5$), and ranged from 2.0 (grade of C) to 4.3 (grade of A+). In all, 80.4% of women had GPAs of 3.0 or above, that is, at least a B average in their calculus classes. The mean math identity score was 2.2 ($SD = 0.9$) and ranged from 0 to 4. Sixty-two percent of the women had math identity scores at or above 2.0, the midpoint of the scale. The correlation between calculus GPA and math identification was .09 ($p = .36$), indicating that these characteristics could be treated as separate moderator variables. Calculus GPA and math identification did not differ by ST condition (ps $> .20$).

H1: Does math identification moderate the relationship between stereotype threat and performance?

To test our first hypothesis that math identification moderated the relation between stereotype threat and test performance, we conducted hierarchical linear regression in which we entered math identification in the first step, two dummy codes for the main effect of stereotype threat in the second step, and the interaction between stereotype threat and math identification in the third step. Math identification was centered at its mean. In the first step, we found a main effect for math identification, $F(1, 100) = 6.39$, $p = .01$, $R^2 = .06$, such that as math identification increased, so too did performance. In the second step, we found a main effect for stereotype threat above and beyond math identification, $F(2, 98) = 3.58$, $p < .05$, $R^2_{change} = .06$. In the third step, we found a marginally significant interaction between math identification and stereotype threat, $F(2, 96) = 2.60$, $p < .08$, $R^2_{change} = .045$. To probe this two-way interaction, we followed recommendations of Aiken and West (1991) to examine the simple slopes of calculus performance on math identification within each of three stereotype threat conditions. As seen by the simple slopes depicted in Figure 1, for women in the ST and GE conditions, there was no significant association between math identity and calculus performance, $B = .39$ ($SE = .43$) and .25 ($SE = .43$), ps $> .35$, respectively. In contrast, in the NM condition, the relationship between math identity and calculus performance was significant, $B = 1.49$ ($SE =$ .42) $p = .001$. The difference in slopes between NM and GE was significant, $p < .05$, whereas that between NM and ST was marginally significant, $p = .07$.

Because we were interested in whether differences between stereotype threat conditions existed as math identification level increased, we followed procedures outlined by Aiken and West (1991, pp. 132–133) to test whether predicted means differed between conditions at low, mean, and high levels of math identification. We tested differences at the arithmetic mean of math identification as well as 1 SD above and 1 SD below the mean. Consequently, high and low math identification signify 1 SD above and below the mean of math

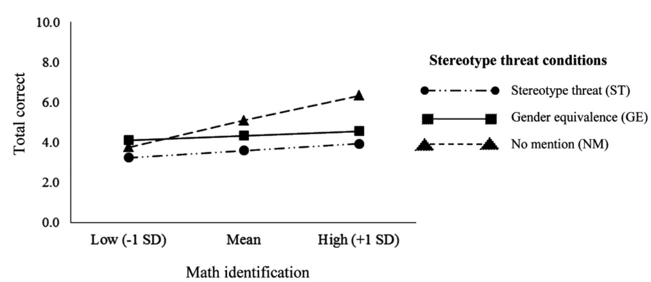

FIGURE 1 Predicted total correct by stereotype threat and math identification.

[1]Some women did not answer all 15 items. If we conduct these analyses using percentage correct of those items answered, analyses do not change for the two-way interaction of stereotype threat with calculus GPA or the three-way interaction with calculus GPA, math identification, and stereotype threat. However, the two-way interaction of stereotype threat with math identification is not significant ($p > .5$).

identity, respectively. In support of H1 (see Table 1 and Figure 1), women with higher math identification were more strongly affected by stereotype threat than women with lower math identification. More specifically, at high math identification, women in the NM condition ($M = 6.4$) performed better than women in both the ST condition ($M = 3.9$) $p = .001$ and in the GE condition ($M = 4.6$) $p = .02$, and the difference between the means in the latter two conditions was not significant, $p > .30$. At mean math identification, there was one significant difference—women in the NM condition ($M = 5.1$) once again performed better than women in the ST condition ($M = 3.6$) $p < .01$. Together, these results suggest that the least threatening condition was the one in which no mention of gender was made.

H2: Does calculus GPA level moderate the relationship between stereotype threat and performance?

To test our second prediction that calculus GPA moderated the relation between stereotype threat and test performance, as predicted in H2, we carried out a similar hierarchical linear regression model, except that we replaced math identification with calculus GPA. In Step 1, we found a significant main effect of calculus GPA, $F(1, 100) = 4.43$, $p < .05$, $R^2 = .04$, such that as calculus GPA increased so too did calculus performance. In Step 2, we found a significant main effects of stereotype threat above and beyond calculus GPA, $F(2, 98) = 3.35, p < .05$, $R^2_{change} = .06$. Finally, in Step 3, we found a significant interaction between stereotype threat and calculus GPA, $F(2, 96) = 3.93$, $p = .02$, $R^2_{change} = .07$. To obtain the simple slopes within conditions and predicted mean at various levels of calculus GPA, we followed the same procedures just described. Simple slopes analyses revealed that the slope of the regression of total correct on calculus GPA was negative (though not significantly so) in the ST condition, $B = -.36$ ($SE = .70$) $p = .60$; rose to positive (though not significantly so) in the GE condition, $B = 1.09$ ($SE = .70$) $p = .13$; and rose to significantly positive in the NM condition, $B = 2.44$ ($SE = .72$) $p = .001$. Only the slopes between the NM and ST conditions differed significantly, $p = .006$.

In support of H2 (see Table 2 and Figure 2), women with higher calculus GPAs were more strongly affected by stereotype threat than women with lower calculus GPAs. More specifically, at high calculus GPA, women in the ST condition ($M = 3.5$) performed worse than women in the GE condition ($M = 4.9$) $p < .05$, and in the NM condition ($M = 6.4$) $p < .0005$, and the difference between the means in the latter two conditions was marginally significant, $p < .07$. At mean calculus GPA, there was one significant difference—women in the NM condition ($M = 5.1$) once again performed better than women in the ST condition ($M = 3.7$), $p = .01$.

H3: Do calculus GPA and math identification synergistically interact with stereotype threat to influence performance?

To test our third hypothesis that math identification and calculus GPA synergistically moderated the effect of stereotype threat on test performance, we again used a hierarchical linear regression model. In this model, the order of entry of the terms was as follows: (a) the main effects of centered calculus GPA and centered math identification; (b) the main effect of stereotype threat; (c) the two-way interactions between (i) the two continuous measures and (ii) each of the continuous measures and the categorical variable of stereotype threat; and (d) the three-way interaction between stereotype threat, calculus GPA, and math identification. In Step 1, we found significant prediction from math identification and calculus GPA, $F(2, 99) = 5.14$, $p = .008$, $R^2 = .09$, with positive associations between math identification and total correct, $B = .60$ ($SE = .25$) $p < .02$, and between calculus

TABLE 1
Predicted Total Correct by Stereotype Threat Condition and Math Identification

	Total Correct		
	Math Identification		
Stereotype Threat Condition	*Low*	*Mean*	*High*
Stereotype threat	3.2 (0.6)	3.6_a (0.4)	3.9_a (0.5)
Gender equivalence	4.1 (0.5)	$4.3_{a,b}$ (0.4)	4.6_a (0.5)
No mention	3.8 (0.5)	5.1_b (0.4)	6.4_b (0.6)

Note. Low, mean, and high are 1 *SD* below the mean, the mean, and 1 *SD* above the mean, respectively. Values in parentheses are standard errors. Within math identification levels, values with different subscripts represent significant mean differences between conditions ($p < .05$).

TABLE 2
Predicted Total Correct by Stereotype Threat Condition and Calculus GPA

	Total Correct		
	Calculus GPA Level		
Stereotype Threat Condition	*Low*	*Mean*	*High*
Stereotype threat	3.9 (0.8)	3.7_a (0.5)	3.5_a (0.5)
Gender equivalence	3.8 (0.9)	$4.4_{a,b}$ (0.6)	4.9_b (0.5)
No mention	3.8 (0.6)	5.1_b (0.5)	6.4_b (0.6)

Note. Low, mean, and high are 1 *SD* below the mean, the mean, and 1 *SD* above the mean, respectively. Values in parentheses are standard errors. Within calculus grade point average (GPA) levels, values with different subscripts represent significant mean differences between conditions ($p < .05$).

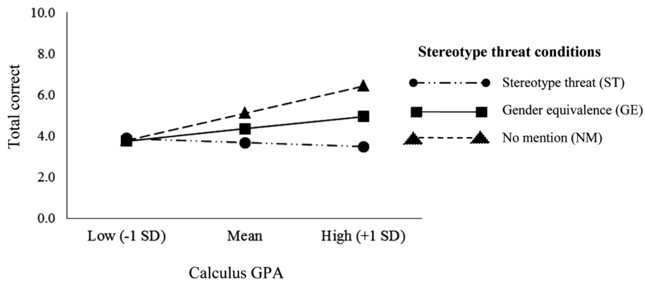

FIGURE 2 Predicted total correct by stereotype threat and calculus grade point average (GPA).

GPA and total correct, $B = .80$ ($SE = .42$) $p < .06$. In Step 2, we found a significant effect of stereotype threat, $F(2, 97) = 4.46$, $p = .01$, $R^2_{change} = .08$. In Step 3, the addition of the three two-way interactions significantly increased the amount of variance explained, $F(5, 92) = 2.31$, $p = .05$, $R^2_{change} = .09$. In support of H3, these effects were further qualified by a significant Stereotype Threat × Calculus GPA × Math Identification three-way interaction effect, $F(2, 90) = 7.81$, $p = .001$, $R^2_{change} = .11$.

To examine the form of the interaction, we generated the simple slopes of calculus performance on calculus GPA within stereotype threat condition and at low, mean, and high levels of math identification following the aforementioned procedures outlined by Aiken and West (1991), where low and high signify 1 *SD* above the mean of math identification. The simple slopes are graphed in Figure 3. At low math identification (Figure 3a), two marginally significant slopes were found in the NM, $B = 1.35$ ($SE = .73$) $p < .07$, and ST conditions, $B = 1.87$ ($SE = .97$) $p < .06$; the slope in the GE condition was not significant, $B = .20$ ($SE = 1.00$) $p > .80$. There were no significant differences in slopes between conditions, *p*s > .20. At the sample mean level of math identification, total correct increased significantly as a function of GPA in the NM condition, $B = 2.79$ ($SE = .73$) $p < .0005$, and marginally significantly in the GM condition, $B = 1.25$ ($SE = .65$) $p < .06$. In contrast, there was a slightly (though nonsignificantly) negative slope in the ST condition, $B = -.18$ ($SE = .64$) $p = .78$. The slopes between the ST and NM conditions differed significantly, $p = .003$. At high math identification (Figure 3c), there were dramatically different slopes; there was a significant positive slope in the NM condition, $B = 4.24$ ($SE = 1.19$) $p = .001$, and a marginally significant positive slope in the GE condition, $B = 2.29$ ($SE = 1.18$) $p < .07$. In contrast, performance declined significantly as a function of calculus GPA in the ST condition, $B = -2.22$ ($SE = .86$) $p = .01$. Differences in slopes were found between the ST and the GE, $p = .003$, and NM conditions, $p < .0005$.

To examine whether the differences between conditions at various levels of math identification and calculus GPA were in accord with H3, we estimated predicted means at all combinations of math identification (-1 *SD, M*, $+1$ *SD*) and calculus GPA (-1 *SD, M*, $+1$ *SD*) for each stereotype threat condition. Figure 3 and Table 3 present the predicted mean test performance by stereotype threat and calculus GPA at low, mean, and high levels of math identification. At high levels of both math identification and calculus GPA, women in the ST condition ($M = 3.1$) answered fewer items correctly than did women in both the NM ($M = 8.8$) $p < .0005$, and GE conditions ($M = 5.7$) $p < .01$. In addition, women in the GE condition answered fewer items correctly compared to women in the NM condition, $p < .01$. In sum, the combination of high math identification with high GPA in the ST condition led to notably low performance (see Table 3 and Figure 3c). Similar findings were obtained among women at high calculus GPA and mean math identification and among women at mean calculus GPA and high math identification, though the differences by stereotype threat conditions were smaller than when individuals were high on both characteristics.

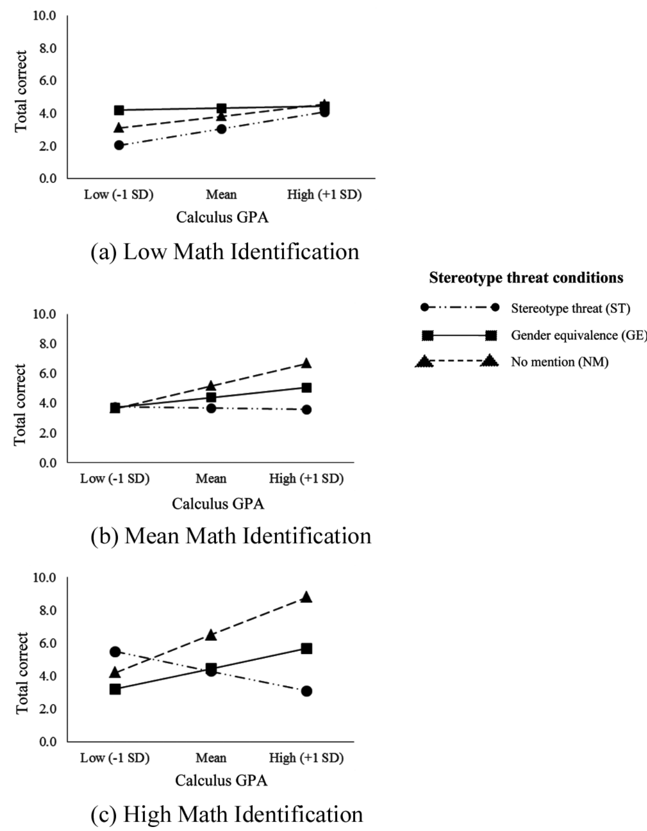

FIGURE 3 (a) Predicted total correct by stereotype threat and calculus grade point average (GPA) at low math identification. (b) Predicted total correct by stereotype threat and calculus GPA at mean math identification. (c) Predicted total correct by stereotype threat and calculus GPA at high math identification.

TABLE 3
Predicted Total Correct by Stereotype Threat Condition, Math Identification Level, and Calculus GPA Level

Math Identity Level	*Stereotype Threat Condition*	*Calculus GPA Level*		
		Low	*Mean*	*High*
Low	Stereotype threat	2.0 (0.8)	3.0 (0.5)	4.1 (0.7)
	Gender equivalence	4.2 (0.8)	4.3 (0.5)	4.4 (0.6)
	No mention	3.1 (0.5)	3.8 (0.5)	4.6 (0.7)
M	Stereotype threat	3.8 (0.5)	3.7_{a} (0.3	3.6_{a} (0.5)
	Gender equivalence	3.7 (0.5)	$4.4_{a,b}$ (0.3)	5.1_{b} (0.5)
	No mention	3.7 (0.5)	5.2_{b} (0.4)	6.7_{c} (0.6)
High	Stereotype threat	5.5 (0.7)	4.3_{a} (0.5)	3.1_{a} (0.6)
	Gender equivalence	3.2 (0.9)	4.4_{a} (0.5)	5.7_{b} (0.7)
	No mention	4.2 (0.8)	6.5_{b} (0.5)	8.8_{c} (0.9)

Note. Low, mean, and high are 1 *SD* below the mean, the mean, and 1 *SD* above the mean, respectively. Values in parentheses are standard errors. Within math identification and calculus grade point average (GPA) levels, values with different subscripts represent significant mean differences between conditions ($p < .05$).

At mean levels of both math identification and calculus GPA, women in the NM condition ($M = 5.2$) once again performed better than women in the ST condition ($M = 3.7$, $p < .01$), whereas those in the GE condition ($M = 4.4$) did not significantly differ from women in either of the other conditions. Of interest, at low levels of both math identification and calculus GPA, women in the GE condition ($M = 4.2$) tended to perform better than women in the ST condition ($M = 2.0$) $p = .06$. Furthermore, among women with low calculus GPA and high math identification, there was a tendency for women in the ST condition ($M = 5.5$) to answer *more* items correctly than women in the GE condition ($M = 3.2$) $p < .06$.

DISCUSSION

The purpose of the current study was to test three hypotheses derived from stereotype threat theory pertaining to whether the effect of stereotype threat on calculus performance is amplified by domain identification, domain skill, and domain identification and skill together. Our focus was on women persisting in STEM domains.

Consistent with our first two hypotheses, we found that domain identification and skill each moderated (or tended to moderate) the relation between stereotype threat and calculus test performance such that those with higher calculus GPAs and greater math identification were more affected by stereotype threat than those with lower calculus GPAs and math identification. The Stereotype Threat × Calculus GPA interaction suggests that researchers collecting information on previous skill (or achievement) should test whether it moderates the effects of stereotype threat rather than simply use it as a covariate.

In support of our third hypothesis, the three-way interaction between stereotype threat, calculus GPA, and math identification was significant and in the expected direction. Consistent with stereotype threat theory, women at high levels of both characteristics were most strongly affected by stereotype threat, whereas women at a high level of one characteristic and at the mean level of the other characteristic were also negatively affected although to a lesser degree. Another way to interpret these findings is to compare the slopes between calculus GPA and total correct across stereotype threat conditions at high levels of math identification. In the NM condition, among strongly math identified women, there was a strong positive association between calculus GPA and total correct. However, in the ST condition, among strongly math-identified women, there was a significant negative association between calculus GPA and task performance. The inverse relationship observed between calculus GPA and performance in the ST condition is a novel finding in the stereotype threat literature and suggests another way that high math identification has a negative effect in ST conditions: It makes the expected positive association between previous calculus GPA and current task performance negative. This may occur among those who care greatly about performing well in the domain, because as domain skill level increases they may experience more anxiety and pressure to perform well, thereby causing performance to decrease.

Women at mean levels of both characteristics were also negatively affected, though to a lesser degree than when both characteristics were at high levels or one characteristic was at a high and the other was at a mean level. Relative to women at high or mean levels of both characteristics, at low levels of both characteristics women were less negatively affected. In fact, performance among those in both the ST and NM conditions was the lowest for those with low levels of both of these characteristics. This suggests that those with low levels of both characteristics were not motivated to perform well in the NM or ST conditions.

In contrast, those with low calculus GPA seemed motivated to perform well in the context of high math identification and stereotype threat. Indeed, the highest performance of women in the ST condition was among those with low calculus GPA and high math identity, and those in the ST condition tended to perform better than women in the GE condition. In the current sample, the women with the lowest calculus GPA received a C average in their calculus classes. This suggests that the performance of women with low calculus GPAs is malleable, provided they also have high math identity. It may be that these women are motivated but do not feel pressure to perform well in the stereotype threat situation (because they expect themselves to perform poorly), try harder, and consequently do better on the task.

In addition, similar to other research which has found that those with low math identification have a tendency to not suffer stereotype threat effects (e.g., Aronson et al., 1999; Keller, 2002), we found this tendency only among women with low math identification and high calculus GPA, no differences by stereotype threat condition. This suggests that one strategy to buffer the negative effects of stereotype threat, for women in STEM fields who have high skill level, may be to temporarily disengage from the math domain (Major & Schmader, 1998; Major, Spencer, Schmader, Wolfe, & Crocker, 1998) or to discount the validity of the task (Nussbaum & Steele, 2007). Arming women with the capacity to temporarily disengage may create a less threatening situation in the context of stereotype threat, thereby reducing the negative effects.

In the current study, the NM condition appeared to be the least threatening condition, suggesting that the safest environment for women in STEM fields may be one in which no attention is drawn to women's math performance relative to men's. Given the current well-publicized goal of increasing women in STEM fields in the United States, it is unlikely that such an environment can be created. Instead, we are more likely to create an environment in which women believe that no gender differences exist. Fortunately, those in the GE condition generally performed better than women in the ST condition, and at higher levels of the individual characteristics, the difference between these conditions were statistically significant.

Although stating that men perform better than women may be a blatant cue of stereotype threat (Keller, 2002), stating that men and women perform equally may be a subtle cue of stereotype threat for women in STEM fields because on a daily basis they witness the dearth of women in their fields. According to Stone and McWhinnie (2008), subtle cues may lead people to use cognitive resources to reduce uncertainty about the presence or not of bias, resulting in working memory and performance deficits. Thus, because it conflicts with their everyday experiences, information indicating the lack of gender differences in performance may capture cognitive resources. Future research could investigate whether this explanation accounts for why women in STEM fields are adversely affected in the no gender difference condition, which is generally regarded as the safest environment for women targeted by the negative stereotype about women's math abilities.

Conclusions

In support of hypotheses derived from stereotype threat theory, we showed that among women persisting in upper-level undergraduate math and science fields, the inimical effect of stereotype threat on performance were greatest for those with higher skill and math identification and decreased as skill level and math identification decreased. In addition, in contrast to other research, it appears that telling women in STEM fields that men and women perform equally may be a subtle cue which distracts women from the task at hand. These results suggest that creating a safe and nondistracting environment for women in STEM fields, particularly for those who have both the motivation and skill to perform well, allows them to perform significantly better than a threatening or distracting environment. Alternatively, teaching these women to temporarily disengage from the domain may boost their performance.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression:Test and interpreting interactions*. Newbury Park, CA: Sage.

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

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, *39*, 626–633.

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.

Campbell, S. M., & Collaer, M. L. (2009). Stereotype threat and gender differences in performance on a novel visuospatial task. *Psychology of Women Quarterly*, *33*, 437–444.

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

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

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Development*, *29*, 17–28.

Hollis-Sawyer, L. A., & Sawyer, T. P. (2008). Potential stereotype threat and face validity effects on cognitive-based test performance in the classroom. *Educational Psychology*, *28*, 291–304.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365–371.

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

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, *93*, 544–564.

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

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.

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

Lesko, A. C., & Corpus, J. H. (2006). Discounting the difficult: How high math identified women respond to stereotype threat. *Sex Roles*, *54*, 113–125.

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

Major, B., & Schmader, T. (1998). Coping with stigma through psychological disengagement. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 219–241). San Diego, CA: Academic Press.

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

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

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, *28*, 1183–1193.

National Science Foundation, Division of Sciences Resources Statistics. (2009). *Women Minorities, and Persons with Disabilities in Science and Engineering: 2009* (NSF Publication No. 09–305). Retrieved from http://www.nsf.gov/statistics/wmpd/about.htm

Nussbaum, A. D., & Steele, C. M. (2007). Situational disengagement and persistence in the face of adversity. *Journal of Experimental Social Psychology*, *43*, 127–134.

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

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

Rydell, R. J., Rydell, M. T., & Boucher, K. (2010). The effect of negative performance stereotypes on learning. *Journal of Personality and Social Psychology*, *99*, 883–896.

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

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440–452.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview.

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

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

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

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.

Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology*, *75*, 1191–1197.

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., & Aronson, J. A. (2004). Stereotype threat does not live by Steele and Aronson (1995) alone. *American Psychologist*, *59*, 47–55.

Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology*, *44*, 445–452.

Vick, S. B., Seery, M. D., Blascovich, J., & Weisbuch, M. (2008). The effect of gender stereotype activation on challenge and motivational states. *Journal of Experimental Social Psychology*, *44*, 624–630.

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

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