

# Converging Evidence That Stereotype Threat Reduces Working Memory Capacity

Toni Schmader and Michael Johns  
University of Arizona

Although research has shown that priming negative stereotypes leads to lower performance among stigmatized individuals, little is understood about the cognitive mechanism that accounts for these effects. Three experiments tested the hypothesis that stereotype threat interferes with test performance because it reduces individuals' working memory capacity. Results show that priming self-relevant negative stereotypes reduces women's (Experiment 1) and Latinos' (Experiment 2) working memory capacity. The final study revealed that a reduction in working memory capacity mediates the effect of stereotype threat on women's math performance (Experiment 3). Implications for future research on stereotype threat and working memory are discussed.

One of our brightest and most motivated undergraduate research assistants was recently lamenting her upcoming date with the Graduate Record Examination (GRE). In describing her past struggles with these sorts of standardized tests, she stated that as soon as she sees the math problems, she becomes intensely fascinated by the physical details of her pencil or any other proximal stimulus, so long as it is not the actual math problem she is meant to solve. Although her experience could represent nothing more than active avoidance of an aversive stimulus, we wondered if it might also illustrate one of the ways that stereotype threat interferes with performance on complex tests of cognitive abilities. Cognitive psychology has identified working memory capacity as the ability to focus one's attention on a given task while keeping task-irrelevant thoughts at bay (Engle, 2001). Thus, one explanation for our research assistant's frustration with standardized math tests is that she experiences lower levels of working memory capacity in these testing situations, perhaps because gender stereotypes place an extra burden on her cognitive resources. In the research presented here, we set out to test the hypothesis that stereotype threat interferes with performance on complex cognitive tasks by reducing individuals' working memory capacity. In this article, we

integrate existing research on stereotype threat with what is known about working memory capacity. We then present the results of three experiments that examine whether manipulations of stereotype threat reduce working memory capacity.

## Stereotype Threat and Performance

Stereotype threat refers to the phenomenon whereby individuals perform more poorly on a task when a relevant stereotype or stigmatized social identity is made salient in the performance situation. Steele and his colleagues (Steele, 1997; Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002) maintained that this reduced performance results from an added pressure or concern that a poor performance could be seen as confirming a negative social stereotype about their ingroup. Thus, in sharp contrast to socialization theories, inherent ability theories, or even educational resource theories for why men outperform women on math tests or why European Americans outperform African Americans on standardized tests, stereotype threat offers a uniquely situational explanation for these group-performance differences (see Steele, 1997, for review).

In support of the stereotype threat explanation, research shows that group-performance differences can be eliminated when the same test is given in a stereotype-free context (see Steele et al., 2002, for a review). For example, African Americans show increased stereotype activation and perform worse than their White peers when the task they are performing is described as diagnostic of intellectual ability (Steele & Aronson, 1995). However, when the same task is framed as unrelated to intelligence, levels of stereotype activation are much lower and African American students perform equally to White students. Similarly, women perform worse than men on a math test when they are told that the test has revealed gender differences in the past, but they perform equally to men when they are told that the test is "gender fair" (Spencer, Steele, & Quinn, 1999). The ease with which stereotype threat can be created in testing situations is demonstrated further by evidence that White men, a group that is not traditionally thought of as being negatively stereotyped as being poor at math, perform more poorly on a math test when they believe they will be compared with Asian men (Aronson et al., 1999).

---

Toni Schmader and Michael Johns, Department of Psychology, University of Arizona.

Portions of this research were presented at the 14th Annual Meeting of the American Psychological Society, New Orleans, Louisiana, June 2002, and at the Fourth Annual Meeting of the Society for Personality and Social Psychology, Los Angeles, February 2003. This research was supported in part by National Science Foundation Grant BCS-0112427 and by a faculty small grant awarded to Toni Schmader from the University of Arizona Foundation. We are indebted to the countless hours put in by the research assistants who ran these studies—Kristen Hooks, Amy Hooks, Jodie Deutsch, Ryan Hendrix, Mike Stanfill, David Trotter, Yassiin Nasser, and Martin Kaasa—and also to Amy Baesler for her help preparing materials. We also thank Marchelle Barquissau and Brian Lickel for their valuable help and advice in planning this research.

Correspondence concerning this article should be addressed to Toni Schmader, Department of Psychology, University of Arizona, Tucson, Arizona 85721. E-mail: schmader@u.arizona.edu

Taken together, these findings suggest that activating negative stereotypes about a social identity that one possesses can create an extra situational burden that interferes with the ability to perform as well at a mental task as might otherwise be possible. Although the body of research establishing the existence of these effects is ever expanding, the processes by which performance is reduced have remained elusive in most studies. The assertion that reduced performance results from an added pressure or concern (Steele, 1997; Steele & Aronson, 1995) has led many researchers to examine the role of several affective mechanisms, such as anxiety, evaluation apprehension, or physiological arousal, in producing stereotype threat. For example, several studies have demonstrated that stereotype threat conditions lead to higher levels of anxiety that parallel performance decrements on complex tasks (Aronson et al., 1999; Spencer et al., 1999; Stone, Lynch, Sjomeling, & Darley, 1999). In addition, Spencer et al. (1999) found evidence that self-reported anxiety partially mediated the effects of a stereotype threat manipulation on women's math performance. However, many other stereotype threat studies have reported no differences in self-reported anxiety between stereotype threat conditions (Gonzales, Blanton, & Williams, 2002; Schmader, 2002; Schmader & Johns, 2003; Steele & Aronson, 1995).

Although self-reported measures of anxiety show mixed results, there is some evidence to suggest that stereotype threat involves heightened levels of anxiety and arousal. For example, Blascovich, Spencer, Quinn, and Steele (2001) reported that African Americans show increases in blood pressure under conditions of stereotype threat, and in our previous research (Schmader & Johns, 2003), women who thought that their test performance would be used as an indicator of women's math ability, in general, felt that it was more important that they do well on the test. In addition, manipulations designed to reduce anxiety or arousal, such as engaging in self-affirmations or expressive writing (Martens, Johns, Greenberg, & Schimel, 2002), or providing a cue to misattribute arousal (Stone et al., 1999), seem to diffuse the effects of stereotype threat on performance.

The above findings provide converging evidence that stereotype threat involves affective experiences associated with increased anxiety and apprehension. However, these past studies also highlight the extent to which researchers interested in the underlying mechanism involved in stereotype threat have focused almost exclusively (with one notable exception, Quinn & Spencer, 2001, discussed below) on the affective side of the threat equation. In contrast, our goal in the present research was to examine how the fear of confirming a negative stereotype about a salient group identity disrupts cognitive processing. Because most stereotype threat studies involve evidence of reduced performance on a complex cognitive task, we were interested in understanding the nature of the cognitive disruption that stereotype threat causes. Specifically, we were interested in the effects of stereotype threat on working memory capacity.

### Working Memory Capacity

The contemporary conceptualization of working memory capacity has its roots in theory and research on short-term memory, or the type of memory used to retain and manipulate information for immediate or near-immediate use. Baddeley and Hitch (1974) expanded earlier theory on short-term memory by proposing a

model of working memory that distinguished between three distinct but interactive cognitive functions: the phonological loop, the visuospatial sketchpad, and the central executive processor. The current concept of working memory capacity is an articulated version of the central executive processor (Engle, 2001), and refers to that type of memory that is used to focus attention on temporarily activated information of interest while inhibiting other information that is irrelevant to the task at hand. Thus, working memory capacity includes both the temporary storage of information as well as an attentional capability (Engle, Tuholski, Laughlin, & Conway, 1999). People with higher working memory capacity are better able to suppress task-irrelevant information (Rosen & Engle, 1998) as evidenced by their lower susceptibility to the cocktail party effect (Conway, Cowan, & Bunting, 2001).

Research on working memory has focused almost exclusively on assessing individual differences, the basic assumption being that interpersonal variation in working memory capacity is predictive, if not indicative, of variation in fluid intelligence (Engle et al., 1999). Thus, much of the research that has been conducted on working memory capacity has been aimed at developing construct valid measures that can predict performance on complex cognitive tasks (Engle et al., 1999; Klein & Fiss, 1999; Oberauer, Süß, Schulze, Wilhelm, & Wittman, 2000). This research indicates that individuals with higher levels of working memory capacity perform better on a variety of cognitive ability tests (e.g., La Pointe & Engle, 1990; Turner & Engle, 1989). Similarly, the measure of working memory capacity used in the present research has been shown to correlate significantly with both verbal ( $r = .34$ ) and quantitative ( $r = .33$ ) Scholastic Aptitude Test (SAT) scores (Turner & Engle, 1989).

The test of working memory capacity that we used in our research is the operation-span task developed by Turner and Engle (1989). The test consists of two separate tasks that are performed concurrently. One task is a processing task in which participants are presented with a mathematical equation [e.g.,  $(2 \times 3) - 5 = 1$ ] and must decide whether the answer given is correct or incorrect. The second task is a memory span task in which participants are given a word to recall at a later point. Each word is presented after an equation. Thus, participants might be presented with five equations and word pairings before being cued to recall the five words. Participants' ability to correctly recall the words provides an index of working memory capacity in that it reflects the ease with which they can both process the math problems while simultaneously holding the words in their mind.

Although working memory capacity research has not traditionally examined how situational manipulations can reduce working memory capacity temporarily, there is evidence to suggest that chronic levels of stress and anxiety might be associated with lower levels of working memory capacity (Eysenck & Calvo, 1992). For example, individuals who score high in trait anxiety or report experiencing more life stress perform worse than their less-stressed counterparts on measures of working memory capacity (e.g., Derakshan & Eysenck, 1998; Klein & Boals, 2001). Klein and Boals (2001) argued that life stress reduces working memory capacity because people under stress dedicate some of their mental resources to suppressing unwanted negative thoughts and feelings that intrude during other tasks. In addition to the direct relationship between general stress and working memory capacity, other research suggests that stressful performance situations reduce the

working memory capacity of those who are high in trait anxiety (Sorg & Whitney, 1992) or math anxiety (Ashcraft & Kirk, 2001).

This collection of findings is consistent with the supposition that working memory capacity might be reduced in stressful testing situations because stress-related thoughts consume valuable cognitive resources. We conceptualize stereotype threat as a stressor in that a negative social stereotype that is primed in a performance situation poses a threat to one's social identity (Schmader, 2002). Furthermore, given evidence that stereotype threat involves increased levels of stereotype activation (Davies, Spencer, Quinn, & Gerhardstein, 2002; Steele & Aronson, 1995), one might imagine that cognitive resources are being expended processing this information, leaving fewer resources for the task at hand (Spencer, 2003). Quite simply, the present research was designed to test the idea that a manipulation of stereotype threat would result in lower working memory capacity among individuals targeted by that stereotype.

Although working memory capacity, per se, has not been previously examined in stereotype threat research, work by Quinn and Spencer (2001) does highlight some of the cognitive effects of stereotype threat. They found that although women performed worse than men on a set of complex word problems, they performed equally to men when solving the same problems with the key equations extracted from the text of the word problem. A second study also demonstrated that women under stereotype threat had greater difficulty generating strategies they could use to solve the word problems. Quinn and Spencer interpreted these results as evidence that stereotype threat interferes with problem-solving ability, perhaps because it results in diminished cognitive resources. Our research was designed to test this possibility directly.

### Overview of Research

The primary goal of this research was to investigate the impact of stereotype threat manipulations on working memory capacity. In two of the three studies presented here, we focused on women who are targeted by the stereotype that women are inferior to men in mathematical ability. Of interest, the role of working memory capacity in solving mathematical equations has been studied (e.g., Ashcraft & Kirk, 2001), and there is evidence that individuals who have higher trait math anxiety also score lower on an individual-difference measure of working memory capacity (Ashcraft, 2002). However, working memory capacity has not been examined as an explanation for gender differences in math performance, let alone differences that are produced by experimentally manipulated conditions of stereotype threat.

Experiment 1 tested the hypothesis that women (but not men) would show reduced working memory capacity when a testing situation was framed as measuring math ability. Experiment 2 was a conceptual replication of Experiment 1, comparing the working memory capacity of Caucasian and Latino participants when the task was said to be related to general intelligence. In the final study, we used a different manipulation of stereotype threat and a measure of working memory that does not involve math to directly test the hypothesis that a reduction in working memory mediates the effects of stereotype threat on women's performance on a standardized math exam.

## Experiment 1

In our first study, women and men completed a measure of working memory capacity under conditions of stereotype threat or not. We hypothesized that if stereotype threat interferes with working memory, then women who complete a working memory test described as a measure of mathematical ability—a stereotype-relevant ability domain—would show lower working memory scores compared with men and compared with women in a control condition. This pattern of results would provide initial evidence that making the stereotype about women's math abilities applicable to test performance impairs working memory capacity—a psychological capacity critical to performing well on complex intellectual tasks (e.g., Engle et al., 1999).

### Method

#### Participants and Design

The participants in this study were 40 male and 35 female undergraduate psychology students who participated for credit toward a course requirement or for \$10. Following past stereotype threat research on women and math (Schmader, 2002; Spencer et al., 1999), we selected participants who indicated in a previous mass survey session that they had scored 500 or higher on the quantitative section of the SAT (or equivalent converted American College Test score). In the same mass survey, we assessed stereotype knowledge with the question, "Regardless of what you personally believe, do you think there is a stereotype about women having less mathematical ability than men?" (rated on a 7-point scale from 1 = *not at all* to 7 = *definitely*). Only those who responded at or above the scale midpoint of 4 were recruited to participate in the study. Participants were randomly assigned to one of two test description conditions in a 2 (male or female)  $\times$  2 (stereotype threat or control) factorial design. Data from 9 men and 7 women in the working memory test description condition were excluded from analyses because of lost data, leaving a final sample of 31 men and 28 women.<sup>1</sup>

#### Materials

**Working memory test.** To measure working memory capacity, we adapted a dual-processing test called the operation-span task that has been developed and used extensively by Engle and his colleagues to assess working memory (e.g., La Pointe & Engle, 1990; Turner & Engle, 1989). In this task, participants evaluate mathematical equations while memorizing words for later recall. The equations begin with the multiplication or division of two positive integers [e.g.,  $9 \times 6$ ]. The product of this operation is then added to, or subtracted from, another positive integer. The answer for the entire operation is included in the expression, and the participant is asked to evaluate whether the equation is correct or incorrect [e.g., Is  $(9 \times 6) - 4 = 50?$ ]. A word is presented after each mathematical equation and at the end of a series of equation/word combination trials (i.e., a set) participants are asked to recall as many of the words from the preceding series as possible. The math equations are merely meant to engage participants in a certain amount of cognitive processing; working memory capacity is indexed as the number of words that participants recall correctly from each equation/word set.

<sup>1</sup> The lost data were caused by a computer programming error when we initially began the study. The computer failed to record the data from one block of trials in the control condition, resulting in missing data for this subset of participants. Including these 16 participants in the primary analyses does not change the results.

The mathematical equations used in the present study were presented in the same format as just described and were designed to be moderately challenging. For equations requiring multiplication, the integers used in this expression ranged from 5 to 12. For equations requiring division, the integers used were multiples of the integers ranging from 5 to 16 [e.g., 128/16]. The integer added to, or subtracted from, the solution to the first expression could range from 5 to 14. The answer presented with each equation was either correct or incorrect by plus or minus one and could be positive or negative. A pool of 72 equations was generated using these criteria (36 correct and 36 incorrect). The 72 words used in the test were randomly selected from a pool of one-syllable words used by La Pointe and Engle (1990).

The test we created included 18 sets of equation/word trials that contained three to five equation/word combinations (i.e., there were six blocks of each set size for a total of 72 equation/word trials). The sets were presented in random order so participants did not know how many equations and words they would be required to evaluate and recall at the beginning of each set. The equations and words were presented in random order within each set, but equations and words were assigned to the same set for each test. The test was administered on a computer and the participant controlled the presentation of the stimuli with their responses. Participants were instructed to evaluate the correctness of the equations quickly and accurately while also remembering the words for later recall. Each set began with the presentation of an equation to be evaluated for correctness. After recording their evaluation using the keyboard (1 = correct, 2 = incorrect), participants were presented a to-be-remembered word for 2 s. A blank screen lasting 1 s separated the presentation of each equation and word. After presentation of all equation/word combinations in a set, participants were prompted to recall all of the words in that set. Each set was separated by the prompt "next set," which was displayed for 3 s. The computer recorded the words recalled, participants' correct and incorrect responses to the equations, and the time spent on each equation.

### Procedure

Two female research assistants ran the experimental sessions in same-gender groups ranging from 2 to 4 participants. Upon entering the lab, participants were seated in individual rooms and asked to listen to, and read along with, a prerecorded description of the study broadcast on an intercom system and displayed on the computer monitor. This study description served as the manipulation of stereotype threat and was delivered by a male who identified himself as a researcher in the psychology department. In the control condition, he described the test as a reliable measure of working memory capacity. He went on to describe working memory capacity as the ability to hold different pieces of information simultaneously while trying to process one specific piece of information. In the stereotype threat condition, the male researcher described the test as a reliable measure of "quantitative capacity." Quantitative capacity was described as the ability to solve complex mathematical equations while trying to process multiple pieces of information related to the problem-solving task. To prime the stereotype related to women's math ability, participants were also told that gender differences in math performance might stem from underlying gender differences in quantitative capacity. To increase the personal relevance of the performance situation, participants in both conditions were informed that they would receive feedback about their performance.

After the stereotype threat manipulation, participants were presented with a general overview of the testing procedure. They were told that their performance on the test would be based on both how accurately they evaluated the equations and the number of words they recalled correctly. After the instructions, participants were allowed to complete a practice set of three equation/word trials. Before beginning the test, participants in both conditions were asked to indicate whether they were right-handed or left-handed, and participants in the stereotype threat condition were also asked to indicate their gender and their last name.

Following the test, participants completed a brief test experience questionnaire that contained an anxiety scale and items related to perceptions of the test and testing situation. Anxiety during the tests was measured using questions adapted from the Spielberger State Anxiety Scale (Spielberger, Gorsuch, & Lushene, 1970). Participants rated on a 7-point scale (ranging from 1 = *not at all* to 7 = *very much*) how much they felt *anxious, comfortable, jittery, worried, at ease, nervous, relaxed, and calm* while taking the test. The items on the scale were averaged (after reverse scoring *comfortable, at ease, relaxed, and calm*) to form an index of state anxiety, where higher numbers indicated more anxiety ( $\alpha = .90$ ). Participants were also asked to rate the difficulty of the working memory test on a 7-point scale ranging from 1 (*extremely easy*) to 7 (*extremely difficult*).

To determine if participants perceived the test description in a manner consistent with stereotype threat, we asked them to rate how much they agreed with the following statements (tailored to participant gender): "I am concerned that the researcher will judge [women/men], as a whole, based on my performance on this test" and "The researcher will think that [women/men], as a whole, have less math ability if I did not do well on this test." Responses were recorded using a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*) and were averaged to create a reliable index of gender identity threat ( $\alpha = .90$ ). After completing this questionnaire the participants were debriefed and thanked.

### Results

An initial Gender  $\times$  Stereotype Threat analysis of variance (ANOVA) on participants' quantitative SAT scores revealed that men's SAT scores ( $M = 620$ ) were higher than women's ( $M = 579$ ),  $F(1, 55) = 4.60, p < .05$ . Because previous research has found that performance on the operation-span task is significantly correlated with SAT scores (Turner & Engle, 1989), we controlled for these group differences in SAT scores by conducting 2 (gender)  $\times$  2 (stereotype threat) analyses of covariance (ANCOVAs) controlling for SAT when SAT was found to be a significant covariate.

### Working Memory Capacity

**Absolute span score.** In past research, working memory capacity is assessed either as the total number of words recalled or the number of words recalled taking into account only sets recalled perfectly (called the *absolute span score*). The absolute span score is derived by summing the total number of words from only those sets of words where all the words in the set were recalled correctly. So, if a participant only recalled three words from a four-word set, then these three words would not count toward the total score. If, however, all four words were recalled correctly, all four words would count toward the final score. In this way, the absolute span score is thought to provide a more sensitive measure of working memory capacity (La Pointe & Engle, 1990). In all of the studies reported here, we used the absolute span score as the assessment of working memory capacity; however, in all three studies, analyses using the total words recalled yielded the same conclusions.

An ANCOVA on the absolute span score yielded significant main effects of gender,  $F(1, 54) = 4.81, p < .05$  (women  $M = 44.23$  and men  $M = 49.80$ ), and stereotype threat,  $F(1, 54) = 23.84, p < .001$  (stereotype threat  $M = 41.05$  and control  $M = 52.98$ ), and the predicted two-way interaction,  $F(1, 54) = 15.69, p < .001$ . SAT was a significant covariate,  $F(1, 54) = 4.98, p < .05$ . Simple main effects tests revealed that women in the stereotype threat condition ( $M = 33.44$ ) recalled

fewer words than men in the stereotype threat condition ( $M = 48.66$ ,  $F(1, 54) = 19.38$ ,  $p < .001$ ,  $d = 1.19$ ), and than women in the control condition ( $M = 55.03$ ,  $F(1, 54) = 37.49$ ,  $p < .001$ ,  $d = 1.66$ ). The span score for men in the control condition ( $M = 50.94$ ) was not significantly different from the score of men in the stereotype threat condition and women in the control condition ( $F_s < 1.5$ ). The means are displayed in Figure 1. This pattern of results supports our prediction that stereotype threat reduces women's working memory capacity.

**Equation evaluation.** Although working memory is assessed as a function of performance on the word recall, we also analyzed performance on the equations to assess whether there were any significant differences due to the stereotype threat manipulation. Prior research indicates that manipulations of stereotype threat only reduce women's performance on very complex tests of their math ability (Spencer et al., 1999), typically those that require extracting an equation from a word problem (Quinn & Spencer, 2001). Thus, we did not expect the stereotype threat manipulation to affect women's ability to evaluate the equations correctly. In fact, Quinn and Spencer (2001) found that stereotype threat does not reduce women's ability to solve basic equations like those used in our task. In line with these past results, there were no significant effects of gender or stereotype threat on percentage of equations solved correctly (Grand mean ( $GM$ ) = 78%;  $F_s < 1.5$ ). Analysis of the average amount of time (in seconds) spent on each equation revealed a marginal main effect of stereotype threat,  $F(1, 54) = 3.44$ ,  $p < .10$ . SAT was a significant covariate,  $F(1, 54) = 4.41$ ,  $p < .05$ , but no other effects were significant ( $F_s < 1.5$ ). Regardless of gender, participants in the stereotype threat condition tended to spend more time evaluating each equation ( $M = 8.53$ ) than did participants in the control condition ( $M = 7.42$ ).

### Test Experience Questionnaire

**Anxiety.** Analysis of the anxiety measure did not yield any significant main effects or interactions ( $F_s < 2$ ).

**Perceived difficulty.** Analysis of the difficulty ratings revealed a significant two-way interaction,  $F(1, 54) = 4.71$ ,  $p < .05$ . SAT was a significant covariate,  $F(1, 54) = 4.04$ ,  $p < .05$ ; neither main effect was significant ( $F_s < 1.5$ ). Simple effects analyses showed

that within the stereotype threat condition, women rated the test as more difficult ( $M = 4.63$ ) than did men ( $M = 3.54$ ),  $F(1, 54) = 5.70$ ,  $p < .05$ . The difficulty ratings of women ( $M = 3.96$ ) and men ( $M = 4.26$ ) in the control condition were not significantly different ( $F < 1$ ).

**Gender identity threat.** Analysis of the gender identity threat composite yielded a main effect for the stereotype threat manipulation,  $F(1, 55) = 8.29$ ,  $p < .01$ . Both women and men in the stereotype threat condition expressed greater concern ( $M = 2.47$ ) that the researcher would evaluate their performance in terms of their gender identity compared with participants in the control condition ( $M = 1.61$ ). No other effects were significant ( $F_s < 2$ ). This mean pattern reveals that the manipulation did produce some conscious awareness that the researcher might use their gender as a factor to evaluate their performance, but this concern was not unique to women.

### Discussion

The results of this first study provide initial support for the hypothesis that a stereotype threat manipulation—framing a task in terms of a stereotyped group deficiency—can lead to a measurable decrease in cognitive resources. As predicted, women completing a working memory test described as a test related to mathematical ability showed reduced cognitive capacity, as measured by the number of words they were able to recall within the task. Men, and women in a nonthreat control condition, did not exhibit this working memory decrement and were able to recall an equal number of words. This is the exact pattern of performance that would be predicted by stereotype threat. Measures designed to assess the psychological experience created by the testing situation indicated that women experienced the test as more difficult than men in the stereotype threat condition despite the finding that both men and women indicated that they were concerned their performance on the test might be evaluated as a function of their gender identity. This would seem to indicate that although gender was salient for both women and men in the stereotype threat condition, it was only women, who are specifically targeted by the stereotype, that were affected by the fear of confirming the researchers' expectations. As in some other stereotype threat research (e.g., Schmader, 2002; Steele & Aronson, 1995), however, participants did not report differential levels of anxiety as a function of the test description.

### Experiment 2

The results of Experiment 1 represent an important step in understanding exactly how stereotype threat works to undermine the performance of negatively stereotyped group members on challenging academic tasks. In the stereotype threat condition, both women and men expressed apprehension that the researcher would evaluate their performance in relation to their gender identity, but only the performance of women suffered under this concern. This pattern of results is consistent with our hypothesis that having to contend with negative group stereotypes places an additional cognitive burden on people that can interfere with their ability to perform up to their potential on complex cognitive tasks. One further question is whether the same effect can be found with a different stereotyped group in a different domain. Thus, the second study was designed to replicate the negative effects of

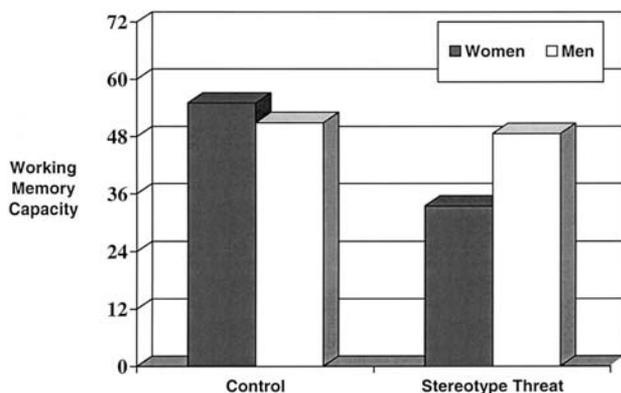


Figure 1. Experiment 1: The effects of gender and stereotype threat on participants' working memory capacity.

stereotype threat on the working memory capacity of Latinos who are negatively stereotyped on intellectual tasks (e.g., Gonzales et al., 2002).

An additional goal of our second experiment was to address an alternative explanation for the results of Study 1. The stereotype threat manipulation we used in Experiment 1 involved two components: We described the test as a measure of math ability, and we also explicitly mentioned the idea that there are gender differences in math performance. One could argue that this direct priming of gender stereotypes might have created an experimental demand that produced the results we observed rather than actual effects on cognitive capacity. The fact that we did not observe effects on women's performance when evaluating the math equations—both accuracy and time spent evaluating the equations—speaks against the idea that participants' behavior was primarily affected by expectancies; however, we wanted to address this possibility by using a more subtle manipulation of stereotype threat that did not explicitly mention group differences in performance.

For this study we compared the performance of Latino students to the performance of White students under conditions of stereotype threat. On the basis of the cultural stereotype that Latinos are less intelligent than Whites (Gonzales et al., 2002), we created stereotype threat by framing the working memory test as a task that was highly predictive of general intelligence. Rather than explicitly mentioning ethnic differences in performance, we primed ethnic identity in the stereotype threat condition by asking participants to identify their ethnicity on a demographic questionnaire prior to completing the working memory measure (Steele & Aronson, 1995). We expected that Latino students would show evidence of reduced working memory capacity compared with Whites when the test was described as a measure related to intelligence. We did not expect any performance differences when the test was described as a measure of working memory.

## Method

### Participants and Design

We recruited 33 Latino (20 women, 13 men) and 40 White (27 women, 13 men) psychology students only on the basis of their self-reported ethnicity in an earlier mass survey. All participants received credit for a research requirement. Participants were randomly assigned to one of two conditions in a 2 (Latino or White)  $\times$  2 (stereotype threat or control) factorial design. The data of 1 White participant were lost because of a computer error, leaving a final sample of 72.

### Materials and Procedure

Two White female experimenters conducted the sessions in mixed ethnic groups ranging from 2 to 4 participants.<sup>2</sup> The materials and procedures were the same as Study 1 except for the stereotype threat manipulation. The stereotype threat manipulation was again presented in a prerecorded description delivered by a male "researcher," but unlike the previous study, the researcher described the test as a reliable measure of working memory capacity in both the control condition and the stereotype threat condition. In the stereotype threat condition, the researcher went on to say that research had shown that performance on the working memory test is "highly predictive" of performance on intelligence tests and that their performance on the test would be used to "help establish norms for different groups." The test description in both conditions ended by inform-

ing participants that they would receive feedback about their performance. As in Study 1, this was done to increase the personal relevance of the testing situation. In addition, to further prime ethnic identity in the stereotype threat condition, participants in this condition were also asked to indicate their ethnicity before beginning the test.

After the task, all participants completed a test experience questionnaire containing the same measure of anxiety ( $\alpha = .87$ ) and difficulty used in the previous study. The questionnaire also included the two identity threat items from the previous study modified to assess whether participants thought that the researcher would evaluate them on the basis of their ethnicity. Responses were averaged to create a reliable index of ethnic identity threat ( $\alpha = .90$ ). We also had data on participants' self-reported verbal and quantitative SAT scores from a mass survey that they had completed at the beginning of the semester. Participants were debriefed and thanked after completing the final questionnaire.

## Results

Because of the quasi-experimental design of this study, we conducted initial 2 (ethnicity)  $\times$  2 (gender)  $\times$  2 (stereotype threat) ANOVAs on participants' quantitative and verbal SAT scores to discern whether there were any group differences in these variables that should be controlled in our analyses. There were no significant differences in quantitative SAT (all  $F$ s  $<$  2), but there was a marginal ethnic difference in verbal SAT scores,  $F(1, 59) = 2.60, p = .10$ . No other effects were significant ( $F$ s  $<$  1.5). Because prior research reveals a significant relationship between verbal SAT and working memory capacity (Turner & Engle, 1989), we controlled for this variable in all analyses where it was a significant covariate. Degrees of freedom for these analyses are lower because controlling for verbal SAT resulted in the loss of 5 additional participants who did not report their SAT scores.

### Working Memory Capacity

**Absolute span score.** Analysis of the absolute span score produced the predicted Ethnicity  $\times$  Stereotype Threat interaction,  $F(1, 58) = 4.48, p < .05$  (see Figure 2). Verbal SAT was a significant covariate,  $F(1, 58) = 11.83, p < .001$ . Simple effects testing revealed that Latinos in the stereotype threat condition ( $M = 41.29$ ) recalled significantly fewer words than did Whites in the stereotype threat condition ( $M = 51.93$ ),  $F(1, 58) = 6.45, p < .05, d = .66$ , and Latinos in the control condition ( $M = 50.65$ ),  $F(1, 58) = 4.19, p < .05, d = .55$ . Recall by Whites in the control condition ( $M = 48.10$ ) was equivalent to that of Latinos in the control condition and to the recall of Whites in the stereotype threat condition ( $F$ s  $<$  1). The analysis also yielded an unexpected main effect of gender,  $F(1, 58) = 4.10, p < .05$  (men recalled more words than women), but no other effects were significant ( $F$ s  $<$  2).

**Equation evaluation.** Accuracy of responses to the equations did not vary as a function of ethnicity, stereotype threat, or the interaction of the two ( $F$ s  $<$  1). There was, however, a significant gender difference in equation accuracy,  $F(1, 58) = 4.97, p < .05$  (men were more accurate than women), that was qualified by an unexpected Gender  $\times$  Stereotype Threat interaction,  $F(1,$

<sup>2</sup> We ran mixed ethnic groups in this study in contrast to homogeneous groups that we had used for Experiment 1 because we were concerned that it would be seen as unusual to have only four Latinos in a session and that this alone might prime ethnicity.

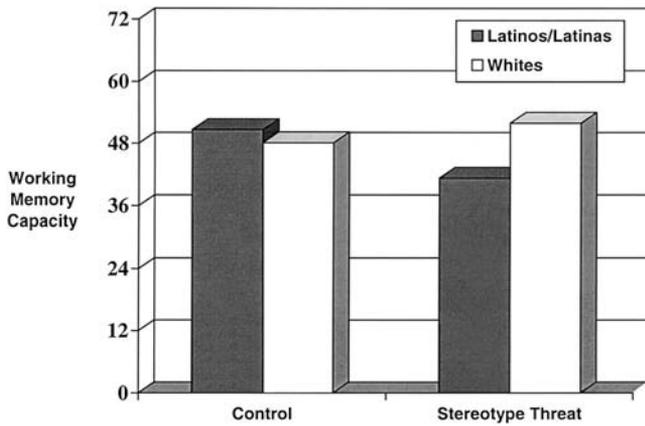


Figure 2. Experiment 2: The effects of ethnicity and stereotype threat on participants' working memory capacity.

58) = 5.21,  $p < .05$ . No other effects were significant ( $F_s < 1$ ). Verbal SAT was a marginally significant covariate,  $F(1, 58) = 3.34, p < .08$ . Simple effects tests of this interaction showed that although men and women were equally accurate in their evaluation of the problems in the control condition (both  $M_s = 81%$ ;  $F < 1$ ), women ( $M = 76%$ ) were less accurate than men ( $M = 86%$ ) in their evaluations of the problems when the test was described as predictive of intelligence,  $F(1, 58) = 11.21, p < .001$ . There were no significant differences in the average time (in seconds) participants spent evaluating each equation ( $GM = 9.02$ ;  $F_s < 2.5$ ).

### Test Experience Questionnaire

**Anxiety.** Analysis of the anxiety composite yielded a marginal Gender  $\times$  Ethnicity interaction,  $F(1, 64) = 3.63, p < .10$ , and a significant Ethnicity  $\times$  Stereotype Threat interaction,  $F(1, 64) = 5.07, p < .05$ . Simple effects analysis indicated that Latinos in the stereotype threat condition reported significantly more anxiety ( $M = 4.05$ ) compared with Latinos in the control condition ( $M = 3.14$ ),  $F(1, 64) = 4.84, p = .05$ , whereas the stereotype threat manipulation had no effect on self-reported anxiety of Whites ( $M = 3.41$  in stereotype threat;  $M = 3.78$  in control;  $F < 1$ ).

**Test difficulty.** Analysis of the perceived difficulty of the test yielded a gender main effect,  $F(1, 64) = 6.82, p < .01$  (women perceived the test to be more difficult,  $M = 4.37$ , than did men,  $M = 3.65$ ), and a marginal Ethnicity  $\times$  Stereotype Threat interaction,  $F(1, 64) = 3.66, p < .06$ . No other effects were significant ( $F_s < 1.5$ ). Simple effects tests revealed that Whites ( $M = 4.19$ ) and Latinos ( $M = 3.99$ ) saw the test as equally difficult in the control condition ( $F < 1$ ), whereas Latinos ( $M = 4.44$ ) rated the test to be more difficult than did Whites ( $M = 3.54$ ) under stereotype threat,  $F(1, 64) = 4.96, p < .05$ .

**Ethnic identity threat.** Analysis of the ethnic identity threat measure yielded only a marginal Ethnicity  $\times$  Stereotype Threat interaction,  $F(1, 64) = 2.69, p = .11$ . Latinos reported slightly more ethnic identity threat in the stereotype threat condition ( $M = 1.91$ ) than in the control condition ( $M = 1.47$ ), whereas Whites reported slightly less ethnic identity threat in the stereotype

threat condition ( $M = 1.47$ ) than in the control condition ( $M = 1.97$ ), although none of the simple effects were significant. No other effects were significant ( $F_s < 1$ ).

### Discussion

The results of the second study provide further evidence for the negative effects of stereotype threat on working memory capacity. When the working memory test was described as a measure related to intelligence, Latinos recalled fewer words compared with Whites and compared with Latinos in the nonthreat control group. These results demonstrate that the working memory task we chose is not uniquely relevant to stereotypes about math performance but can be used to assess the capacity deficits experienced by other stereotyped groups during a cognitively taxing test. This study also provided a stronger test of our hypothesis because, although participants were told that their performance on the test would be used to establish norms for different groups, the issue of group differences in the stereotyped performance domain was never mentioned explicitly.

The less direct nature of the manipulation might explain why there were no significant differences in the self-reported ratings of ethnic identity threat across test description and ethnicity. Both Latinos and Whites indicated on a 7-point scale that they were fairly unconcerned ( $M_s < 2$ ) that the researcher would evaluate their performance in terms of their ethnicity. However, unlike our previous studies focusing on gender stereotypes, this experiment did show anxiety effects that paralleled the effects on working memory capacity. When the working memory test was described as being related to intelligence, Latinos reported feeling more anxious than Whites did. However, further analysis shows that participants' self-reports of anxiety were uncorrelated with their working memory score ( $r = .01$ ), highlighting a disconnection between the processing effects of stereotype threat and the conscious experience of stereotype threat.

### Experiment 3

The studies presented so far provide consistent support for the proposition that stereotype threat impairs performance on challenging cognitive tests by reducing the cognitive resources available to the test taker. However, these studies have only shown that a manipulation of stereotype threat reduces working memory capacity. We know from other research that manipulations of stereotype threat also reduce performance on standardized tests (e.g., Spencer et al., 1999) and that there is a significant relationship between working memory capacity and performance on standardized tests (Turner & Engle, 1989). Thus, the next step was to conduct an experiment designed to test whether working memory capacity mediates the effect of stereotype threat on academic test performance. To test for mediation, we had women complete both the working memory measure and a standardized math test under stereotype threat or nonthreat conditions. We expected that if working memory capacity mediates stereotype threat effects, then scores on the working memory test should account for performance decrements on a standardized math test.

To test our hypothesis more precisely, two other changes were made in this final study. First, we modified our measure of working memory capacity to rule out an alternative explanation for our

results. In the previous studies, the processing component of the working memory task required solving mathematical equations. This task is obviously related to the stereotyped domain in Experiment 1 and also could be perceived as related to the stereotyped domain in Experiment 2 (i.e., intelligence). This situation raises the possibility that when the stereotype was made salient in the performance domain, participants might have assumed that the processing component of the working memory measure was more important for their performance on the test. As a consequence, they might have shifted more of their attentional focus to evaluating the equations rather than remembering and recalling the words. If this were the case, then lowered word recall would occur as a function of shifting focus and not a general reduction in working memory. To address this artifactual explanation, we used a measure of working memory capacity that required a similar amount of processing time but did not involve mathematical equations. In this study, the processing component of the task involved counting the number of vowels in a sentence. We reasoned that if the effect of stereotype threat on word recall is caused by a disproportionate focus on the processing task, then making the task unrelated to the stereotyped domain should reduce any influence of this potential artifact.

In addition to modifying the measure of working memory, we also adopted a different manipulation of stereotype threat. In the previous two experiments, we manipulated stereotype threat by telling participants that the working memory task was in fact measuring, or was highly related to, the relevant stereotyped ability (i.e., math or intelligence). We did this in an effort to assess the effects of stereotype threat online, that is, while participants actually believed they were in the performance situation. However, connecting the stereotype threat manipulation to the working memory test directly allows for the possibility that reduced working memory capacity simply represents another stereotype threat effect and not necessarily a mediating mechanism. Our final experiment was designed to address this conceptual ambiguity. Specifically, we sought to demonstrate that stereotype threat reduces working memory even when the threat is not directly connected to the measure of working memory. To test this possibility, we used a manipulation that would create a general atmosphere of stereotype threat during the session. Specifically, women assigned to the stereotype threat condition participated in a session as the solo woman in a session of other male participants (Inzlicht and Ben-Zeev, 2000; Sekaquaptewa & Thompson, 2003) and were told that they would be taking a math exam later in the session (i.e., after they completed the measure of working memory capacity). Women in the control condition were run in a session of only other women and were told that they would be completing a problem-solving task later in the session. Previous research has shown that priming stereotype threat can reduce performance expectancies (Stangor, Carr, & Kiang, 1998) and can also lead to self-handicapping before the critical performance (Stone, 2002). These results suggest that stereotype threat can influence behavior even in anticipation of the performance situation. Given these past findings, we expected that women would exhibit a reduction in working memory capacity after stereotype threat was primed but before beginning the stereotype-relevant task and that this reduction would mediate the effect of stereotype threat on performance.

## Method

### Participants and Design

The participants were 31 female undergraduates who were randomly assigned to complete two tasks (the working memory test and a standardized math test) under stereotype threat or control conditions. The participants were selected on the basis of the same criteria and method used in Experiment 1 (math SAT  $\geq$  500 and self-reported knowledge of the stereotype) and completed the study in exchange for course credit or \$10. Data from 3 participants in the control condition were lost, 1 because of a computer malfunction and 2 because of a procedural error, leaving a final sample of 28.

### Materials

For this study, we created a version of the working memory test that did not involve mathematical equations. Instead, the processing component of the test required participants to count the number of vowels contained within a given sentence. The sentences were 7–12 words long and contained an average of approximately 10 vowels. After being presented with each sentence, participants were given a word to recall. Furthermore, to minimize fatigue we reduced the total number of equation/word combination trials from 72 to 60. The set sizes for this test were increased to range from four to six equation/word combination trials (four blocks of each set size) to account for any loss of sensitivity caused by reducing the total number of words.

The standardized math test we used consisted of 30 multiple-choice word problems taken from the quantitative section of practice GREs. Because these problems had been included in actual GRE tests, we had data on performance norms as indexed by the proportion of test takers who answered each problem correctly in previous administrations. To ensure that the test was difficult but appropriate for the skill level of our sample, we selected problems that ranged from 44% to 80% accuracy, with a mean accuracy of 64%. The final test was administered to participants in paper-and-pencil format.

### Procedure

*Stereotype threat manipulation.* To prime and maintain the salience of stereotype threat during the completion of the working memory test, we adapted procedures used by Inzlicht and Ben-Zeev (2000; see also Sekaquaptewa & Thompson, 2003) to test the effects of solo gender status on women's math performance. Participants in both conditions completed the study in groups of 3 people. Women in the stereotype threat condition completed a session conducted by a male experimenter with 2 other male confederates acting as participants. Women in the control condition completed the working memory test in a session conducted by a female experimenter along with 2 other female participants.

Upon entering the lab, the participants assembled in a large room where they received a brief overview of the experimental session and signed the consent form. Participants in both conditions were told that they would be completing two completely unrelated tasks. In the stereotype threat condition, the experimenter explained that the main goal of the study was to administer a test of mathematical aptitude to collect normative data on men and women. In the control condition, participants were told that the primary purpose of the study was to administer a problem-solving exercise to collect normative data on college students. After this introduction, the experimenter explained that because of a room scheduling conflict they would first need to complete a brief computer task in another part of the lab. The participants were then escorted to individual rooms equipped with computers where they completed the vowel-counting version of the working memory test. There was no mention of math or math ability in connection with this test.

*Working memory questionnaire.* Once all the participants completed the working memory test, they were given a brief questionnaire designed to get a sense of how they perceived the test. To assess whether participants might have seen the test as a measure linked to math ability, we asked them to rate the extent to which they thought performance on the test would be related to mathematical ability and memory ability on a 7-point scale ranging from 1 (*not at all related*) to 7 (*highly related*). To address the possibility that participants in the stereotype threat condition might have shifted more of their attention to the processing task because it seemed most related to math ability, we asked participants to rate on an 11-point scale how important they thought each part of the task (counting vowels and remembering words) was for determining their overall performance (1 = *counting vowels is most important*, 6 = *words and vowels are equally important*, 11 = *remembering the words is most important*). Participants also indicated how they allocated their attention while completing the task using an 11-point scale ranging from 1 (*counting the vowels*) to 6 (*focused equally on counting the vowels and remembering the words*) to 11 (*remembering the words*). Finally, to assess the possibility that stereotype threat reduces women's expectations for their test performance, participants were asked to rate how well they expected to perform on the upcoming math test/problem-solving exercise using a 9-point scale ranging from 1 (*poor*) to 9 (*excellent*).

*Math test.* Participants were then escorted back to the large room where they were seated individually at small tables and introduced to the math test. In the control condition, the test was described as a pilot test of materials to be used in future research on problem-solving processes. No mention of math was made. In the stereotype threat condition, participants were told the test was a reliable measure of mathematical aptitude. To avoid demand characteristics, there was no explicit mention of gender differences in performance or ability. Participants in both conditions were told that the test would be difficult but within their range of ability. They were asked to give a genuine effort and to avoid random guessing and were told that they would receive feedback about their performance afterwards. Before beginning, women in the stereotype threat condition were asked to indicate their gender and provide their last name on the cover of the test booklet. Participants were given 20 min to work on the test and then completed a brief posttest questionnaire that included the measures of anxiety, perceived difficulty (of the math test), and gender identity threat ( $\alpha = .84$ ) used in Experiments 1 and 2. To further assess expectancies, we also asked participants to rate how they thought the researcher expected men and women to do relative to each other on the math test, using a 7-point scale ranging from 1 (*men will score better than women*) to 4 (*men and women will score the same*) to 7 (*women will score better than men*). After completing the questionnaire, participants were probed for suspicion, thoroughly debriefed, and thanked for their participation.

## Results and Discussion

### Working Memory Capacity

*Absolute span score.* As predicted, women in the stereotype threat condition ( $M = 26.78$ ) recalled fewer words on the working memory test than did women in the control condition ( $M = 38.86$ ),  $t(26) = 3.13$ ,  $p < .01$ ,  $d = 1.19$ .

*Vowel counting.* The accuracy of the number of vowels counted correctly did not vary as a function stereotype threat ( $GM = 87\%$ ;  $t < 1$ ), and there were no significant differences in the average time (in seconds) participants spent counting the vowels ( $GM = 7.60$ ;  $t < 1$ ). It is also important to point out that the average amount of time participants spent counting the vowels was not substantially different than the average amount of time participants spent evaluating the difficult equations in Experiments 1 and 2 ( $M = 8.32$ ).

### Perceptions of the Working Memory Task

Women's perceptions of what performance on the working memory task would be related to (mathematical ability and memory ability) were analyzed with a 2 (stereotype threat)  $\times$  2 (ability type) mixed-factors ANOVA. This analysis revealed only a main effect of ability type,  $F(1, 26) = 58.76$ ,  $p < .001$ . Participants in both conditions indicated that they perceived performance on the working memory test to be more related to memory ability ( $M = 6.14$ ) than mathematical ability ( $M = 3.71$ ).

Additional analyses revealed that participants in the control condition ( $M = 8.20$ ) and the stereotype threat condition ( $M = 7.46$ ) did not differ significantly in their perceptions of which part of the working memory task was more important (vowel counting or word recall;  $p > .20$ ); and participants in both conditions generally viewed word recall to be the more important portion of the test (compared with the midpoint = 6),  $t(26) = 6.31$ ,  $p < .001$ . There were also no significant differences between conditions in what part of the test participants reported focusing on (stereotype threat  $M = 8.00$ ; control  $M = 7.31$ ;  $p > .15$ ), and participants focused more on remembering the words than counting the vowels (compared with the midpoint = 6),  $t(26) = 6.52$ ,  $p < .001$ . Together these results reduce the plausibility of the alternative explanation that lowered word recall in the stereotype threat condition resulted from participants simply shifting their focus to the processing task and away from memorizing and recalling the words.

### Math Test Performance

Because very few participants were able to answer all 30 questions in the time allotted for the test, performance on the math test was analyzed as women's accuracy on the test, that is, the number of problems answered correctly divided by the number of problems attempted (Inzlicht & Ben-Zeev, 2000; Marx & Roman, 2002). Replicating previous stereotype threat research, women in the stereotype threat condition ( $M = 0.49$ ) were less accurate on the math test than women in the control condition ( $M = 0.66$ ),  $t(26) = 2.38$ ,  $p < .05$ ,  $d = .88$ . There were no significant differences in the number of math problems attempted by women in the stereotype threat condition ( $M = 16.15$ ) and women in the control condition ( $M = 15.47$ ;  $t < .50$ ), suggesting that women in both conditions expended comparable effort on the test.

### Test Experience Questionnaire

*Anxiety, perceived difficulty, and gender identity threat.* As in Experiment 1, there were no significant differences in self-reported anxiety between the two conditions ( $t < 1$ ). The average anxiety rating for both conditions was at the midpoint of the scale ( $GM = 4.11$ ). However, whereas stereotype threat had a significant effect on women's difficulty ratings and perceptions of gender identity threat in Experiment 1, the stereotype threat manipulation in Experiment 3 did not affect either their difficulty ratings for the math test ( $GM = 4.93$ ) or their ratings of gender identity threat ( $GM = 2.75$ ; both  $ts < 1$ ). This inconsistency in the conscious experience of stereotype threat across the two studies with women probably reflects the fact that our manipulation of threat was more

explicit in the first experiment in which we explicitly told women that there were gender differences on the test.

*Performance expectancies on the math test/problem-solving exercise.* After the working memory task, we asked participants to rate how well they expected to do on the upcoming task (described as either a math test in the stereotype threat condition or as a problem-solving exercise in the control condition). There were no significant differences in the expectancies of participants in the stereotype threat condition ( $M = 5.62$ ) and participants in the control condition ( $M = 6.00$ ;  $t < 1$ ). After completing the math test, participants also rated how they thought the researcher expected men and women to do relative to each other on the task. Women in the stereotype threat condition ( $M = 3.00$ ) were no more likely to believe that the researcher expected gender differences than were women in the control condition ( $M = 3.47$ ;  $p = .19$ ). Overall, women in both conditions ( $GM = 3.25$ ) tended to believe that the researcher expected men to outperform women (compared with the midpoint = 4),  $t(26) = -4.27$ ,  $p < .001$ .

### Mediational Analyses

To test whether the reductions of working memory mediated the effect of stereotype threat on test performance, we computed a series of regression equations as prescribed by Baron and Kenny (1986). According to this approach, three relationships between the target variables must be demonstrated to establish a basis for testing mediation. The independent variable must predict both the dependent and the mediator variable and the mediator must predict the dependent variable. Once these conditions are established, the dependent variable is regressed onto the independent variable and mediator in a final regression analysis. Support for mediation is obtained by demonstrating that the effect of the independent variable (stereotype threat) on the dependent variable (math test performance) is significantly reduced when accounting for the effect of the hypothesized mediator (working memory capacity). The results of these analyses are shown in Figure 3.

As reported above, stereotype threat had a significant negative effect on women's working memory capacity (the mediator),  $\beta = -.52$ ,  $t(26) = -3.13$ ,  $p < .01$ , and math test performance (the

dependent variable),  $\beta = -.42$ ,  $t(26) = -2.38$ ,  $p < .03$ . A third regression analysis established that working memory capacity was a significant predictor of accuracy on the math test,  $\beta = .64$ ,  $t(26) = 4.28$ ,  $p < .001$ . Finally, when performance on the math test was regressed onto both stereotype threat and working memory capacity, stereotype threat was no longer a significant predictor of math test performance,  $\beta = -.12$ ,  $t(26) = -.66$ ,  $p > .50$ , whereas working memory capacity remained significant in the equation,  $\beta = .58$ ,  $t(26) = 3.26$ ,  $p < .01$ . A Sobel test of the reduction in the direct stereotype threat effect was significant ( $Z = 2.26$ ,  $p < .02$ ), providing support for our hypothesis that stereotype threat interferes with women's math test performance by reducing their working memory capacity.

To rule out other explanations for these relationships (e.g., model misspecification), we also conducted a reverse mediation analysis with math test performance serving as the mediator and working memory serving as the dependent variable. In contrast to the primary mediation analysis, stereotype threat remained a marginally significant predictor of working memory when controlling for math test performance,  $\beta = -.31$ ,  $t(26) = -1.94$ ,  $p = .06$ . A Sobel test confirmed that the reduction in the stereotype threat effect when controlling for performance on the math test was not significant ( $Z = 1.51$ ,  $p > .10$ ). In sum, these analyses provide greater support for the hypothesis that working memory mediates effects on math test performance than for an alternative model in which math test performance mediates stereotype threat effects on working memory capacity.

### General Discussion

Previous stereotype threat research has predominantly examined affective variables that are thought to account for the performance decrements observed by individuals who face the possibility of confirming a negative stereotype (e.g., Spencer et al., 1999). Relatively less work has examined the effects that stereotype threat manipulations might have on cognitive processing per se (see Quinn & Spencer, 2001, for an exception). The experiments reported here were designed to test the hypothesis that stereotype threat reduces an individual's performance on a complex cognitive test because it reduces the individual's working memory capacity. Results of three experiments provide support for this hypothesis. Experiments 1 and 2 demonstrated that manipulations of stereotype threat led to lower working memory scores among individuals who are targeted by the stereotype (women and Latinos) but had no effect on those who are not targeted by the stereotype (men and Whites). Results of the third experiment reveal that the reductions in working memory capacity observed under stereotype threat mediate the reductions in performance on a standardized test. Taken together, these findings suggest that members of stigmatized groups perform poorly on cognitive tests when negative stereotypes have been primed because this added information interferes with their attentional resources.

Although the pattern of results on the primary variable of interest (i.e., working memory capacity) was quite consistent across three studies using different groups, different manipulations, and different measures, the patterns of data on participants' self-reported experiences were more variable. For example, al-

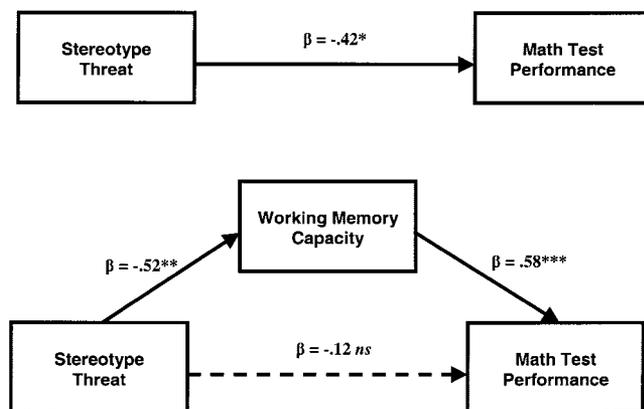


Figure 3. Experiment 3: Tests of working memory capacity as a mediator of stereotype threat effects on women's math test performance. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

though Latinos in Experiment 2 reported feeling more anxious under conditions of stereotype threat, the same pattern was not observed among women in Experiments 1 or 3. Even though Latinos' reports of anxiety paralleled their mean differences for working memory capacity, there was no correlation between these two measures. Similarly, women reported more gender identity threat in the stereotype threat condition in Experiment 1 but not in Experiment 3. The inconsistency of self-reported experiences paired with consistent performance decrements suggests that a conscious awareness of stereotype threat may not be a necessary component of the phenomenon. In practical terms, the apparent dissociation between processing effects and the conscious experience of anxiety or identity threat implies that most individuals who suffer the deleterious effects of stereotype threat are not necessarily aware of their predicament. Although they might not experience a test as more difficult or become more anxious than other nonstigmatized students, the salience of negative stereotypes might nevertheless consume critical cognitive resources and lead to a poor performance that is unrepresentative of their true abilities. Without awareness that a problem exists, individuals under stereotype threat are unable to do anything to circumvent these effects. Of course, this concern highlights a need for future research to investigate the strategies that an individual might use to combat the effects of stereotype threat.

We want to state explicitly that our emphasis on the cognitive deficits associated with stereotype threat is not meant to imply that negative affect does not contribute to the effect of stereotype threat on performance. As Steele et al. (2002) have noted recently, stereotype threat appears to be a multifaceted experience that does not lend itself to one phenomenological explanation. Furthermore, it would be difficult to argue in light of research from both the stereotype threat literature and the working memory literature that anxiety does not play a role in interfering with performance on cognitive tasks. However, we would suggest that future research use more novel approaches to measuring affective processes that do not rely on self-report measures exclusively. Our findings highlight the need to approach the "threat" from a more implicit angle given that stereotype threat does not appear to produce consciously reportable affective states on a consistent basis or even lowered expectations of performance.

The findings from these experiments not only advance our understanding of stereotype threat, but represent an advance for research on working memory capacity as well. As stated earlier, working memory capacity has been examined exclusively as an individual-difference measure that has been assumed to be relatively stable over time. Our demonstration that manipulations of stereotype threat can produce situational reductions in working memory underscores the need for researchers to use caution when assuming that such measures index a stable and unchanging ability. Furthermore, these studies offer an important new methodological tool for measuring the effects of situational manipulations on working memory capacity. Presumably the working memory measures that we used in our studies could be used in other social psychological research areas (e.g., research on persuasion and stereotype application) that posit a role for cognitive capacity in mediating judgments.

### *Processes That Might Deplete Cognitive Resources Under Stereotype Threat*

Although these studies constitute a direct examination of the hypothesis that stereotype threat reduces cognitive capacity, future research will be needed to determine exactly where attentional capacity is diverted when stereotype threat occurs. There are several possible processes that might be examined. One possibility is that cognitive capacity is being consumed by the attempt to suppress negative group stereotypes. Current conceptualizations of working memory capacity assume that this cognitive system is intimately involved in directing attention to the task at hand while suppressing intrusive or irrelevant information (Engle, 2001; Rosen & Engle, 1998). For example, Klein and Boals (2001) posited that individuals high in life stress have lower levels of working memory capacity because they are chronically trying to suppress unwanted negative thoughts and feelings that they have. Thus, one can imagine that once group stereotypes have been activated by the task description in our experiments, some amount of working memory capacity is being devoted to screen out this information during the task.

Consistent with these ideas, recent evidence confirms that constructs associated with female stereotypes are activated among women during stereotype threat and that the accessibility of these constructs partially mediates their lowered performance on a math test (Davies et al., 2002). Furthermore, Spencer (2003) has conducted research concurrent with our own that suggests that women might try to suppress these stereotypic constructs. For example, he has found that adding further cognitive load to women who are already experiencing stereotype threat leads to a heightened activation of stereotype related constructs, supporting the notion that the load interfered with their attempts to suppress this information. He also showed that giving women instruction to replace stereotypic thoughts that they might have during the test with less threatening thoughts eliminates the negative effects of stereotype threat on performance. What is remarkable is that these replacement thoughts can be something self-affirming, such as an important social identity, or something quite irrelevant, such as a red Volkswagen.

Another mechanism to explore in future research is the role of heightened arousal in impairing working memory. Given research showing that participants under stereotype threat show heightened physiological arousal (Blascovich et al., 2001) as well as the finding that stress impairs working memory (e.g., Klein & Boals, 2001), it seems quite reasonable to assume that physiological changes occurring during stressful situations, in general, and under stereotype threat, more specifically, impair cognitive functioning. Still, more work is needed to understand the specific processes by which this happens. In one promising study on this topic, Croizet, Desprès, Gauzins, Hugué, and Leyens (2003) have examined heart-rate variability as a physiological variable that is predictive of cognitive resources (i.e., higher variability indicates higher resources). Their research shows that both stigmatized and nonstigmatized individuals show decreased heart-rate variability when told that a test is a diagnostic measure of their intelligence. However, these changes in heart-rate variability predict and mediate test performance only among stigmatized individuals. Although these results represent an important step forward in linking physiology and cognition, still more research is needed that inte-

grates the affective, cognitive, and physiological processes underlying the experience of stereotype threat.

### Conclusions

The primary purpose of the research reported here is to advance our knowledge of the ways in which negative stereotypes exert their influence on the individuals they target. Following the work of Steele et al., we approached this issue from the perspective that negative social stereotypes can create an added psychological burden in situations where one's behavior might be interpreted as evidence for the validity of such belief systems. Across three studies, we provide converging evidence that performing under the specter of a negative stereotype can deplete the cognitive resources of stigmatized group members and impair performance on challenging academic tasks. These findings represent an important advance in understanding exactly how stereotypes work to undermine the talents and abilities of individuals who might otherwise meet their full potential in high-stakes performance situations, such as college entrance exams. Beyond simply advancing our theoretical knowledge, however, we hope these findings also meet their full potential and contribute to broader efforts aimed at developing strategies to combat the deleterious effects of stereotype threat on performance.

### References

- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35, 29–46.
- Ashcraft, M. H. (2002, October). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, 11, 181–185.
- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, 130, 224–237.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.
- 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.
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, 12, 225–229.
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, 8, 331–335.
- Croizet, J. C., Desprès, G., Gauzins, M., Huguet, P., & Leyens, J. (2003). *Stereotype threat undermines intellectual performance by triggering a disruptive mental load*. Unpublished manuscript, Université Blaise Pascal, Clermont-Ferrand, France.
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardtstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28, 1615–1628.
- Derakshan, N., & Eysenck, M. W. (1998). Working memory capacity in high trait-anxious and repressor groups. *Cognition and Emotion*, 12, 697–713.
- Engle, R. W. (2001). What is working memory capacity? In H. L. Roediger III & J. S. Nairne (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 297–314). Washington, DC: American Psychological Association.
- Engle, R. W., Tuoholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.
- Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion*, 6, 409–434.
- Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, 28, 659–670.
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why women are susceptible to experience problem-solving deficits in the presence of men. *Psychological Science*, 11, 365–371.
- Klein, K., & Boals, A. (2001). The relationship of life event stress and working memory capacity. *Applied Cognitive Psychology*, 15, 565–579.
- Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, and Computers*, 31, 429–432.
- La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 1118–1133.
- Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, 28, 1183–1193.
- Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2002). *Combating stereotype threat: The effects of self-affirmation and emotional expression on women's math performance*. Unpublished manuscript, University of Arizona.
- Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittman, W. W. (2000). Working memory capacity – Facets of a cognitive ability construct. *Personality and Individual Differences*, 29, 1017–1045.
- Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, 57, 55–71.
- Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language*, 39, 418–436.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38, 194–201.
- Schmader, T., & Johns, M. (2003). *Suspicious minds: Do beliefs about gender stereotypes moderate stereotype threat effects?* Unpublished manuscript, University of Arizona.
- Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, 39, 68–74.
- Sorg, B. A., & Whitney, P. (1992). The effect of trait anxiety and situational stress on working memory capacity. *Journal of Research in Personality*, 26, 235–241.
- Spencer, S. (2003, February). *Media images and stereotype threat: How activation of cultural stereotypes can undermine women's math performance*. Paper presented at the annual meeting of the Society for Personality and Social Psychology, Universal City, CA.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28.
- Spielberger, C. D., Gorsuch, R. R., & Lushene, R. (1970). *The State-Trait Anxiety Inventory (STAI) test manual*. Palo Alto, CA: Consulting Psychologists Press.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology*, 75, 1191–1197.

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

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

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

Stone, J. (2002). Battling doubt by avoiding practice: The effects of

stereotype threat on self-handicapping in white athletes. *Personality and Social Psychology Bulletin*, 28, 1667–1678.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, 77, 1213–1227.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.

Received May 21, 2002  
 Revision received May 13, 2003  
 Accepted May 14, 2003 ■



**AMERICAN PSYCHOLOGICAL ASSOCIATION  
 SUBSCRIPTION CLAIMS INFORMATION**

Today's Date: \_\_\_\_\_

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do **NOT** duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION _____		MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL) _____
ADDRESS _____		DATE YOUR ORDER WAS MAILED (OR PHONED) _____
CITY _____ STATE/COUNTRY _____ ZIP _____		<input type="checkbox"/> PREPAID <input type="checkbox"/> CHECK <input type="checkbox"/> CHARGE CHECK/CARD CLEARED DATE: _____
YOUR NAME AND PHONE NUMBER _____		(If possible, send a copy, front and back, of your cancelled check to help us in our research of your claim.) ISSUES: <input type="checkbox"/> MISSING <input type="checkbox"/> DAMAGED
<b>TITLE</b> _____	<b>VOLUME OR YEAR</b> _____	<b>NUMBER OR MONTH</b> _____
_____	_____	_____
_____	_____	_____

*Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4–6 weeks.*

(TO BE FILLED OUT BY APA STAFF)

DATE RECEIVED: _____	DATE OF ACTION: _____
ACTION TAKEN: _____	INV. NO. & DATE: _____
STAFF NAME: _____	LABEL NO. & DATE: _____

Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242

**PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.**

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.