Female Role Models: Protecting Women's Math Test Performance

David M. Marx
Jasmin S. Roman
Harvard University

Recent theory and research suggest that certain situational factors can harm women's math test performance. The three studies presented here indicate that female role models can buffer women's math test performance from the debilitating effects of these situational factors. In Study 1, women's math test performance was protected when a competent female experimenter (i.e., a female role model) administered the test. Study 2 showed that it was the perception of the female experimenter's math competence, not her physical presence, that safeguarded the math test performance of women. Study 3 revealed that learning about a competent female experimenter buffered women's self-appraised math ability, which in turn led to successful performance on a challenging math test.

Few people would disagree with the idea that role models can have a profoundly positive impact on a person's life. Role models may differ in terms of their age, race, and/or sex, yet there is one feature that they share: They are all perceived to be competent in their respective areas (Lockwood & Kunda, 1997). For instance, female role models in math-related domains might be particularly helpful for math-talented women because they represent stereotype-disconfirming evidence about women's inferior math ability, so that women's math test performance is protected after encountering or learning about a female role model. In addition, if a low score on a standardized math test is one of the reasons why women are seriously underrepresented in math and engineering, then the benefits of having female role models for female students in those academic domains may be considerable.

GENDER DIFFERENCES IN MATH

Over the years, a vast amount of research has been devoted to investigating gender differences in math. Even though much of this work shows that such differences are relatively scarce (Hyde, Fennema, & Lamon, 1990; Kimball, 1989), there is a general pattern to the differences that do exist. Women, for instance, perform as well as men on less advanced math problems, but when the math problems become more advanced, women do not perform as well as men (Gallagher et al., 2000; Halpern, 1992; Kimball, 1989). This difference often does not emerge until high school (Hyde et al., 1990), precisely when performance on standardized math tests—tests that contain more advanced math problems—may matter the most.

But perhaps the most far-reaching gender difference is the fact that female students, unlike male students, routinely encounter negative stereotypes about their math ability (Halpern, 1992; Quinn & Spencer, 2001), which has been shown to disrupt their math test performance in the laboratory (Quinn & Spencer, 2001; Spencer, Steele, & Quinn, 1999) and presumably in standardized testing situations. The threat of negative stereotypes also can disrupt women's choice of major (Boswell, 1985; Hackett, 1985). Female students who do stay in math-related majors, however, still underperform, relative to men, on standardized math tests (Bridgeman & Wendler, 1991; Graduate Record Examinations Board,

Authors' Note: The research reported in this article was supported by a Grants-in-Aid award from The Society for the Psychological Study of Social Issues (SPSSI) to David M. Marx. We thank Macarena Correa, Alisha Creel, Morgan Goulet, Michelle Jamnongjit, Uthra Raghunathan, David Roddenberry, Sally Schoenvogel, Chris Soto, and Pam Vohra for their able assistance with the execution of these studies. Finally, we are grateful to Geoff Cohen, Sei Jin Ko, Sean Massey, Steve Most, Bill Swann, and the University of Colorado Stereotyping and Prejudice (CUSP) laboratory for their insightful comments on earlier versions of this article. Correspondence concerning this article should be addressed to David M. Marx, who is now at the Department of Psychology, The University of Colorado at Boulder, CB 345, Boulder, CO 80309; e-mail: dmarx@psych.colorado.edu.

PSPB, Vol. 28 No. 9, September 2002 1183-1193

1997). It seems then that even equivalent math backgrounds cannot override the impact of the negative gender stereotype. Indeed, simply being in a situation that is diagnostic of ability can remind women about this negative stereotype and thus impair their test performance (Marx, Brown, & Steele, 1999; Steele, 1997).

This predicament, known as stereotype threat (Steele, 1997), occurs not from internal doubt (i.e., believing the stereotype to be true of one's group) but from the concern of confirming the stereotype through one's actions. Take the case of highly skilled and motivated women who are just about to take a challenging standardized math test. Naturally, most students in this situation would feel discouraged if they perform poorly. But women face an additional threat—that of confirming a negative gender stereotype alleging their quantitative inferiority. In the end, this concern may be disturbing enough to harm the test performance of women who are otherwise motivated to perform well, particularly if performing well is also esteem-relevant.

Given the very real negative consequences of the gender stereotype on women's math test performance, it seems sensible to identify ways to protect their performance from being undermined by the negative stereotype. In this article, we argue that one way this may occur is by increasing the number of female role models in math-related domains. In fact, in a survey by Boswell (1985), she found that female mathematicians mentioned that female students "should be exposed to more positive female role models in mathematics" (p. 183; see also Gilbert, Gallessich, & Evans, 1983, for a related discussion). Having examples of women, who represent stereotype-disconfirming information, might mean that mathematically talented female students would not be as disturbed by the gender stereotype, thus buffering their math test performance.

FEMALE ROLE MODELS

Individuals become role models when others choose to emulate them. The search for role models is generally a selective, social comparison process whereby certain attributes are copied and others excluded (Collins, 1996; Javidan, Bemmels, Devine, & Dastmalchian, 1995). Of course, the attributes chosen will be ones that are also important to the chooser. Female students who care about mathematics, for instance, may be especially attuned to other women who have (or at least appear to have) expertise in that same domain, with the end result being that those women may become role models for the female students. Finally, the role models themselves play a passive part in this process and may not even be aware of having been chosen as role models by others (Fisher, 1988).

To date, very little research has actually been conducted on role models. The limited research, however, does outline some important benefits of comparing oneself to a similar and outstanding other. For instance, under particular conditions, role models can be inspirational (Lockwood & Kunda, 1997; Tesser, 1986), enhance self-evaluations and motivation (Blanton, Crocker, & Miller, 2000; Collins, 1996; Major, Sciacchitano, & Crocker, 1993; Taylor & Lobel, 1989), and even guide students' academic aspirations (Hackett, 1985; Lockwood & Kunda, 1997). Given the many positive effects that role models have on students, it is surprising then that no research has investigated how role models affect students' academic performance in the face of negative stereotypes. To the best of our knowledge, the research reported here is the first investigation of this sort.

Social comparisons with outstanding ingroup members can be beneficial, but simply being from the same ingroup is not enough. We believe that there need to be at least three other criteria before a female role model will have a positive impact on women's math test performance, particularly in situations where a negative stereotype about women's math ability is made salient. First, the female role model's success needs to be perceived as attainable; otherwise, the result may be a negative social comparison that leads to contrast effects (Collins, 1996; Dijksterhuis et al., 1998). Second, the female role model needs to be perceived as similar—similarity in our studies is operationalized as shared physical (e.g., same gender) and psychological (e.g., similar academic interests) attributes between the perceiver and the role model—to the student because that will increase the likelihood that a female student will seek out the female role model as a comparison standard (Lockwood & Kunda, 1997; Tesser, 1986). Third, the student needs to care about her performance in a particular domain. We included domain identification (Steele, 1997), defined as the extent to which students care about their performance in a particular domain, because this factor may moderate the positive effects of female role models.

OVERVIEW OF THE PRESENT STUDIES

In this article, we discuss how female role models protect the math test performance of women, who are highly skilled in math, from the adverse effects of the gender stereotype. Study 1 examined whether, in a seemingly diagnostic testing situation, a female experimenter (i.e., a female role model) buffered women's test performance compared to a male experimenter. For Studies 2 and 3, we manipulated the level of math competence of a fictitious female experimenter by providing detailed biographical information about her math experience and ability. Moreover, in Studies 2 and 3, we investigated whether the physical presence of a female role model is necessary because it may not be her competence per se but other interactional factors (i.e., nonverbal behavior on the part of the female experimenter that provides some type of reassurance) that protect women's test performance. Finally, in Study 3, we were interested in how a female role model influences certain aspects of the self that pertain to academics, namely, self-appraised math ability.

In addition to math test performance, we examined how female role models affected women's performance state self-esteem. Given that interacting with or learning about a female role model buffers women's math test performance, it also may be the case that women's performance state self-esteem will be protected. That is, we expected similar patterns of results for these variables because past research has shown them to be positively correlated (e.g., Brockner & Elkind, 1985; Heatherton & Polivy, 1991).

STUDY 1: WOMEN, MEN, AND MATH TESTS

The primary purpose of Study 1 was to test the link between the presence of female role models (operationalized as a highly math-competent female experimenter) and women's math test performance. Accordingly, we predicted that women would perform as well as men on a difficult math test, even in a situation designed to remind women about the negative gender stereotype, when the experimenter is female, rather than male. We anticipated, however, that women would perform worse than men when a male experimenter gave the math test. Finally, we expected that women and men's performance state self-esteem would be equivalent when a female experimenter, instead of a male experimenter, administered the math test.

Method

PARTICIPANTS AND DESIGN

Participants in this study included 22 women and 21 men who participated in this study. Only those students who were motivated and identified with math as an academic domain were eligible to participate. Identification with math was assessed, during e-mail screening sessions, by asking the students four questions about their interest and ability in math, such as "How important is it for you to do well on math exams?" and "How good are you at math?"[1] The students responded by indicating a number from 1 (*not at all/very bad*) to 5 (*very important/excellent*), and only those who responded to each of the questions with a 3 or greater were eligible to participate. The e-mail message further asked students to report their math SAT score and the number of math classes taken in college. Therefore, in addition to assessing student's identification with math, we used a minimum score (650) from the SAT along with the requirement that each student had taken at least one math course in college to ensure that they had the skills to succeed on the math test. Students' math SAT score also served as the covariate for all of the studies reported here. For this study, we used a 2 (sex of student: female or male) × 2 (sex of experimenter: female or male) between-subjects design.[2]

PROCEDURE

Students reported to the laboratory individually, where they were greeted by one of three female ($n = 21$) or male ($n = 22$) experimenters who were blind to the hypotheses.[3] The experimenters made it clear to the students that they were investigating the math test performance of undergraduates. Moreover, to create an impression of expertise and competence, the experimenters also explained to the students that they would be taking a challenging diagnostic math test created by them and that they would provide feedback about their mathematical strengths and weaknesses at the end of the study session. In short, all of this was done to ensure that the experimenters were perceived as competent and to create a situation similar to that of a typical standardized test administration.

Measuring math test performance. Students had 25 min to complete the math test. The test format resembled a standard Graduate Record Exam (GRE) math section and consisted of 25 problems, which were selected based on the percentage of students from an earlier sample who answered these problems correctly (all questions fell within the range of 17% to 45%; Educational Testing Service, 1994). Students were not provided with performance feedback in any of the studies reported here.

Measuring performance state self-esteem. After taking the math test, performance state self-esteem was measured using the performance subscale from the State Self-Esteem Scale (Heatherton & Polivy, 1991). This subscale measures short-lived changes in how confident people feel about performing well in future situations and has been demonstrated to be sensitive to manipulations that temporarily alter their performance state self-esteem. Respondents rate how they are feeling "right now" for seven performance items, such as "I feel that I have less scholastic ability right now than others." Each item is scored on a 5-point scale ranging from 1 (*not at all*) to 5 (*extremely*), with possible scores ranging from 7 to 35. Finally, all students were debriefed and thanked for their participation.

TABLE 1: Mean Math Test Performance and Performance State Self-Esteem Scores of Female and Male Students as a Function of Sex of Experimenter

	Sex of Experimenter	
Sex of Student	*Female*	*Male*
Math test performance		
Female	14.46_a (1.33)	9.88_b (1.32)
Male	13.56_a (1.43)	15.02_a (1.33)
Performance state self-esteem		
Female	29.19_a (1.36)	23.76_b (1.34)
Male	27.35_a (1.45)	27.40_a (1.35)

NOTE: The means for math test performance and performance state self-esteem are adjusted according to the students' math SAT score. All means that do not share a subscript differ at $p < .05$. Numbers in parentheses are standard errors for the adjusted means.

Results and Discussion

EXCLUDED DATA

Of the 22 women and 21 men who took part in the study, 1 woman and 1 man failed to complete all the measures, and 1 woman in the male experimenter condition scored a zero on the math test. These students' data were not analyzed.[4]

MATH TEST PERFORMANCE

Women and men each received a math test performance score based on the number of problems they answered correctly (see Table 1). We then subjected their math test performance scores, while controlling for math SAT, to a 2 (sex of student) × 2 (sex of experimenter) ANCOVA. This analysis revealed a main effect for the covariate, $F(1, 35) = 14.59$, $p < .01$, and a marginal main effect for sex of student, $F(1, 35) = 3.07$, $p = .09$, $\eta = .28$. As expected, the Sex of Student × Sex of Experimenter interaction was reliable, $F(1, 35) = 4.75$, $p < .04$, $\eta = .35$.

To clarify the nature of this interaction, we tested simple effects by sex of student within sex of experimenter. Results show that women scored just as well as men when a female experimenter administered the math test, $F(1, 35) = 0.30$, $p = .59$, $\eta = .09$, but women reliably underperformed compared to men when the test was given by a male experimenter, $F(1, 35) = 7.72$, $p < .01$, $\eta = .43$.

Having established an effect of the experimenter's sex on women's math test performance, we next examined whether this effect was caused by a decrease in the number of problems answered or to a decrease in performance accuracy. Accordingly, we conducted two separate ANCOVAs on the number of problems answered and the students' performance accuracy (computed by dividing the number of problems the students answered correctly by the number of problems attempted, expressed as a percentage). The ANCOVA on the number of problems answered only revealed a marginal main effect for the covariate, $F(1, 35) = 3.59$, $p < .07$. None of the other effects were reliable, Fs < 2.42, ps $> .13$. In terms of students' performance accuracy, the ANCOVA revealed a main effect for the covariate, $F(1, 35) = 11.08$, $p < .01$, and a reliable Sex of Student × Sex of Experimenter interaction, $F(1, 35) = 5.87$, $p = .02$, $\eta = .38$, such that women who were given the test by a male experimenter had the lowest performance accuracy ($M = 51\%$) compared to the students in the remaining three conditions (e.g., men with a male experimenter = 68%, men with a female experimenter = 62%, and women with a female experimenter = 69%). The pattern of performance differences among the women who were given the math test by a male experimenter therefore appears to be due to a decrease in their performance accuracy rather than to the number of problems answered.

PERFORMANCE STATE SELF-ESTEEM

The students' responses on the performance state self-esteem scale were summed to form a composite score (Cronbach's $\alpha = .87$), and then this score was submitted to a 2 (sex of student) × 2 (sex of experimenter) ANCOVA, with women and men's math SAT as the covariate (see Table 1). The ANCOVA revealed main effects for the covariate, $F(1, 35) = 6.82$, $p < .02$, and sex of experimenter, $F(1, 35) = 4.50$, $p = .04$, $\eta = .34$. The expected interaction was only marginally reliable, $F(1, 35) = 3.78$, $p = .06$, $\eta = .31$. We next tested simple effects by sex of student within sex of experimenter.

Women's performance state self-esteem scores were just as high as men's scores when a female experimenter administered the math test, $F(1, 35) = 0.96$, $p = .35$, $\eta = .16$, which is to be expected considering that women and men performed equally as well. Women, however, had lower performance state self-esteem scores compared to men when a male experimenter gave them the test, $F(1, 35) = 3.74$, $p = .06$, $\eta = .31$, as was anticipated because women underperformed relative to men.

To summarize, having a female experimenter administer a test to highly motivated and math-identified women in a situation designed to activate the gender stereotype about women's inferior math ability protects women's math test performance and allows them to perform at the same level as equally talented men. In contrast, women's math test performance suffered when a male experimenter administered the math test. Of importance, the lowered test results of women who were given the test by a male experimenter were not a result of answering fewer problems. Rather, they were due to a decrease in performance accuracy. In terms of performance state self-esteem, we found the same pattern of results. Women and men had similar levels of performance state self-esteem when a female, rather than a male, experimenter administered the test.

Despite the promising nature of these results, there were some issues that were left unaddressed. First, because we did not manipulate math competence, it remained unclear if the effects we found with our female experimenters were due to the perception of the female experimenters' math competence or their physical presence. This issue seems important in light of research showing that subtle nonverbal cues can profoundly affect people's behavior (Word, Zanna, & Cooper, 1974). To rule out this possibility, we removed the female experimenter's physical presence from Study 2's procedure.

Second, we did not include a role model manipulation check. However, we still contend that our female experimenters were perceived as role models. We say this because our experimenters were self-described as competent in math, were investigating math test performance (an area in which women often deal with negative stereotypes about their math ability), and had conveyed an air of expertise and competence in an area that was relevant and attainable to the students. Indeed, if all of these factors are present, then this can lead one to be viewed as a role model (Javidan et al., 1995; Lockwood & Kunda, 1997), and this may be particularly true when the role model and the students in question are from the same group.

STUDY 2: WOMEN, MEN, AND PERCEPTIONS OF MATH COMPETENCE

In this study, we examined whether female role models need to be physically present to protect women's math test performance. We also tested whether women's math test performance is differentially affected by a female experimenter's level of math competence. Finally, we only included a female experimenter in this study because we were specifically interested in how a female role model affects women's test performance and self-esteem.

According to our thinking about female role models, we expected that women would perform as well as men on the math test when they perceived the female experimenter to be very competent, rather than not very competent, in math. We also anticipated that women and men would have equivalent levels of performance state self-esteem when the female experimenter is described as highly math competent but that women would have lower performance state self-esteem when the female experimenter was not described as highly math competent.

Method

PARTICIPANTS AND DESIGN

In this study, there were 24 women and 22 men who met the same selection criteria from Study 1. We used a 2 (sex of student: female or male) × 2 (level of experimenter math competence: high or low) between-subjects design for this study.

PROCEDURE

Women and men reported to the laboratory individually. On arrival, the students found a note taped to the testing room door explaining that the female experimenter (a fictitious student named Catherine) was late but that they should begin the study anyway. After entering the testing room, a computer screen displayed instructions about how to complete the study packet. This packet was placed next to the computer and contained all the dependent measures except for the math test, which was given on the computer. Toward the end of the study session, a female research assistant informed the students that they were to bring the completed study packet to an adjacent room where they would meet with Catherine and discuss the details of the study.

Manipulating math competence. We manipulated math competence by providing a brief biographical sketch of the female experimenter that the students used as a guide to help them complete their own biographical information. Based on this information, the students were led to believe that the female experimenter was a senior in college and had either a high or low level of competence in math. Students randomly assigned to the high math competence condition ($n = 23$) learned that the female experimenter was majoring in math and psychology, had taken six challenging math classes in college, and planned on earning a Ph.D. in quantitative psychology. Those students assigned to the low math competence condition ($n = 23$) read that the female experimenter was majoring in English and psychology, had taken two moderately challenging math classes in college, and planned on pursuing a career in acting.

Thus, in the high math competence condition, we attempted to create a role model who was similar to the students (a peer in the form of a talented college senior) whose success seemed attainable (she had taken challenging math classes at the same college as the students) as well as relevant (recall that we only selected students who were highly motivated and identified with math as an academic domain). In contrast, the female experimenter in the low math competence condition simply appeared as one of the students' peers.

Measuring math test performance and performance state self-esteem. Following the math competence manipulation, students had 25 min to complete the same math test used in Study 1. The students timed themselves using a countdown timer that was preset for 25 min. Preliminary testing revealed that no student had difficulty with the timing procedure and all stopped when 25 min had elapsed. After the math test, students completed the same performance state self-esteem scale from Study 1.

Manipulation check. To test for the effectiveness of the math competence manipulation, students were presented with two items: "The experimenter's score on this test would be higher than that of the average student's score" and "The experimenter's math SAT score would be higher than that of the average student's score," to which they responded on a 7-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*). After completing these items, the students were debriefed and thanked for their participation.

Results and Discussion

EXCLUDED DATA

Of the 46 students, 4 women and 2 men did not complete all the measures; hence, they were excluded from the analyses.

MANIPULATION CHECK

The students' perceptions of the female experimenter's math competence were summed and then submitted to a 2 (sex of student) × 2 (level of experimenter math competence) ANCOVA. This analysis revealed a main effect for sex of student, $F(1, 36) = 4.12$, $p = .05$, η = .32 (M_{men} = 9.55, M_{women} = 8.40), and level of experimenter math competence, $F(1, 36) = 11.85$, $p < .01$, η = .50, such that students in the high math competence condition (M = 9.95) rated the female experimenter as more competent than did students in the low math competence condition (M = 8.00), indicating strong support for the effectiveness of this manipulation.

MATH TEST PERFORMANCE

As in Study 1, the students' math test performance scores (adjusted by their math SAT) were subjected to a 2 (sex of student) × 2 (level of experimenter math competence) ANCOVA (see Table 2). The analysis revealed main effects for the covariate, $F(1, 35) = 15.71$, $p < .01$, and sex of student, $F(1, 35) = 10.68$, $p < .01$, η =.48. The Sex of Student × Level of Experimenter Math Competence interaction also was reliable, $F(1, 35) = 7.92$, $p < .01$, η = .40. To interpret this interaction, we tested simple effects.

Women in the high math competence condition scored better on the math test than did women in the low math competence condition, $F(1, 35) = 4.20$, $p < .05$, η = .33, but men had lower test scores when they perceived

TABLE 2: Mean Math Test Performance and Performance State Self-Esteem Scores of Female and Male Students as a Function of Level of Experimenter Math Competence

	Level of Experimenter Math Competence	
Sex of Student	*High*	*Low*
Math test performance		
Female	12.18_a (1.04)	9.23_b (1.12)
Male	12.79_a (1.02)	15.90_c (1.03)
Performance state self-esteem		
Female	24.14_b (1.30)	28.74_a (1.40)
Male	28.52_a (1.28)	28.50_a (1.29)

NOTE: The means for math test performance and performance state self-esteem are adjusted according to the students' math SAT score. All means that do not share a subscript differ at $p < .05$. Numbers in parentheses are standard errors for the adjusted means.

the female experimenter to be competent rather than not very competent in math, $F(1, 35) = 4.66$, $p < .05$, η = .34. Our reasoning for this result is that men showed a contrast effect when they were exposed to an exemplar of math competence, who also disconfirmed the negative gender stereotype about women and math (Dijksterhuis et al., 1998).[5] We next examined whether women performed worse than men in the low math competence condition. The results supported our hypothesis, $F(1, 35) = 21.45$, $p < .01$, η = .62.

Again, we investigated if this performance difference occurred for women because they answered fewer problems or because of a decrease in their performance accuracy. The ANCOVA on the number of problems answered only revealed a marginally reliable Sex of Student × Level of Experimenter Math Competence interaction, $F(1, 35) = 3.38$, $p = .07$, η = .30, showing that men in the low math competence condition (M = 23.43) answered slightly more problems then did students in the other three conditions (e.g., men in the high math competence condition = 20.13, women in the high math competence condition = 21.36, and women in the low math competence condition = 20.17). The ANCOVA on the students' performance accuracy showed a main effect for the covariate, $F(1, 35) = 16.68$, $p < .01$, and sex of student, $F(1, 35) = 8.49$, $p < .01$, η = .39, such that men (M = 66%) were more accurate than were women (M = 53%). In addition, the Sex of Student × Level of Experimenter Math Competence interaction was reliable, $F(1, 35) = 7.02$, $p = .01$, η = .41, indicating that men in the low math competence condition had the highest performance accuracy (M = 69%), whereas women in this condition had the lowest (M = 46%). There was no reliable difference between women and men's performance accuracy in the high math competence condition (M_{men} = 63%, M_{women} = 59%). Thus, in this study, the pattern of performance differences in the low math competence condition appears to be a result of both an increase in the number of problems answered by men and a decrease in performance accuracy among women.

PERFORMANCE STATE SELF-ESTEEM

The students' performance state self-esteem scores were summed and submitted to a 2 (sex of student) × 2 (level of experimenter math competence) ANCOVA, with math SAT as the covariate (see Table 2). This analysis revealed a marginal main effect for sex of student, $F(1, 35) = 2.93$, $p < .10$, $\eta = .28$, and a marginally reliable Sex of Student × Level of Experimenter Math Competence interaction, $F(1, 35) = 2.91$, $p < .10$, $\eta = .28$, indicating that women in the high math competence condition had lower performance state self-esteem compared to the students in the remaining three conditions. This unexpected result seems to be due to the fact that after controlling for math SAT and level of experimenter math competence, the correlation between women's math test performance and performance state self-esteem was not significant, $r = .17$, $p = .26$ (one-tailed). Women's performance state self-esteem therefore seems not to be strongly related to their test performance. Moreover, because performance state self-esteem is contingent on actual performance, these highly motivated and math-identified women may have felt frustrated about not having performed as well as they believed they could have performed. It is also possible that when women compared their math test performance to how they believed the female experimenter could have performed, this ingroup upward social comparison harmed their performance state self-esteem (see Major et al., 1993; Tesser, 1986, for related discussions).

The results from this study are somewhat mixed. On one hand, a female experimenter who is described as competent in math buffers women's math test performance but does not protect their performance state self-esteem. Men, on the other hand, showed the opposite pattern of performance results, presumably because they were threatened by the female experimenter's math competence. In other words, these men may have felt extra pressure to live up to the positive stereotype about their math ability in the high math competence condition, whereas in the low math competence condition, they were presumably freed from this burden of proof (Brown & Josephs, 1999). Finally, men's performance state self-esteem did not seem to be affected by the female experimenter's level of math competence.

But why would a female role model protect women's math test performance? We argue that in situations thought to affect women's self-appraised ability, female role models may actually remind women, who are already extremely motivated and identified with math, of their own high ability in that domain. Indeed, assimilation effects can occur for women when the comparison target is an ingroup member (e.g., a female role model) who represents stereotype-disconfirming evidence (e.g. high math competence) on a task (e.g., math) that is esteem-relevant (Blanton et al., 2000; Brewer & Weber, 1994).

Support for this notion can be found in a series of studies by Lockwood and Kunda (1997) in which they demonstrated that role models enhance the way students see their academic abilities, especially when the role model's success, in the form of a talented graduating college senior, seemed attainable and relevant to the students. Furthermore, students are more likely to benefit in terms of their self-appraisals when they believe that the role model is similar to them (e.g., from the same gender group) and possesses a desirable characteristic, such as talent in a particular academic domain (Collins, 1996; Pelham & Wachsmuth, 1995). Consistent with this, Schmader and Major (1999) found that if female college students learn about another female students' success on a task, then this will have a profoundly positive effect on the way these female students view their own ability on that same task.

STUDY 3: WOMEN, MATH TESTS, AND SELF-APPRAISED MATH ABILITY

In Study 3, we tested if female role models, even in seemingly diagnostic situations, protect women's self-appraised math ability, which could then lead to successful math test performance. For this study, we only used a female experimenter and female students because we were specifically interested in how female role models affect women's self-evaluations.

We expected that women would perform better when they perceived the female experimenter to be very competent in math. We also anticipated that women's self-appraised math ability would be buffered in the high instead of the low math competence condition. Moreover, we hypothesized that self-appraised math ability would partially mediate the relationship between the female experimenter's level of math competence and the women's math test performance. Finally, because the performance state self-esteem results from Study 2 were unexpected, we conducted the same analyses on the students' performance state self-esteem scores to investigate whether the results from Study 2 would be repeated.

Method

PARTICIPANTS AND DESIGN

Participants included 44 women who met the same selection criteria from Studies 1 and 2. This study used a one-way ANCOVA design with level of experimenter math competence as the between-subjects variable.

PROCEDURE

The procedure was identical to that described in Study 2 except for the addition of the self-appraised math ability measure and the inclusion of another item to assess the women's perceptions of the female experimenter's math competence (i.e., "Do you think the experimenter could be successful in math?"). After learning that the female experimenter was either very competent (n = 23) or not very competent (n = 21) in math, the women were asked to respond to four items about their math ability (Cronbach's α = .84), each on a scale from 1 (*not at all*) to 5 (*extremely true*) (viz. "I deal poorly with challenges in math"; "In the future, I feel that I will not succeed at math"; "Relative to other women, I perform inadequately in many important math situations"; and "In the future, I will not have much to be proud of in math"). Following this, women took a 15-problem math test that consisted of a random subset of problems from the math test used in Studies 1 and 2. Due to time constraints, we shortened the math test so that the study lasted about 45 min instead of an hour. After the math test, women completed the performance state self-esteem scale and were then debriefed and thanked for their participation.

Results and Discussion

EXCLUDED DATA

Of the 44 women, 4 failed to complete all the measures; hence, they were excluded from the analyses.

MANIPULATION CHECK

The women's perceptions of the female experimenter's math competence were submitted to a one-way ANCOVA with the level of experimenter math competence as the between-subjects variable. This analysis revealed a main effect for level of experimenter math competence, such that women in the high math competence condition (M = 17.95) rated the female experimenter as reliably more competent in math than did women in the low math competence (M = 13.73) condition, $F(1, 37) = 54.13$, $p < .01$, $\eta = .77$.

MATH TEST PERFORMANCE

Women's math test performance scores were subjected, while controlling for math SAT, to a one-way ANCOVA with level of experimenter math competence as the between-subjects variable (see Table 3). This analysis revealed main effects for the covariate, $F(1, 37) = 14.60$, $p < .01$, and level of experimenter math competence, $F(1, 37) = 11.34$, $p < .01$, $\eta = .48$, indicating that women in the high math competence condition performed reliably better on the math test than did those in the low math competence condition.

TABLE 3: Mean Math Test Performance, Self-Appraised Math Ability, and Performance State Self-Esteem of Women as a Function of Level of Experimenter Math Competence

	Level of Experimenter Math Competence	
Measure	*High*	*Low*
Math test performance	8.02 (0.50)	5.63 (0.50)
Self-appraised math ability	18.61 (0.56)	16.82 (0.56)
Performance state self-esteem	27.17 (1.02)	22.85 (1.02)

NOTE: The means for math test performance, self-appraised math ability, and performance state self-esteem are adjusted according to the women's math SAT score. All means differ at $p < .05$. Numbers in parentheses are standard errors for all the adjusted means.

We next examined whether this effect was a result of a difference in the number of problems answered or a decrement in performance accuracy. None of the effects from the one-way ANCOVA on the number of problems answered were reliable, Fs < 0.34, ps > .57. In contrast, the results from the one-way ANCOVA on the women's performance accuracy revealed a main effect for the covariate, $F(1, 37) = 17.29$, $p < .01$, and level of experimenter math competence, $F(1, 37) = 12.60$, $p < .01$, η = .50, revealing that women in the high math competence condition were more accurate (M = 63%) than were women in the low math competence condition (M = 48%). The performance results from this study therefore appear to be due to a decrease in women's performance accuracy in the low math competence condition and not to a difference in the number of problems answered.

SELF-APPRAISED MATH ABILITY

We added the women's responses from the self-appraised math ability measure to form a single self-appraised math ability score. Next, we submitted their scores to a one-way ANCOVA with level of experimenter math competence as the between-subjects variable and math SAT as the covariate (see Table 3). The ANCOVA revealed a marginal main effect for the covariate, $F(1, 37) = 3.90$, $p < .06$, and a main effect for level of experimenter math competence, $F(1, 37) = 5.03$, $p = .03$, $\eta = .35$, showing that women in the high math competence condition had significantly higher self-appraised math ability scores than did women in the low math competence condition. This result suggests that learning about a female experimenter who is competent in math buffers women's self-appraised math ability, thereby leading to better math test performance. In fact, the correlations (controlling for math SAT and level of experimenter math competence) between women's math test performance and self-appraised math ability, $r = .28$, $p < .05$ (one-tailed), and women's perceptions of the female experimenter's math competence and self-appraised math ability, $r = .31$, $p < .03$ (one-tailed) also lend support to this notion. Thus, it appears that women are assimilating the female experimenter's positive characteristics into their self-concept. This, in turn, protects them from the negative effects of the gender stereotype and as a result leads them to perform better in a situation where women have traditionally underperformed.

To test our hypothesis that the positive effects of female role models on women's math test performance could, in part, be accounted for by women's self-appraised math ability, we followed Baron and Kenny's (1986) guidelines for conducting mediational analyses.[6] We first found that the relationship between the female experimenter's level of math competence and women's math test performance was reliable, $\beta = .43$, $p < .002$, $R^2 = .23$. This relationship represents the effect to be mediated. However, when we tested this relationship again, after controlling for women's self-appraised math ability, it remained reliable, $\beta = .35$, $p = .01$, $R^2 = .16$; hence, women's self-appraised math ability was not a significant mediator of this relationship. It should be stressed, though, that women's self-appraised math ability was reliably higher in the high math competence condition compared to the low math competence condition.[7]

PERFORMANCE STATE SELF-ESTEEM

We subjected the women's performance state self-esteem scores to a one-way ANCOVA with level of experimenter math competence as the between-subjects variable and math SAT as the covariate (see Table 3). The ANCOVA only yielded a condition main effect, $F(1, 37) = 8.71$, $p < .01$, $\eta = .44$, indicating that women in the high math competence condition had higher performance state self-esteem relative to the women in the low math competence condition. Here, unlike in Study 2, but consistent with Study 1, we found that women's performance state self-esteem did reliably covary with their math test performance, presumably because these women may have performed at the level that they had expected to perform.

In short, we examined whether learning about a female experimenter who is described as competent in math would protect women's self-appraised math ability. The results suggest that women's self-appraised math ability is protected by a female role model; yet, counter to our predictions, self-appraised math ability was not a reliable mediator. As before, we found that women in the high math competence condition performed better on the math test than did women in the low math competence condition. Finally, consistent with related research on performance state self-esteem (e.g., Blanton et al., 2000), women had higher performance state self-esteem scores when they perceived the female experimenter to be highly competent, instead of not very competent, in math.

GENERAL DISCUSSION

According to recent National Science Foundation (1996) statistics, women constitute only 35% of computer science, math, and physics majors, and less than 10% of women enter graduate programs in these areas. Why are there so few female students in math-related domains? Despite the varied explanations for this state of affairs, there is one that remains consistent: Women's performance on standardized math tests is slightly lower than men's (e.g., Gallagher et al., 2000) regardless of there being few other differences among women and men in math achievement and classroom performance (Hyde et al., 1990; Kimball, 1989). Perhaps this test score difference is one of the roadblocks that is turning away highly motivated and math-identified women from degree programs that would allow them to become mathematicians and/or engineers.

But the situation seems encouraging considering the positive impact of female role models for the women in our studies, who were motivated and identified with math. If more female role models were present in math-related domains, would this then eliminate the gender gap on standardized math tests? The results from the three studies presented here suggest that this may be the case. The first study showed that highly motivated and math-identified women performed better on a math test, even in a situation that has traditionally undermined their test performance, when a female experimenter (i.e., a female role model) instead of a male experimenter gave the test. Study 2 indicated that the perceived competence of a female experimenter, and not her physical presence, protected women's math test performance. Finally, Study 3 revealed that the perceived competence of a female experimenter buffered women's self-appraised math ability so that they performed better on a difficult math test. But, in contrast to our hypothesis, women's self-appraised math ability did not serve as a reliable mediator of the relationship between the female experimenter's level of math competence and women's math test performance.

One reason for this result may be that protecting women's self-appraised ability from the negative effects of a gender stereotype is not the only positive effect of a female role model on women's self-evaluations; rather, there may be a variety of additional positive effects, such as increased inspiration (Lockwood & Kunda, 1997; Tesser, 1986) and enhanced task motivation (e.g., Collins, 1996). But before a complete list of positive effects can be generated, we feel it is important to consider that the level of identification with one's group and the domain in which a female role model excelled are potentially significant factors that could alter the positive effects of female role models. In other words, depending on a female student's level of identification with the female role model and the domain in which the female role model has excelled, very different effects could occur. Clearly, the strength of group identification and type of domain need to be explored further before any concrete claims about the positive effects of female role models can be made, particularly because it may be the case that if women are not strongly identified with their gender group and/or the domain in question, they may not derive the same benefits from female role models (see Schmader, in press, for a related discussion about gender identification and stereotype threat).

Another issue that deserves mention is the fact that we did not find a consistent result for women's performance state self-esteem. Indeed, we found that in Study 2, even though women performed better after learning about a female role model, they had lower performance state self-esteem. Our interpretation of this result is that these highly math-skilled women may have felt as if they did not perform as well as they thought both they and the female role model presumably could. Thus, this ingroup upward social comparison led to contrast effects in terms of their performance state self-esteem. This idea is also similar to Tesser's (1986) work on self-evaluation maintenance, such that unfavorable comparisons on esteem-relevant tasks can harm a person's self-evaluations. Taken together, however, the results from these studies show that if women are able to perform up to their potential, then their performance state self-esteem is protected. Evidence for this can be found by the fact that when we combined the performance state self-esteem results across the three studies (after assigning a negative weight to Study 2's results), we still found that the combined p value supported our hypothesis: Female role models protect women's performance state self-esteem, $Z = 1.78$, $p < .04$ (one-tailed).

Finally, it should be stressed that we are not making the argument that all female experimenters are threatening to men or that all male experimenters are threatening to women. Rather, only in particular situations do we feel that this would be the case. For instance, we believe that because our female experimenters were peer role models instead of more abstract role models (i.e., older and perhaps less similar role models to the students), this is why men may have felt threatened by our competent female experimenter. Moreover, recent research shows that if the female role model is described as a highly competent math tutor, rather than a peer role model, both women and men benefit as long as she is still perceived to be similar (in terms of shared interests) to the students (Marx, Urland, & Overbeck, 2002).

IMPLICATIONS BEYOND THE LABORATORY

Extrapolating findings from the laboratory to the real world must always be done with caution. The research presented here could, nevertheless, have far-reaching implications for educational policy and practice. Specifically, this work may demonstrate the importance of female role models for female college students who are motivated and identified with math as an academic domain, particularly in situations where negative gender stereotypes apply, such as in math and engineering (Crocker, Major, & Steele, 1998; Halpern, 1992; Quinn & Spencer, 2001). Of importance, related research suggests that women are less likely to become threatened by a stereotype indicating their mathematical inferiority when they outnumber men rather than when they are outnumbered by men (Inzlicht & Ben-Zeev, 2000). Could this effect be enhanced by the presence of female role models? It certainly seems as if it could, in light of the promising effects that we found with a single female role model. In the end, increasing the number of female role models in math and engineering classes may allow female students to view the negative gender stereotypes that confront them as surmountable barriers rather than ones that are insurmountable and therefore potentially inspire more women, who may not be initially identified with math, to pursue careers in these academic areas.

NOTES

1. We obtained students' e-mail addresses from recruitment flyers posted around the psychology department. The e-mail message simply described the nature of the study and then asked the students to indicate their responses to the identification questions and provide their math SAT score and the number of math classes taken in college. Even though all students had already agreed to participate by writing their e-mail address down on the recruitment flyer, only those who met the selection criteria actually did.

2. We had originally included a stereotype threat manipulation, but due to the fact that our manipulation check did not yield a reliable difference between the high and low stereotype threat conditions (for the low stereotype threat condition, students simply read that the test was challenging with no mention of gender or diagnosticity of the test), we dropped the low stereotype threat condition and only present results from the high stereotype threat condition. In this way, we maintain consistency across the three studies. Furthermore, because high stereotype threat situations have been shown to lower women's math test performance (e.g., Spencer, Steele, & Quinn, 1999), we were most interested in how female role models protect women's math test performance in this situation in particular.

3. No differences emerged on any of the dependent measures among the three female experimenters, Fs < 1.27, ps > .29, with the same being true among the three male experimenters, Fs < 1.41, ps > .26. Thus, the results of Study 1 surfaced not because of any unique characteristics the experimenters may have had.

4. To have the same number of students in each condition, we continued to run students until we had 10 (Studies 1 and 2) and 20 (Study 3) in each. For all studies reported here, we excluded students if they did not complete all the measures or scored a zero on the math test because it was quite clear that they did not take the testing situation seriously.

5. Further support for this notion comes from a study in which men and women were asked to make predictions about their math test per-

formance after they saw pictures of either math competent or neutral female students who were also the students' peers (Marx, Urland, & Overbeck, 2002). Results show that college men had lower performance predictions when they saw pictures of math competent female students and higher performance predictions when they saw pictures of neutral female students, whereas college women showed the opposite pattern of performance predictions, $F(1, 60) = 5.91$, $p < .02$, $\eta = .30$.

6. All mediational analyses were conducted after adjusting for the women's SAT scores.

7. The female experimenter's level of math competence predicted women's self-appraised math ability, $\beta = .34$, $p = .03$, $R^2 = .12$, and women's self-appraised math ability predicted their math test performance, $\beta = .37$, $p < .01$, $R^2 = .16$.

REFERENCES

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Blanton, H., Crocker, J., & Miller, D. T. (2000). The effects of in-group versus out-group social comparison on self-esteem in the context of a negative stereotype. *Journal of Experimental Social Psychology*, *36*, 519-530.

Boswell, S. J. (1985). The influence of sex-role stereotyping on women's attitudes and achievement in mathematics. In S. F. Chipman, L. R. Brush, & D. M. Wilson (Eds.), *Women and mathematics: Balancing the equation* (pp. 175-197). Hillsdale, NJ: Lawrence Erlbaum.

Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology*, *66*, 268-275.

Bridgeman, B., & Wendler, C. (1991). Gender differences in predictors of college mathematics performance. *Journal of Educational Psychology*, *83*, 275-284.

Brockner, J., & Elkind, M. (1985). Self-esteem and reactance: Further evidence of attitudinal and motivational consequences. *Journal of Experimental Social Psychology*, *21*, 346-361.

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.

Collins, R. L. (1996). For better or worse: The impact of upward social comparisons on self-evaluation. *Psychological Bulletin*, *119*, 51-69.

Crocker, J., Major, B., & Steele, C. M. (1998). Social Stigma. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 504-553). New York: McGraw-Hill.

Dijksterhuis, A., Spears, R., Postmes, T., Stapel, D., Koomen, W., Knippenberg, A., et al. (1998). Seeing one thing and doing another: Contrast effects in automatic behavior. *Journal of Personality and Social Psychology*, *75*, 862-871.

Educational Testing Service. (1994). *GRE: Practicing to take the general test* (9th ed.). Princeton, NJ: Author.

Fisher, B. (1988). Wandering in the wilderness: The search for women role models. *Signs: Journal of Women in Culture and Society*, *13*, 211-233.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, *75*, 165-190.

Gilbert, L. A., Gallessich, J. M., & Evans, S. L. (1983). Sex of faculty role model and students' self-perceptions of competency. *Sex Roles*, *9*, 597-607.

Graduate Record Examinations Board. (1997). *Sex, race, ethnicity and performance on the GRE General Test* (Tech. Rep.). Princeton, NJ: Educational Testing Service.

Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of Counseling Psychology*, *32*, 47-56.

Halpern, D. F. (1992). *Sex differences in cognitive abilities* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale measuring state self-esteem. *Journal of Personality and Social Psychology*, *60*, 895-910.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139-155.

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

Javidan, M., Bemmels, B., Devine, K. S., & Dastmalchian, A. (1995). Superior and subordinate gender and the acceptance of superiors as role models. *Human Relations*, *48*, 1271-1284.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychology Bulletin*, *105*, 198-214.

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, *73*, 91-103.

Major, B., Sciacchitano, A. M., & Crocker, J. (1993). In-group vs. out-group comparisons and self-esteem. *Personality and Social Psychology Bulletin*, *19*, 711-721.

Marx, D. M., Brown, J. L., & Steele, C. M. (1999). Allport's legacy and the situational press of stereotypes [Special issue: Prejudice and intergroup relations: Papers in honor of Gordon W. Allport's centennial]. *Journal of Social Issues*, *55*(3), 491-502.

Marx, D. M., Urland, G. R., & Overbeck, J. R. (2002). *Seeing is believing: Protecting students' test performance through images of success.* Manuscript in preparation.

National Science Foundation. (1996). *Women, minorities, and persons with disabilities in science and engineering: 1996* (Publication No. 96-311). Arlington, VA: Author.

Pelham, B. W., & Wachsmuth, J. O. (1995). The waxing and waning of the social self: Assimilation and contrast in social comparison. *Journal of Personality and Social Psychology*, *69*, 825-838.

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies [Special issue: Stigma: An insider's perspective]. *Journal of Social Issues*, *57*(1), 55-71.

Schmader, T. (in press). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology.*

Schmader, T., & Major, B. (1999). The impact of ingroup vs. outgroup performance on personal values. *Journal of Experimental Social Psychology*, *35*, 47-67.

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.

Taylor, S. E., & Lobel, M. (1989). Social comparison activity under threat: Downward evaluation and upward contacts. *Psychological Review*, *96*, 569-575.

Tesser, A. (1986). Some effects of self-evaluation maintenance on cognition and action. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 435-464). New York: Guilford.

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in inter-racial interactions. *Journal of Experimental Social Psychology*, *10*, 109-120.

Received September 17, 2001
Revision accepted January 24, 2002