Journal of Social Issues, Vol. 57, No. 1, 2001, pp. 55–71

The Interference of Stereotype Threat With Women's Generation of Mathematical Problem-Solving Strategies

Diane M. Quinn*
University of Connecticut

Steven J. Spencer
University of Waterloo

At the highest levels of math achievement, gender differences in favor of men persist on standardized math tests. We hypothesize that stereotype threat depresses women's math performance through interfering with their ability to formulate problem-solving strategies. In Study 1, women underperformed in comparison to men on a word problem test, however, women and men performed equally when the word problems were converted into their numerical equivalents. In Study 2, men and women worked on difficult problems, either in a high- or reduced-stereotype-threat condition. Problem-solving strategies were coded. When stereotype threat was high, women were less able to formulate problem-solving strategies than when stereotype threat was reduced. The effect of stereotype threat on cognitive resources and the implications for gender differences in mathematical testing are discussed.

> *"I have no idea how to do this one, it just doesn't seem like there is enough information. I have no idea how to set up a formula for this. I really don't know what to do with that one."*

---

*We would like to thank Christina Brearley, Nancy Faulk, Marissa Przybylo, and Sabrina Voelpel for their help in data collection and coding.

Correspondence concerning this article should be addressed to Diane Quinn, Department of Psychology, University of Connecticut, 406 Babbidge Road, U-1020, Storrs, CT, 06269 [e-mail: diane.quinn@uconn.edu].

> *"Hmm, I'm thinking that I don't know where to start . . . I don't know, I don't see, I don't understand any of them, so I'll skip it."*

The above quotes illustrate how difficult and frustrating trying to solve mathematical problems can be. What makes these quotes particularly disheartening is that they all come from college women with strong mathematical backgrounds who are trying to solve high-level Scholastic Assessment Test (SAT) problems. Why are these women having a difficult time? What is interfering with their abilities to formulate problem-solving strategies? We propose that these women are in a situation in which cultural stereotypes about their math abilities are depressing their performance and interfering with their problem solving.

Stereotypes about academic skills are well known in our culture (Eagly & Wood, 1991; Eccles, Jacobs, & Harold, 1990; Levine & Ornstein, 1983; Maccoby & Jacklin, 1974; Swim, 1994). According to these stereotypes, men/boys are better at math and science domains, whereas women/girls are better at English and reading domains. These stereotypes are transmitted in the culture in a variety of ways, including through mass media, books, parents, peers, and teachers. They may affect children through the toys they play with, the books they choose to read, the way they are treated in class, and, eventually, the classes they choose to take and the careers they pursue (Biernat, 1991; Constantinople, Cornelius, & Gray, 1988; Eccles, 1987; Hewitt & Seymour, 1991; Leinhardt, Seewald, & Engel, 1979; Martin, Wood, & Little, 1990; Meece, Eccles, Kaczala, Goff, & Futterman, 1982).

One of the most basic ways in which stereotypes about women's math abilities are promulgated is through parents' and teachers' expectations. Jacobs and Eccles (1992) found that the stereotypic beliefs of mothers tended to color the mothers' perceptions of their daughters' and sons' math abilities. In addition, work by Frome and Eccles (1998) showed that mothers tended to underestimate the mathematical abilities of their sixth-grade daughters and overestimate the math abilities of their sons. In a study of high school students, Hyde, Fennema, Ryan, Frost, and Hopp (1990) found that high school boys reported greater favorable attitudes from their mothers, fathers, and teachers about their math ability than high school girls. These stereotypical expectancies and beliefs from parents, teachers, and the culture certainly have effects on girls' attitudes toward math. Jacobs and Eccles (1992) found that mothers' perceptions of their children's abilities have a greater influence on children's perception of their own abilities than do past grades. In addition, mothers' low expectations for their daughters are related to girls' having lower math performance expectations, which in turn leads to intentions to take fewer math courses (Eccles & Jacobs, 1986).

All of this research highlights the many powerful ways that stereotypes about gender and math are transmitted in the culture. What is the effect of these stereotypes on girls' and women's actual math performance and achievement? Examination of gender differences in math achievement largely contradicts the stereotypes.

In the elementary school years, girls are performing better than boys on tests of computation and equally with boys on word problem tests (Hyde, Fennema, & Lamon, 1990). Again, at middle school age, girls are performing higher on computation tests and equally on word problem tests (Armstrong, 1981; Hyde et al., 1990). Girls are also getting higher grades in math classes than boys (Frome & Eccles, 1998; Kimball, 1989). Starting approximately at junior high, however, and continuing more strongly in high school, the pattern of achievement reverses. Boys and girls perform equally on computational tests, but boys outperform women on math word problem tests. This gap on math word problems continues and widens into college and adulthood (Armstrong, 1981; Hyde et al., 1990). Although recent work that has looked at national samples has found the achievement gap between high school men and women has narrowed over the past 20 years (Cole, 1997; Feingold, 1988), significant gaps are still quite evident when the highest achievers are examined. Indeed, on standardized tests such as the SAT and GMAT (Graduate Management Admissions Test), it is at the highest level, within the group of high scorers, that women are underrepresented in comparison to men, even though they may have equal math backgrounds and class grades (Cole, 1997; Stockard & Wood, 1984). In addition, women are still underrepresented in college majors and professional careers that stress mathematical skills, such as engineering and computer science (Ramist, Lewis, & McCamley-Jenkins, 1994; Strenta, Elliot, Adair, Scott, & Matier, 1993).

Why are women who have achieved the most, who have the strongest math skills, underperforming in comparison to their male peers? Why is this gender gap more likely to show up in standardized testing than in class grades? We believe the answers to these questions lie in examining the interaction between cultural stereotypes and the test-taking situation, what we call a “stereotype threat” situation. Stereotype threat occurs when a person is in a situation in which a negative stereotype about that person (or that person’s group) could be applied to the person and used to judge the person’s behavior (Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995; Steele, 1997a). In the case of gender and math, imagine a boy and girl sitting down to take the SAT for the first time. Both are straight-A students. Both care about education and want to go to the best colleges possible. Taking the SAT is a tense, sometimes frustrating experience for both of them. However, as the girl is taking the test she has an extra worry to contend with that the boy does not: A stereotype that she, as a girl, has inferior math skills. As she experiences frustration and difficulty with the problems, she has the burden of knowing that her difficulty could be judged as proof of the veracity of the stereotype. The boy has none of these doubts or thoughts to interrupt his performance.

We have tested the stereotype threat hypothesis in a series of studies (Spencer et al., 1999). In the first of these studies we brought men and women college students with equivalent math backgrounds into the laboratory to take either an easy or a difficult math test. We found that women underperformed only on the difficult test. To demonstrate that it was the threat of the stereotype that caused this underperformance, we (Spencer et al., 1999, Study 2) gave men and women a second difficult math test. In order to make stereotypes about math explicit, half of the participants were told that the test had shown gender differences in the past. In order to eliminate a stereotype-based interpretation of the situation, the other half of the participants were told that the test had been shown to be gender fair—that men and women performed equally on this test. In line with our predictions, when the stereotype was not said to be applicable to the situation, men and women performed equally on the test. However, when stereotype threat was high, when men and women were told that the test had shown gender differences in the past, women underperformed in comparison to the men. Thus it seems that being in a situation in which the stereotype could be applied was suppressing women's scores, whereas when that stereotype was removed, their performance improved. In Study 1 and replicated in Study 3, it is notable that women perform worse than men when there is *no mention* of stereotypes or gender. Hence, we believe that on difficult math tests, the "normal" or standard state of affairs is for the situation to be high in stereotype threat.

The results of this research are consistent with findings about how stereotype threat affects African Americans' test performance. Steele and Aronson (1995) found that making the stereotype inapplicable to a particular test by taking away the threat of judgment increased African Americans' test performance. (The threat of judgment was manipulated by telling participants that the test was either diagnostic or nondiagnostic of verbal ability.) African American students who had underperformed in comparison to European American students when their performance could be judged performed equally when stereotype threat was reduced. In addition, priming race led to a decline in performance by African Americans (Steele & Aronson, 1995). Crozet and Claire (1998) found that when participants in France were reminded of their socioeconomic status (SES), low-SES students performed worse than high-SES students. Likewise, Shih, Pittinsky, and Ambady (1999) found that for Asian American women, when their Asian identity was primed, they performed better on a math test than a control group, whereas when their female identity was primed, they performed worse than the control group. These studies demonstrate that identities, stereotypes, and situations can interact in many powerful ways to reduce or increase stereotype threat, which consequently has a strong effect upon performance.

The research conducted on stereotype threat has revealed several important points about when, or under what conditions, stereotype threat depresses performance. First, the test must be difficult: at the very edge of the students' abilities. If the test is easy, and the test taker experiences little difficulty with it, it is unlikely that stereotype threat will interfere with performance (Spencer et al., 1999). If the test is so difficult that the test taker knows it is out of the bounds of his or her skills, stereotype threat is also unlikely to show up in performance differences. Second, students must care about their performance on the test. It may be that for students who have already disengaged from the academic subject of the test (Steele, 1997b; Steele et al., 1998), changing the stereotype threat in the situation will not affect their performance. Considering women and math particularly, the women must still care about math and believe they have the skills to do well on the test for stereotype threat to affect their performance. Unfortunately, many women disassociate themselves from math at an early age, and for these women, a change in the stereotype threat situation will probably not affect their score. Thus, the cultural stereotype about women's inferior math ability works in many ways: It may account for girls' greater anxiety about their math skills (Wigfield & Meece, 1988), it may sway girls toward taking fewer math classes, and fewer girls may value math as an important personal ability. Even for those girls who do care about math, however, and do continue to take math courses and achieve, the cultural stereotype can affect their performance by conjuring up doubts about their ability and by applying additional pressures during difficult testing situations. Thus, it is those girls and women who are the very best at math that may be most affected by stereotype threat while taking a difficult math test.

Research with women in math and African Americans in the verbal domain (Aronson, Quinn, & Spencer, 1998; Steele, 1997a) has shown that stereotype threat can depress test performance. Questions remain, however, about the process by which stereotype threat affects performance. Several studies (Spencer et al., 1999; Steele & Aronson, 1995) have examined cognitive and emotional mediators of stereotype threat. These studies have shown that, when people experience stereotype threat, they tend to have higher levels of anxiety, they try to avoid being characterized in a stereotypic ways, and yet (or perhaps because of their desire to avoid being stereotyped) they show increased cognitive activation of the stereotype. Together these studies suggest that when people experience stereotype threat they experience affective arousal, and their cognitive resources are constrained.

What we have not yet been able to elucidate is how stereotype threat affects women's problem solving on math tests. That is, how does the heightened anxiety and diminished cognitive capacity that accompanies stereotype threat impair women's ability to solve math problems? This is the question that we address in the current studies. We hypothesize that stereotype threat interferes with problem solving by impairing women's ability to formulate strategies. We believe that women have the conceptual or computational skills necessary to solve difficult math problems, but the additional anxiety and diminished cognitive capacity associated with stereotype threat interferes with their ability to strategize, a process that takes focused concentration and attentional resources. We examine this hypothesis in two ways. In Study 1, we examine whether removing the requirement for strategizing improves women's math performance when stereotype threat is high. In Study 2, we examine whether reducing stereotype threat improves women's ability to use problem-solving strategies.

Study 1

As reviewed above, research examining the gender gap in math achievement has shown that one of the first places that gender differences appear is on tests of word problems or conceptual applications (Hyde et al., 1990; Lummis & Stevenson, 1990). From a stereotype threat perspective, this is not surprising. Word problems involve many steps (Bernardo & Okagaki, 1994) that can open up room for doubt and frustration. When doubt and frustration occur, the stereotype that women have inferior math skills is likely to be more available as a potential explanation for women's performance and, therefore, it is on these types of problems that we would expect gender differences to emerge. One could argue the alternative explanation, however, that women simply do not have the mathematical ability to do well on difficult word problem tests. The first study seeks to rule out this interpretation. In this study we gave participants one of two tests. One test was a word problem test. The other test was composed of problems that were the mathematical equivalents of the word problems, but they were already in their correct mathematical form. Thus, both tests required the same mathematical skills, but the word problem test required the extra step of formulating a strategy to solve the problem. We hypothesized that women and men would perform equally on the test that was strictly numerical, whereas women would underperform in comparison to men on the word problem test.

Method

Participants. Fifty-four women and 54 men from the State University of New York at Buffalo participated in the study in return for partial credit toward an Introductory to Psychology requirement.

Materials. Two tests were created. The word problem test was taken from a practice book for the GMAT. A meta-analysis by Hyde et al. (1990) found an effect size of 0.43 favoring men on the 1987 results of the GMAT. The numerical test was created by taking each word problem from the word problem test and converting it to a numerical or algebraic problem. That is, if a word problem called for setting up a series of equations, on the numerical version of the test, the series of equations was already set up. In this way, the actual math skills (i.e., algebra) needed were exactly the same on both tests. However, the word problem test required the extra step of forming a strategy to convert the words into solvable numerical problems. (See Table 1 for an example problem.) Each test was composed of 25 problems, each with five multiple-choice answers. Tests were scored by giving one point for each correct answer and subtracting one fifth of a point for each incorrect answer in order to control for guessing.

Table 1. Example of a Word Problem and an Equivalent Math Problem

From the word problem test:				
2. A sporting goods store sold 64 Frisbees in one week, some for \$3.00 and the rest for \$4.00 each. If receipts from Frisbee sales for the week totaled \$204, what is the fewest number of \$4.00 Frisbees that could have been sold?				
(a) 24	(b) 12	(c) 8	(d) 4	(e) 2
Equivalent problem from the numerical test:				
2. $3(64 - x) + 4(x) = 204$				
$x =$				
(a) 24	(b) 12	(c) 8	(d) 4	(e) 2

Design and procedure. The study had a 2 (gender) × 2 (type of test: word problem or numerical) between-subjects design. Participants were tested in groups by a female experimenter. The groups were of mixed gender, and the experimenter was unaware of which type of test each participant was taking. The type of test was also mixed within groups.

When participants arrived at the study they were simply told they would be taking a math test. All math tests had the same cover page. Participants read on the cover page that they would have 15 min to work on the test. They were also told that the test would be scored to correct for guessing and that they would receive their score at the end of the study. We did not actually score tests during the experimental session. This was included in the instructions to increase the relevance of the test for the participants. After all participants had a chance to read the instructions, the experimenter told them to begin. The experimenter stopped them after 15 min and debriefed them.

Results

In order to examine the results, we conducted an analysis of variance (ANOVA) with score as the dependent variable and gender and type of test as the independent variables. Because of the differences in the types of tests—participants who took the numerical test both scored higher and completed more problems than those who took the word problem test—the assumption of homogeneity of variance was violated (Levene's test for homogeneity of variances, $F(3, 104) = 6.42$, $p < .0001$). Therefore, we submitted the score variable to a logarithmic transformation in order to equalize the variance (Winer, Brown, & Michels, 1991). The ANOVA conducted on the transformed score showed a main effect for condition, $F = 35.7$, $p < .001$, and a main effect for gender, $F = 3.76$, $p = .06$. However, these main effects were qualified by a significant interaction, $F = 4.07$, $p < .05$. As can be seen in Figure 1, men and women did not differ in their performance on the numerical test, $t(54) = .25$, *ns*. On the word problem test, however, women significantly underperformed in comparison to the men, $t(50) = 3.07$, $p < .01$.

Discussion

In Study 1, women did not score as high as men on a word problem test. When the word problems were converted into their numerical mathematical equivalents, however, women and men performed equally well. The only difference between these two types of problems is that participants had to utilize a strategy to convert the word problems into their numerical equivalents. This study demonstrates that women have the mathematical skills and knowledge necessary to solve the problems. Something interfered, however, with women's ability to strategize and convert the problems when they completed the word problem test. Although we did not manipulate the amount of stereotype threat in the test-taking situation, we assume that the amount of stereotype threat in this situation—a difficult word problem test on which one's ability is being judged—should be high and that this high level of stereotype threat impaired women's performance.

In the second study we tested this reasoning more directly. In one condition men and women completed a difficult math test under standard conditions in which stereotype threat should be high. In a second condition, we reduced stereotype threat by telling participants that they were taking a gender-fair test. We then examined how stereotype threat interfered with problem solving. From the first study, all we can know is that women scored lower on the word problem test. We cannot tell whether they tried to strategize and used the wrong strategies or whether they could not strategize at all. In the next study, we took a closer look at the types of strategies used by men and women under varying conditions of stereotype threat.

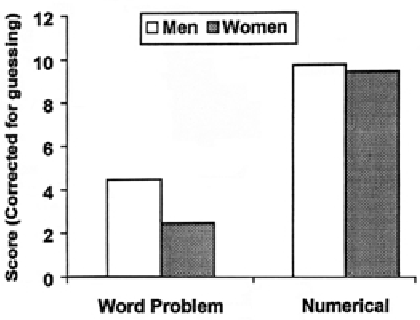

Fig. 1. Performance on word problem test and numerical test by gender.

Study 2

Research by Gallagher and DeLisi (1994) showed that men and women may differ in the types of problem solving strategies they use. These researchers had high school participants who had formerly scored very high (670 and above) on the mathematical portion of the SAT (SAT-M) work on a set of difficult SAT-M word problems while talking out loud about how they were solving the problems. Their problem-solving strategies were then coded. Gallagher and DeLisi found no overall difference in the scores of men and women on this test, but they did find that women tended to use more "conventional" strategies, defined as the type of strategies that are taught and used in math classes—applying known algorithms, plugging in numbers for variables—whereas men tended to use more "unconventional" strategies, such as logic and estimation. Although Gallagher and DeLisi (1994) did not find performance differences in their study, they speculated that differential strategy use might account for gender differences found in math achievement.

In order for us to make a closer examination of problem-solving strategies, we replicated the Gallagher and DeLisi (1994) study, using their test and protocol as well as their coding categories for strategy use. In our study, however, half of the participants were in a condition designed to reduce stereotype threat and the other half were in a control condition in which they were given the standard instructions, a condition presumably high in stereotype threat. In our previous research (Spencer et al., 1999; Spencer, Quinn, Davies, Gerharstein, & Pochert, 2000) we have shown that anxiety and diminished cognitive capacity are associated with stereotype threat. It seems reasonable that such cognitive and emotional states would interfere with women's ability to form strategies and that this difficulty in forming strategies would help account for women's underperformance when stereotype threat is high. We tested this reasoning in the current study.

Method

Participants. In this study we sought to replicate the methods of Gallagher and DeLisi (1994). These researchers had recruited students who had a score of 670 or above on the SAT-M. Although we initially also recruited participants with a score of over 650 on the SAT-M, it became clear that for those with a SAT-M score above 700, the test given was within their ability, and thus they would be unlikely to experience the stereotype threat (those with SAT-M scores above 700 had almost double the score on the test we gave of those with SAT-M scores less than 700). This may explain why Gallagher and DeLisi did not find any gender differences in overall performance. Thus, the data reported below are from 36 students from the University of Michigan who participated in the study in return for credit toward an Introductory to Psychology class and had a SAT-M score between 650 and 700. Although these students are between the 90th and 95th percentile on their SAT-M scores and thus have a very high level of math skills, the difficult problems given to them in the test described below presented a challenge for them.

Materials. The test was composed of 18 difficult multiple-choice SAT-M word problems (Gallagher & DeLisi, 1994), each with five answer choices. The problems could be solved in a number of different ways, and all required some amount of strategizing in order to be solved. The score on the test was computed by giving one point for each problem correct and subtracting a fifth of a point for each wrong answer.

Design and procedure. The design of the study was a 2 (gender) × 2 (high stereotype threat versus reduced stereotype threat) between-participants design. Participants completed the study individually. They were met by a female experimenter and told they would be asked to take a math test. Participants gave their consent to being tape-recorded during the study. They were then given an instruction sheet and told to read it carefully. On the instruction sheet they read that they were going to get a set of math problems developed for the SAT and that we wanted them to think out loud while solving the problems. They were told to solve the problems exactly as if they were taking the real SAT, except that they were to say everything that they were thinking while getting the answer. In addition, they would have 4 min to work on each problem, and the experimenter would not tell them if their answers were right or wrong, but might ask them to explain how they got their answers. The instructions for the high-stereotype-threat condition and the reduced-stereotyped-threat condition were the same, except one additional line was included in the instruction sheet for the reduced-stereotype-threat condition that stated: "Prior use of these problems has shown them to be gender-fair—that is, men and women perform equally well on these problems." The instruction sheets were given such that the experimenter was not aware to which condition each participant was assigned.

As participants worked on the math test, they spoke out loud and wrote notes or equations on paper (each problem was presented on its own sheet of paper). If they were silent for more than 30 s, the experimenter prompted them with "What are you thinking?" If they arrived at an answer without explaining their thought process, the experimenter asked them to explain how they got the answer. If participants had not solved a problem in 4 min, they were asked to move on to the next problem.

In order to examine the strategies used to solve the problems, we used the coding categories developed for the test by Gallagher and DeLisi (1994). There were eight possible coding categories, including three conventional strategies (using algorithms, assigning values to variables, and plugging in the multiple-choice options), two unconventional strategies (insight with algorithm and estimation or insight), as well as a category for guessing, a category for misinterpretation of the problem, and, lastly, a category for those who could not formulate any strategy. One rater (unaware of participant condition) coded all the transcripts, using both the transcript and the written notes of each participant. A second rater coded a random subset of the data. The interrater reliability was .80. All disagreements were resolved by the first rater.

Results

We conducted an ANOVA on score, with gender and condition (high versus reduced stereotype threat) as the between-participants independent variables. We found no main effects, both *F*s < 1, however, there was a marginally significant interaction, $F(1, 32) = 3.56$, $p = .07$. As shown in Figure 2, men ($M = 8.17$) outperformed women ($M = 4.64$) in the high-stereotype-threat condition, $t(32) = 2.0$, $p = .05$. In the reduced-stereotype-threat condition, women ($M = 7.05$) performed slightly better than men ($M = 6.03$), although this difference was not statistically significant, $t(32) = -.62$, *ns*.

In order to examine strategy use, we computed the number of times each type of strategy was used by each participant. For example, a participant might have used a conventional strategy four times, used an unconventional strategy twice, guessed three times, and not been able to formulate any strategy six times. We then computed means for each of the strategy categories for each of the conditions. We found that men and women did not differ in the extent to which they used any of the strategies, with one notable exception: When we examined the number of times participants could not formulate a strategy, we found a significant interaction between gender and condition, $F(1, 32) = 4.87$, $p < .05$. There were no main effects. Simple effects tests confirmed that women in the high-stereotype-threat condition were unable to formulate strategies for more of the problems than women in the reduced-stereotype-threat condition, $t(32) = 2.19$, $p < .05$, and that women in the

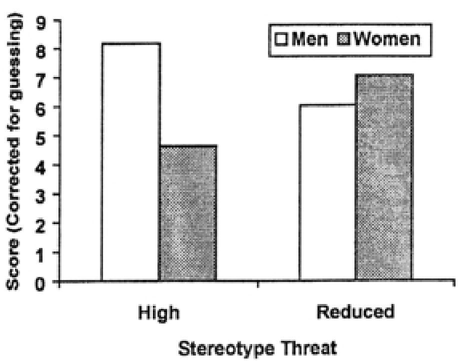

Fig. 2. Performance on math test as function of stereotype threat and gender.

high-stereotype-threat condition were more likely than men in the high-stereotype-threat condition not to be able to strategize, $t(32) = -1.99$, $p < .06$. Table 2 shows the percentage of use for each strategy. In the high-stereotype-threat condition, men could not formulate any strategy only 2% of the time, whereas women could not formulate a strategy 14% of the time. In contrast, in the reduced-stereotype-threat condition, men had no strategy 9% and women 4% of the time. In short, under conditions of high stereotype threat, women underperformed in comparison to men, and they were less likely to be able to formulate strategies. However, when they were told the same test was gender fair, thereby reducing stereotype threat, men and women performed equally on the test and did not differ in the ability to formulate and use strategies.

General Discussion

The quotes given at the beginning of this article depict examples of women who are trying to solve problems but are simply unable to formulate any strategy. The examples were taken from the condition of Study 2 in which stereotype threat was high. In this study, we found that when men and women were told to take a standardized math test—a test that was quite difficult for them—women scored lower and were less able to formulate any strategy on more problems in comparison to men. However, when this same test was portrayed as gender fair, men and women scored equally on the test, and women were just as able to formulate strategies as men. We believe that this small change in the testing situation changed the meaning of the situation for the participants. Although women were taking the same difficult test, when they experienced frustration, they no longer had to be concerned about their ability being judged by the stereotype.

Combined, these two studies answer some of the questions about gender differences in math performance. In Study 1, we showed that women were not scoring lower than men on word problem tests because of a difference in math skills. They were able to perform equally when the test was purely numerical. We hypothesize that on the word problems, stereotype threat interfered with women's ability to

Table 2. Percentage of Strategy Use on Math Problems by Gender and Condition

	Conventional strategy	Unconventional strategy	Guessing	Could not formulate any strategy
Stereotype threat				
Women	52%	26%	8%	14%
Men	55%	32%	11%	2%
Gender fair				
Women	60%	31%	5%	4%
Men	61%	24%	6%	9%

Note. Percentages reflect mean percentage of problems attempted with specific strategy.

strategize successfully to solve the word problems. In order to test this hypothesis more directly, we conducted Study 2. In this study, we found, as predicted, that when stereotype threat was experimentally reduced, women and men performed equally and solved problems using similar strategies. We conclude from these two studies that the knowledge of cultural stereotypes changes the testing situation for women such that their performance is depressed. This depression in performance can be partly explained by the interference of stereotype threat with the ability to formulate problem-solving strategies.

Because we utilized two different tests in these two studies, the role of stereotype threat on performance on word problems, per se, may not be clear. The first study does not rule out the alternative explanation that women are inferior in applying their math skills. Perhaps they scored lower on this test and other tests similar to it, such as the GMAT, because they have lesser abilities to conceptualize word problems. We have addressed this explanation in subsequent studies. In two recent studies, we (Spencer et al., 2000) gave men and women the same word problem test used in Study 1. We then varied the amount of stereotype threat in the test-taking situation. When the applicability of the stereotype was taken out of the situation, for example, by telling participants that the test was nondiagnostic of their ability, women and men performed equally on the test. In addition, when we primed stereotypic and counterstereotypic examples from the culture, the performance of men and women changed. When study participants saw, before taking this same word problem test, television commercials depicting women stereotypically, women performed significantly worse on the test than men. When they first saw television commercials depicting women counterstereotypically (i.e., being conversant in technical terms about automobiles), however, then women performed equally to men. Thus, in these two studies we have shown, using the same word problem test as in Study 1, that women do have the ability to strategize and solve the problems, but this ability can be suppressed by stereotype threat.

In Study 2, we found that stereotype threat interfered with women's ability to strategize. Interestingly, once women were able to formulate a strategy, they did not differ from men in the types of strategies they used. This is somewhat different from the results obtained by Gallagher and DeLisi (1994), who found that women were more likely to use conventional strategies taught in math classes. We are not certain why our results differed from theirs. One possible explanation is that we used college participants who may not have been currently taking math classes, whereas Gallagher and DeLisi (1994) used high school participants who were all enrolled in math classes. It may be that once students stop taking math classes, women's reliance on strategies taught in math classes decreases. A comparative sample of high school and college participants would be necessary to test this hypothesis.

One possible explanation for women's difficulty in formulating strategies when stereotype threat is high is that stereotype threat may reduce the cognitive resources available to generate strategies. When confronted with the possibility of being stereotyped, women might try to suppress these thoughts and thereby experience an increase in cognitive load. This increase in cognitive load might lead to a decrease in cognitive resources available to strategize about how to solve a problem. There is some evidence consistent with this interpretation. Steele and Aronson (1995) showed that when African Americans faced stereotype threat, the stereotypes about their group were activated at an unconscious level. However, these same participants also avoided describing themselves in stereotypic terms. These results seem to suggest that stereotype threat activates stereotypes about one's group that people then try to suppress. Such suppression is likely to take considerable cognitive resources (Wegner, 1994), and perhaps it is this drain that leads to the performance decrements and difficulty formulating strategies seen in the current research. More research is needed to address this question directly.

Consequences for Women's Math Achievement

Recent reviews indicate that the gap between men's and women's math performance is closing (Cole, 1997; Feingold, 1988). In fact, in math classes women's grades are at least as high as, if not higher than, men's (Kimball, 1989). At the same time, research indicates that stereotypes about women's inferior math skills remain pervasive in our culture (Martin et al., 1990). Thus, although women are making strides toward equality in math performance, they still have to confront the stereotype that they are not as good at math as men. And, despite the gains, women still underperform on some standardized tests and still are less likely than men to major in math and science or enter careers that demand these skills (Hewitt & Seymour, 1991; Hilton & Lee, 1988; Hyde et al., 1990).

The current analysis suggests an explanation for this pattern of findings. Women may perform just as well as men in classroom settings where the material is relatively familiar to them and they have practiced solution strategies in nonjudgmental situations, such as on homework problems. When they face new or especially challenging material, however, particularly in stressful, ability-diagnostic situations such as on achievement tests, stereotype threat may undermine women's performance. Confronting the cultural stereotype is likely to be an uncomfortable and anxiety-provoking experience. In time, women may come to distance themselves from situations and domains where they feel continually devalued (Steele, 1997a, 1997b).

What can be done about a cultural stereotype? Some might argue that if the stereotype is "out there" in the culture, there is nothing that can be done to stop its effects. However, we are not so pessimistic. Steele and his colleagues (1998) have recently put their theorizing about stereotype threat to the test in a much larger setting. They developed a program at the University of Michigan that was designed to lower stereotype threat and increase academic identification for African

Americans. Over a 4-year period this program has increased the grades and the retention of African American students.

In our studies we made very simple changes: simply adding a line in the instructions communicating that a test is gender fair or nondiagnostic. If parents and teachers became more aware of the many subtle ways they may shape math situations for girls and women, then we believe even greater changes could be made in women's attitudes toward math and their math performance. Indeed, if girls and women encounter fewer situations in which they experience stereotype threat, their increasing performance may one day break the ugly cycle of the stereotype leading to poor performance and the poor performance in turn feeding the stereotype.

References

Armstrong, J. M. (1981). Achievement and participation of women in mathematics: Results of two national surveys. *Journal for Research in Mathematics Education*, *12*(5), 356–372.

Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 83–103). San Diego, CA: Academic.

Bernardo, A. B. I., & Okagaki, L. (1994). Roles of symbolic knowledge and problem-information context in solving word problems. *Journal of Educational Psychology*, *86*(2), 212–220.

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: A developmental analysis. *Journal of Personality and Social Psychology*, *61*, 351–365.

Cole, N. S. (1997). *The ETS gender study: How females and males perform in educational settings*. Princeton, NJ: Educational Testing Service.

Constantinople, A., Cornelius, R., & Gray, J. (1988). The chilly climate: Fact or artifact? *Journal of Higher Education*, *59*, 527–550.

Crozet, J., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, *24*(6), 588–594.

Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin*, *17*, 306–315.

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

Eccles, J. S., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs: Journal of Women in Culture and Society*, *11*(21), 367–380.

Eccles, J. S., Jacobs, J. E., & Harold, R. E. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, *46*, 183–201.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, *43*(2), 95–103.

Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology*, *74*(2), 435–452.

Gallagher, A. M., & DeLisi, R. (1994). Gender differences in Scholastic Aptitude Test–Mathematics problem solving among high-ability students. *Journal of Educational Psychology*, *86*, 204–211.

Hewitt, N. M., & Seymour, E. (1991). *Factors contributing to high attrition rates among science and engineering undergraduate majors*. Report to the Alfred P. Sloan Foundation, New York.

Hilton, T. L., & Lee, V. E. (1988). Student interest and persistence in science: Changes in the educational pipeline in the last decade. *Journal of Higher Education*, *59*, 510–526.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*(2), 139–155.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., and Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly*, *14*(3), 299–324.

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *Journal of Personality and Social Psychology*, *63*(6), 932–944.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin*, *105*(2), 198–214.

Leinhardt, G., Seewald, A., & Engel, M. (1979). Learning what's taught: Sex differences in instruction. *Journal of Educational Psychology*, *71*, 432–439.

Levine, D. U., & Ornstein, A. C. (1983). Sex differences in ability and achievement. *Journal of Research and Development in Education*, *16*, 62–72.

Lummis, M., & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology*, *26*(2), 254–263.

Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Martin, C. L., Wood, C. H., & Little, J. K. (1990). The development of gender stereotype components. *Child Development*, *61*, 1891–1904.

Meece, J. L., Eccles, J. S., Kaczala, C. M., Goff, S. B., & Futterman, R. (1982). Sex differences in math achievement: Towards a model of academic choice. *Psychological Bulletin*, *91*, 324–348.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic groups* (College Board Report No. 93-1, ETS No. 94.27). New York: College Entrance Examination Board.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*(1), 80–83.

Spencer, S. J., Quinn, D. M., Davies, P. G., Gerharstein, R., & Pochert, A. (2000). *Stereotypic commercials and women's math performance: The effect of priming the cultural stereotype*. Unpublished manuscript, University of Waterloo, Waterloo, Ontario, Canada.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(4), 4–28.

Steele, C. M. (1997a). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.

Steele, C. M. (1997b, October). *Disidentification in the artful dodging of threat and the negotiation of social structure*. Paper presented at the Society of Experimental Social Psychology convention, Toronto, Ontario, Canada.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811.

Steele, C. M., Spencer, S. J., Nisbett, R., Hummel, M., Harber, K., Schoem, D., & Carter, K. (1998). *African American college achievement: A "wise" intervention*. Unpublished manuscript, Stanford University, Stanford, CA.

Stockard, J., & Wood, J. (1984). The myth of female underachievement: A reexamination of sex differences in academic underachievement. *American Education Research Journal*, *21*, 825–838.

Strenta, A. C., Elliot, R., Adair, R., Scott, J., & Matier, M. (1993). *Choosing and leaving science in highly selective institutions*. Unpublished report to the Alfred P. Sloan Foundation, New York.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, *66*, 21–36.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*(1), 34–52.

Wigfield, A., & Eccles, J. S. (1989). Test anxiety in elementary and secondary school students. *Educational Psychologist*, *24*(2), 159–183.

Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology*, *80*(2), 210–216.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

DIANE M. QUINN received her PhD in social psychology from the University of Michigan in 1999. She is now an assistant professor in the department of psychology at the University of Connecticut. Her research interests include ideology, stigma, self-objectification, stereotype threat, and self-esteem.

STEVEN J. SPENCER received his PhD in social psychology from the University of Michigan. He is currently an associate professor at the University of Waterloo in Canada. His research interests include stereotyping, prejudice, and the self-concept.