Pers Soc Psychol Bull OnlineFirst, published on May 6, 2009 as doi:10.1177/0146167209335165

The Effect of Stereotype Threat on the Solving of Quantitative GRE Problems: A Mere Effort Interpretation

Jeremy P. Jamieson
Stephen G. Harkins
Northeastern University

The mere effort account argues that stereotype threat motivates participants to want to perform well, which potentiates prepotent responses. If the prepotent response is correct, performance is facilitated. If incorrect and participants do not know, or lack the knowledge or time required for correction, performance is debilitated. The Graduate Record Examinations (GRE) quantitative test is made up of two problem types: (a) solve problems, which require the solution of an equation, and (b) comparison problems, which require the use of logic and estimation. Previous research shows that the prepotent tendency is to attempt to solve the equations. Consistent with mere effort predictions, Experiment 1 demonstrates that threatened participants perform better than controls on solve problems (prepotent response correct) but worse than controls on comparison problems (prepotent response incorrect). Experiment 2 shows that a simple instruction as to the correct solution approach eliminates the performance deficit on comparison problems.

Keywords: *stereotype threat; motivation; mere effort; math; GRE*

Stereotype threat refers to the concern that is experienced when one feels "at risk of confirming, as self-characteristic, a negative stereotype about one's group" (Steele & Aronson, 1995, p. 797). A wide range of stereotypes have been tested (e.g., women's lack of ability in math and science: O'Brien & Crandall, 2003; African Americans' underperformance on standardized tests: Steele & Aronson, 1995; White males' athletic inferiority: Stone, 2002). In each case, concern about confirming the relevant negative stereotype has been shown to negatively affect the performance of the stigmatized individuals.

Steele, Spencer, and Aronson (2002) have argued that these effects "are likely to be mediated in multiple ways—cognitively, affectively, and motivationally" (p. 397). Consistent with this view, Schmader, Johns, and Forbes (2008) have identified working memory "as a core cognitive faculty that is implicated in cognitive and social stereotype threat effects" (p. 337). Research has also demonstrated the role of affective processes (e.g., anxiety; Bosson, Haymovitz, & Pinel 2004). In the current work, we focus on the contribution of motivation to threat effects, specifically on Jamieson and Harkins's (2007) mere effort account.

The mere effort account was suggested by Harkins's (2006) analysis of the effect of evaluation on performance. This account argues that evaluation motivates participants to want to perform well, which potentiates whatever response is prepotent, or most likely to be produced in a given situation. If the prepotent response is correct, the potential for evaluation facilitates performance. If the prepotent response is incorrect and participants do not know, or lack the knowledge or time required for correction, performance is debilitated. However, if participants are able to recognize that their prepotent tendencies are incorrect and are given the opportunity to correct, performance will be facilitated. Harkins and his colleagues have found support for the mere effort account of the effect of evaluation on performance on the Remote Associates Task (Harkins, 2006), anagrams (McFall, Jamieson, & Harkins, 2009, Experiment 1), the Stroop task (McFall et al., 2009, Experiments 2 and 3), and the antisaccade task (McFall et al., 2009, Experiment 4).

Authors' Note: Correspondence concerning this article should be addressed to Jeremy P. Jamieson, Department of Psychology, 125 Nightingale Hall, Northeastern University, 360 Huntington Avenue, Boston, MA 02115-5000; e-mail: jamieson.jp@gmail.com.

PSPB, Vol. XX No. X, Month XXXX xx-xx
DOI: 10.1177/0146167209335165

Jamieson and Harkins (2007) argued that stereotype threat, like the potential for evaluation, can motivate participants, which may bring into play the same process that Harkins and his colleagues (Harkins, 2006; McFall et al., 2009) implicate in the evaluation–performance relationship. Of course, other stereotype threat research has also argued that the participants subject to threat are motivated to perform well. For instance, Steele and Aronson (1995) suggest that the performance of threat participants may suffer because they alternate their attention between trying to answer the items and trying to assess the self-significance of their frustration. That is, according to this account, they are indeed motivated to solve the problems. However, it is not this motivation in and of itself that affects their performance but rather the time spent assessing their frustration. In contrast, the mere effort account argues that motivation plays a direct role in producing the effects of threat on performance.

Mere effort shares the notion that stereotype threat energizes prepotent or dominant responses with the arousal/drive explanations proposed by O'Brien and Crandall (2003) and Ben-Zeev, Fein, and Inzlicht (2005). For example, O'Brien and Crandall argue that the arousal produced by stereotype threat "is non-specific and serves to energize behavior in a nondirective way. For this reason, arousal enhances the emission of dominant responses" (p. 783). However, as is also the case for the arousal/drive explanations for social facilitation (e.g., Cottrell, 1972), these accounts do not incorporate the correction process proposed by Harkins and his colleagues (Harkins, 2006; McFall et al., 2009).

In addition, it should be noted that the mere effort account does not argue that stigmatized individuals are motivated to seek out situations to demonstrate their proficiency in stereotyped domains. In fact, research has shown that participants subject to threat take advantage of explanations that allow them to deflect responsibility for their performance (e.g., Ben-Zeev et al., 2005; Johns, Schmader, & Martens, 2005; Keller, 2002; Steele & Aronson, 1995; Stone, 2002). Consistent with this view, Harkins and his colleagues have demonstrated that although participants are more motivated to perform well when subject to experimenter evaluation than when they are not (e.g., Harkins & Szymanski, 1988; McFall et al., 2009; White, Kjelgaard, & Harkins, 1995), when given a choice, participants avoid subjecting themselves to the scrutiny of the experimenter (Szymanski & Harkins, 1993). So, the mere effort account applies to situations that do not provide stigmatized individuals with the opportunity to avoid judgment in the context of the stereotype.

Jamieson and Harkins (2007) tested the mere effort account under such circumstances using an inhibition task, the antisaccade task (Hallett, 1978). This task requires participants to respond to a target presented randomly on one or the other side of the display. Before the target appears, a cue is presented on the opposite side of the display. Participants are explicitly instructed to not look at this cue but rather to look to the opposite side of the display where the target will appear. However, a reflexive-like prepotent tendency to look at the cue must be inhibited or corrected to optimize performance.

The mere effort account predicts that because the motivation potentiates the prepotent response, participants under stereotype threat will look the wrong direction, toward the cue, more often than controls. However, because this response is obviously incorrect and participants have the time required for correction, those subject to threat would also be motivated to launch correct saccades (saccades launched toward the target following successful inhibition of the reflexive saccade) and corrective saccades (saccades launched toward the target following reflexive saccades) faster than controls. Once the participants' eyes arrive at the target area, they must then determine the target's orientation and press the appropriate response key. The motivation to perform well would make stereotype threat participants try to respond as quickly as possible. Thus, when the participants see the target, participants subject to threat should respond more quickly than controls. Jamieson and Harkins (2007, Experiment 3) found support for each of these predictions, which taken together, produced faster overall reaction times for participants subject to stereotype threat than for no threat participants with no difference in accuracy.[1]

Jamieson and Harkins's (2007) research supports the mere effort account of performance on the antisaccade task, an inhibition task framed as a measure of visuospatial capacity. However, most stereotype threat research that has focused on the female math-ability stereotype has used problems taken from standardized tests, such as the quantitative section of the Graduate Record Examinations (GRE) general test (e.g., Ben-Zeev et al., 2005; Inzlicht & Ben-Zeev, 2003; Schmader & Johns, 2003; Spencer, Steele, & Quinn, 1999). On this task, like the Remote Associates Test (RAT) and anagrams but unlike inhibition tasks, such as the Stroop and the antisaccade tasks, the correct answer is not apparent. The present research tests the contribution of mere effort to performance on this task.

The GRE quantitative test is primarily made up of two types of problems: solve problems and comparison problems, which differ in the solution approach that tends to be most efficient (see the appendix for examples). Solve problems are standard word problems that tend to be most efficiently solved by using an equation or algorithm. Comparison problems require the test taker to compare the quantity in one column with the quantity in the other. In this case, the most efficient approach tends to be simplifying the terms in each column or using logic, estimation, and/or intuition to find the correct answer.[2]

To test the mere effort account on GRE items, we must first identify the prepotent response. For example, in the case of the RAT, our own research (Harkins, 2006) identified the prepotent response and then demonstrated the effect of evaluation on this response. For anagrams (McFall et al., 2009, Experiment 1), previous research had identified the prepotent response, the tendency to try consonants in the first position of the word (e.g., Witte & Freund, 2001). The Stroop task (McFall et al., 2009, Experiments 2 and 3) and the antisaccade task (Jamieson & Harkins, 2007; McFall et al., 2009, Experiment 4) are inhibition tasks, and as such, the prepotent response (reading the word and looking at the peripheral cue, respectively) is well established.

In the case of quantitative GRE problems, previous research suggests that participants' prepotent tendency is to take a conventional or solving approach (i.e., compute the answer using a rule or an equation). For example, in their research on the performance of high school students on math problems taken from the SAT, Gallagher et al. (2000, Experiment 2) found that participants used a conventional (i.e., solving) approach 55.5% of the time, whereas they used the unconventional (i.e., comparison) approach 10% of the time, with the remainder of the trials comprising guesses (16.5%), omissions (9%), and unknowns (9%). Other research that has examined participants' performance on the same problems used by Gallagher et al. has found the same pattern of results: Participants use the solving approach significantly more often than any other solution approach. For example, Gallagher and De Lisi (1994) found that their participants used the solving approach on 63% of the problems as opposed to the comparison method (32%) or guessing (5%), and Quinn and Spencer (2001, Experiment 2) found that their participants relied on the solving approach on 57% of the problems, as opposed to the comparison approach (28%), guessing (8%), or an unknown approach (7%). Furthermore, this pattern held true whether the participants were working on solve or comparison problems. For instance, Gallagher et al. (2000, Experiment 2) found that on solve problems, participants used the solving approach 66% of the time and the comparison approach 9% of the time. On the comparison problems, participants used the solving approach 45% of the time versus 11% for the comparison approach.

Taken together, these findings strongly suggest that when participants are asked to solve math problems, the prepotent tendency is to attempt to compute the answer using a known formula or an algorithm (i.e., the solving approach), which is consistent with the training in mathematics that they receive. For example, Katz, Bennett, and Berger (2000) point out, "Traditional strategies [the solving approach] are the formal methods traditionally emphasized in U.S. mathematics education" (p. 41). Likewise, Stigler and Hiebert (1999) note that in the United States, "teachers present definitions of terms and demonstrate procedures for solving specific problems. Students are then asked to memorize the definitions and practice the procedures" (p. 27).

Given this prepotent response, mere effort predictions on the GRE quantitative test are straightforward. On solve problems, the prepotent tendency is often the most efficient approach. As long as the test taker knows the correct equation to apply, solving the equation will lead to a correct answer. The stereotype threat manipulation should motivate the females to do well, which should facilitate performance on these problems. However, on comparison problems the prepotent tendency to use the solving approach will be much less efficient and often may not work at all. It is also unlikely that participants will recognize that they are not using the best approach. Thus, on these problems, the potentiation of the prepotent response, solving equations, should debilitate the performance of females working under stereotype threat.

If stereotype threat facilitates performance on solve problems, but debilitates it on comparison problems, how can mere effort account for the finding that stereotype threat debilitates overall performance (e.g., Brown & Pinel, 2003; O'Brien & Crandall, 2003)? In fact, our analysis suggests that the effect of the debilitation of performance on the comparison problems should be more profound than the effect of facilitation on the solve problems. We argue that stereotype threat potentiates the prepotent response, solving equations. Because this response is prepotent, everyone will tend to take the correct approach to solve problems, but females subject to stereotype threat will be more motivated than no threat females to perform the computations. In addition, the females in the stereotype threat and control groups should not differ in their basic knowledge of the relevant formulas or operations. Thus, the effect of the heightened motivation of females subject to stereotype threat is limited to the fact that they will try harder than controls to solve as many of these problems as they can. And, of course, the time limit on the test provides an upper bound for this effort.

In contrast, on the comparison problems, the potentiation of the solving approach will make it highly unlikely that the threatened females will ever even adopt the correct approach to the problem. As a result, they will be stymied from the outset. So, even though they have the same knowledge as females in the control group, this knowledge will be much less likely to be brought to bear on the problems. And the more motivated they are, the worse off they will be as they continue to attempt to use the solving approach on comparison problems (cf. producing close associates on RAT problems; Harkins, 2006). Consequently, stereotype threat females should perform better on solve problems, more poorly on comparison problems, and more poorly overall than females in a control group. This hypothesis was tested in Experiment 1.

EXPERIMENT 1

The GRE test in Experiment 1 included comparison and solve problems. Because mere effort predicts that stereotype threat will potentiate the prepotent response (Jamieson & Harkins, 2007), threatened females are predicted to perform more poorly on comparison problems and better on solve problems than females not subject to stereotype threat. We also predicted that overall performance on the test would be worse for females subject to stereotype threat than females in the control group, because the debilitation of performance on the comparison problems will be greater than the facilitation of performance on the solve problems. The performance of males should not differ as a function of stereotype threat.

Method

Participants

Sixty-four Northeastern University undergraduates (32 male, 32 female) participated in this experiment in exchange for course credit.

Materials

The math test consisted of 30 multiple-choice problems taken from the quantitative section of the GRE. The test included 15 comparison problems and 15 solve problems, and it was presented as a paper-and-pencil test with scratch paper provided for showing work.

Because each problem actually appeared on a GRE test, we obtained performance norms, as indexed by the proportion of test takers answering each problem correctly out of all those who attempted that problem. We selected problems by first randomly picking 12 problems of each type from problems that varied in their solution rates from 35% to 65%. We then picked the final three problems for each type so that mean overall accuracy averaged 50% for each problem set (comparison range = 38% to 60%, solve range = 42% to 63%).

Problems were randomized throughout the test with the constraint that no one type of problem could appear in more than three consecutive problems. This procedure produced a test on which comparison and solve problems were dispersed evenly throughout the test. Of the first 16 problems, 8 were comparison and 8 were solve problems. Participants worked on the test for 20 min and were instructed to complete as many problems as accurately as possible. All participants were given 2 practice problems (1 comparison, 1 solve) before beginning and were instructed not to use calculators.

Procedure

Threat manipulations were adapted from Jamieson and Harkins (2007). In the threat condition participants were instructed they would be taking a math test that had been shown to produce gender differences. In the no threat conditions, participants were also told they were taking a math test but that the test had been shown not to produce gender differences. This explicit stereotype threat manipulation has been shown to produce performance effects in previous research (e.g., Brown & Pinel, 2003; Jamieson & Harkins, 2007; Keller, 2002; Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Spencer et al., 1999). No specific mention was made as to whether men outperformed women or vice versa, only that gender differences did or did not exist on the task. Participants were expected to infer that women would perform more poorly than men based on the societal stereotype that men are superior to women in mathematical ability.

Each participant responded to a questionnaire upon completion of the math test. Two questions allowed us to evaluate the effectiveness of the stereotype threat manipulation: "To what extent are there gender differences in performance on this task?" (1 = *no gender differences*, 11 = *gender differences*) and "Who do you believe performs better on this task?" (1 = *males perform better*, 6 = *males and females perform the same*, 11 = *females perform better*). Participants were also asked to rate how difficult the test was, how interesting the task was, how anxious they felt about their performance, and how much effort they put into the task, all on 11-point scales.

Results

Manipulation Check for Stereotype Threat

The manipulation checks were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (gender: male vs. female) ANOVAs. Condition and gender were both analyzed as between-subject factors. Participants in the stereotype threat condition reported that gender differences existed to a greater extent ($M = 6.63$, $SD = 2.94$) than did participants in the no stereotype

TABLE 1: Experiment 1: Graduate Record Examinations (GRE) Performance as a Function of Gender,Problem Type, and Stereotype Threat Condition

	Performance Variable							
	No. Attempted Comparison		*No. Attempted Solve*		*% Correct Comparison*		*% Correct Solve*	
Condition	M	SD	M	SD	M	SD	M	SD
NST male	8.94	2.98	9.06	2.74	54.14	14.54	53.83	19.14
ST male	9.00	2.25	8.13	2.13	54.92	17.23	57.08	15.72
NST female	8.63	3.03	7.81	2.43	54.33	14.06	54.56	16.23
ST female	8.00	2.83	9.44	1.83	27.07	14.00	55.34	20.55

NOTE: ST = stereotype threat; NST = no stereotype threat.

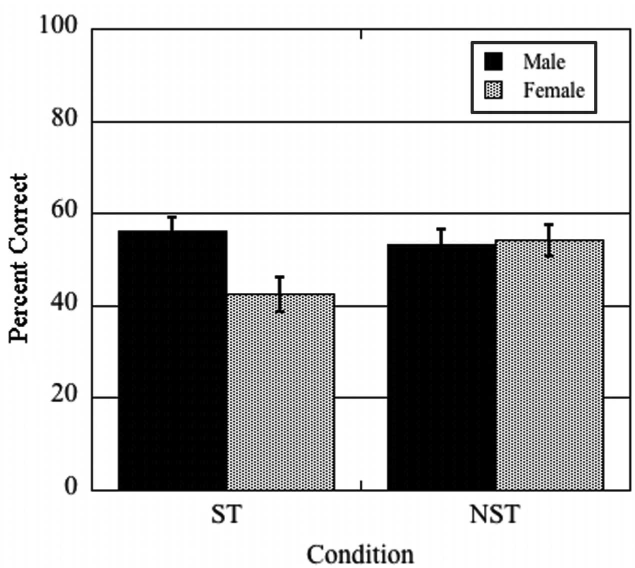

Figure 1 Experiment 1: Total percentage correct as a function of gender and stereotype threat condition.

NOTE: GRE = Graduate Record Examinations; ST = stereotype threat; NST = no stereotype threat.

threat condition (M = 4.03, SD = 2.40), F(1, 60) = 15.58, p < .001, d = 1.02. Participants in the stereotype threat condition also reported that men performed these tasks better than women to a greater extent (M = 3.91, SD = 1.38) than did participants in the control condition (M = 5.44, SD = 1.48), F(1, 60) = 17.91, p < .001, d = 1.09. These results indicate that the stereotype threat manipulation used in the current experiment was successful. Participants in the threat condition were aware of the negative group stereotype, and women were expected to perform more poorly than men.

Math Test Performance

To test for the traditional stereotype threat effect on GRE test performance (e.g., Schmader & Johns, 2003, Experiment 3), we analyzed the total percentage of problems solved (total number solved correctly/total attempted) in a 2 (condition: stereotype threat vs. no stereotype threat) × 2 (gender: male vs. female) ANOVA. Replicating previous research, we found a Stereotype Threat × Gender interaction, F(1, 60) = 4.76, p = .03, d = .56. Pairwise contrasts (Kirk, 1995) revealed that female participants in the stereotype threat condition performed more poorly (M = 42.44%, SD = 15.06%) than females in the no stereotype threat control group (M = 54.16%, SD = 13.85%), F(1, 60) = 6.18, p < .03, d = .64, whereas males did not differ as a function of stereotype threat, p > .50 (see Figure 1). Thus, we were successful in replicating previous findings in this domain.

We also tested for differences in the total number of problems attempted in the same 2 × 2 design and found none, ps > .20. Overall, participants attempted 17.25 problems.

To test for the effects suggested by the mere effort hypothesis, we examined the performance of participants on solve and comparison problems on two measures: the number of problems attempted and the percentage of these problems that were correctly answered. These data were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (gender: male vs. female) × 2 (problem type: comparison vs. solve) ANOVAs with condition and gender as between-subjects factors and problem type as a within-subjects factor (see Table 1 for means and standard deviations). Pairwise contrasts (Kirk, 1995) were used to decompose significant interactions.

Problems attempted. Analysis of the number of problems attempted revealed a significant three-way interaction among gender, stereotype threat, and problem type, F(1, 60) = 12.94, p = .001, d = .93. Females subject to stereotype threat attempted more solve problems (M = 9.44, SD = 1.83) than did females who were not subject to threat (M = 7.81, SD = 2.43), F(1, 60) = 13.25, p < .01, d = .94. In contrast, males in the stereotype threat condition attempted fewer solve problems

($M = 8.13$, $SD = 2.13$) than did males in the no stereotype threat condition ($M = 9.06$, $SD = 2.74$), $F(1, 60) = 4.24$, $p < .05$, $d = .53$.

Neither females nor males differed as a function of stereotype threat in the number of comparison problems attempted, ps > .20. All means and standard deviations are presented in Table 1.

Percentage solved. Analysis of the percentage of problems solved also yielded a significant three-way interaction among stereotype threat, gender, and problem type, $F(1, 60) = 6.28$, $p < .02$, $d = .65$. On the solve problems, there were no differences in the percentage of problems solved by females or males, ps > .20.

However, females in the stereotype threat condition solved a smaller percentage of the comparison problems they attempted ($M = 27.07\%$, $SD = 14.00\%$) than did females in the no stereotype threat control condition ($M = 54.33$ %, $SD = 14.06\%$), $F(1, 60) = 26.45$, $p < .001$, $d = 1.33$. Males did not differ as a function of stereotype threat, $p > .20$. All means and standard deviations are presented in Table 1.

Ancillary measures. Participants did not differ in their ratings of how interesting the task was, how difficult the task was, how much effort they put into the task, or how anxious they felt about their performance, ps > .20.

Discussion

In Experiment 1 we found that overall GRE math performance for females subject to stereotype threat was worse than for females who were not subject to threat. However, we also found that females subject to stereotype threat attempted more solve problems than females under no threat but solved the same percentage of these problems correctly. Thus, threatened females outperformed no threat females (i.e., correctly solved more solve problems). In contrast, on comparison problems, females under threat attempted the same number of problems as control females but solved a smaller percentage of these problems correctly. Therefore, on comparison problems, females under threat performed more poorly than controls.

Consistent with the mere effort account, when the prepotent, conventional approach was correct, stereotype threat participants performed better than controls, whereas when this response was incorrect, their performance was debilitated. Using methods suggested by Meng, Rosenthal, and Rubin (1992) for comparing correlated correlation coefficients, we compared the magnitude of the effect of stereotype threat on comparison problems with the magnitude of the effect of threat on solve problems, ignoring the difference in sign. As predicted, the magnitude of the debilitation effect on the number of comparison problems solved was significantly greater ($r = -.59$) than the magnitude of the facilitation effect on the number of solve problems solved ($r = .21$), $Z = 3.08$, $p < .001$, accounting for the overall finding of poorer performance on the part of females subject to stereotype threat than females not subject to threat.[3]

EXPERIMENT 2

Experiment 1 suggests that stereotype threat motivates participants to want to do well, which potentiates whatever response is prepotent on the given task. However, motivation should also lead to an effort to correct incorrect responses if the participant recognizes that his or her response is incorrect, knows the correct response, and has the opportunity to make it. On inhibition tasks, it is obvious to participants when prepotent responses are incorrect, as is the way in which they can be corrected. Thus, given sufficient time, more motivated participants are able to make the correction and respond more quickly than less motivated participants (see Jamieson & Harkins, 2007).

In contrast, in Harkins's (2006) research on the RAT, regardless of how motivated the participants might have been, they are unlikely to have been aware that their prepotent tendency (generating close associates for triad members) was incorrect. In a pilot study, Harkins (2006) tested the efficacy of a simple instruction on how to approach the task. All participants were told that their performance would be subject to evaluation. One third of these participants were told that if they wanted to succeed, they should refrain from generating close associates. Instead they were to simply register the triad members and then wait for the answer to "pop up." Another third were told that if they wanted to succeed, they should generate as many close associates as possible. The final third, a control condition, were told nothing. Harkins found that participants who were told not to generate close associates but to wait for the answer to emerge outperformed the other two groups, which did not differ from each other.

In Experiment 2, we took a similar approach to improving the performance of females subject to stereotype threat by informing participants as to the correct solution approach to take to each type of problem. We hypothesized that stereotype threat participants' debilitated performance on comparison problems results from reliance on their prepotent tendency to use a solving approach, which is incorrect for these problems. The instructions not only make it more likely that participants will recognize when their solution approach to a problem is incorrect but also suggests the correct approach. If stereotype threat participants are able to reduce their reliance on the solving approach on comparison problems, we would expect that their performance would improve compared to stereotype threat participants not given instructions. On the solve problems, however, threat participants should still outperform controls because the prepotent response, which is potentiated by threat, is correct and the instructions should not reduce their motivation to perform well.

Previous research has also examined intervention strategies for reducing the effect of threat on females' math performance. For example, Johns et al. (2005) taught females about stereotype threat, informing them that any anxiety that they were feeling could be the result of "negative stereotypes that are widely known in society and have nothing to do with your actual ability to do well on the test" (p. 176). They found that these females performed better than females not given this information and argued that teaching threat enabled females to attribute any anxiety felt during the math test to gender stereotypes. The intervention strategy tested in the current research takes a different approach, by providing participants with information that allows them to recognize the behavior that produces the performance debilitation (reliance on the solving approach).

Method

Participants

Sixty-four Northeastern University female undergraduates participated in this experiment in exchange for course credit. Only females were used in this research because the stereotype threat manipulation did not significantly affect the performance of males in Experiment 1.

Materials

All materials were identical to those described in Experiment 1.

Procedure

Manipulations and questionnaires were similar to those described in Experiment 1 except that participants in the instruction condition received instructions before beginning the practice problems. Instructions were adapted from a GRE preparation book (Lurie, Pecsenye, Robinson, & Ragsdale, 2004). Participants receiving instructions were told the following:

> This test consists of two types of problems, comparison and solve problems. You may be familiar with the solve problem format. For these problems, you are given a problem, offered five solutions, and asked to pick one. These problems generally ask you to apply an equation that you have previously learned or to use mathematical rules to work through an equation to obtain a solution. For example, you may be given the dimensions of a geometrical figure and asked to obtain the area. To solve a problem such as this you just need to apply the area formula.
>
> The remaining problems consist of comparison problems, which require you to compare the quantity in Column A to the quantity in Column B. It is important to remember that you do not have to calculate the exact values in each column to solve the problem. After all, your goal is simply to compare the two columns. It's often helpful to treat the two columns as if they were two sides of an equation. Anything you do to both sides of an equation, you can also do to the expressions in both columns. Try planning your approach and simplifying each column rather than calculating.

Participants not assigned to the instruction condition were not given these instructions before working on the practice problems, and their procedure was identical to that described in Experiment 1.

In Experiment 2, participants were again provided with scratch paper. We argue that the prepotent response on the GRE problems is to use a solving approach, which is potentiated by stereotype threat, debilitating performance on comparison problems. Thus, the threat participants should show evidence of using the solving approach more on comparison problems than should control participants. The effect of the instruction should be to reduce the threat participants' reliance on this approach. To test these predictions, a rater blind to condition computed the percentage of problems on which the participants' scratch paper showed evidence that they used the solving approach, regardless of whether the attempt produced the correct answer.[4]

Results

Manipulation Checks

Manipulation checks were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (instruction: instructions given vs. no instructions given) ANOVAs. Both condition and instruction were analyzed as between-subjects factors. Stereotype threat participants rated the test as revealing gender differences to a greater extent (M = 6.47, SD = 2.85) than did no stereotype threat participants (M = 3.56, SD = 2.56), $F(1, 60) = 18.35$, $p < .001$, $d = 1.11$. Participants in the stereotype threat condition also indicated that males performed better than females on this test to a greater extent (M = 4.00, SD = 1.83) than controls (M = 5.30, SD = 1.14), $F(1, 60) = 11.55$, $p = .001$, $d = .88$.

Math Performance

To test for the traditional stereotype threat effect, we analyzed the total percentage solved correctly (total

TABLE 2: Experiment 2: Graduate Record Examinations (GRE) Performance as a Function of Instruction, Problem Type, and Stereotype Threat Condition

	Performance Variable							
	No. Attempted Comparison		*No. Attempted Solve*		*% Correct Comparison*		*% Correct Solve*	
Condition	M	SD	M	SD	M	SD	M	SD
NST no instruction	8.75	2.82	7.94	2.44	51.39	11.43	47.10	20.75
ST no instruction	8.25	2.24	9.00	1.51	23.49	14.48	53.09	17.37
NST instruction	9.25	1.53	8.56	1.55	46.71	12.73	51.80	19.18
ST instruction	8.63	2.33	9.81	2.32	44.77	5.64	59.01	17.37

NOTE: ST = stereotype threat; NST = no stereotype threat.

number correct/total attempted) in a one-way (stereotype threat/no instructions vs. no stereotype threat/no instruction) ANOVA. As in Experiment 1, stereotype threat participants solved a smaller percentage of problems (M = 39.23%, SD = 9.25%) than did no stereotype threat participants (M = 49.53%, SD = 3.09%), $F(1, 60) = 7.12$, $p < .02$, $d = .97$. We also tested for differences in the total number of problems attempted in the same one-way design and found none, $F < 1$. Overall, participants attempted 16.97 problems.

Problems attempted and percentage solved were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (instruction: instructions given vs. no instructions given) × 2 (problem type: comparison vs. solve) ANOVAs. Condition and instruction were analyzed as between-subjects factors whereas problem type was analyzed as a within-subjects factor. Pairwise contrasts (Kirk, 1995) were used to decompose significant interactions. All means and standard deviations are presented in Table 2.

Problems attempted. Analysis of the problems attempted revealed a Stereotype Threat × Problem Type interaction, $F(1, 60) = 37.66$, $p < .0001$, $d = 1.58$. Participants in the stereotype threat condition attempted more solve problems (M = 9.41, SD = 1.97) than did participants in the no threat condition (M = 8.25, SD = 2.03), $F(1,60) = 29.74$, $p < .0001$, $d = 1.41$, whereas participants in the no stereotype threat condition attempted more comparison problems (M = 9.00, SD = 2.24) than did participants in the stereotype threat condition (M = 8.44, SD = 2.26), $F(1,60) = 8.03$, $p < .02$, $d = 1.03$. No other effects were reliable.

Percentage solved. Analysis of the percentage of problems solved correctly yielded a reliable Stereotype Threat × Instruction × Problem Type interaction, $F(1, 60) = 6.18$, $p < .02$, $d = .64$. As in Experiment 1, when not given instructions, participants in the stereotype threat condition did not differ from no stereotype threat participants in the percentage of solve problems solved ($M_{stereotype\ threat}$ = 53.09%, SD = 17.37%; $M_{no\ stereotype\ threat}$ = 47.10%, SD = 20.75%), $p > .20$. Replicating Experiment 1, under no instructions, participants in the stereotype threat condition solved a smaller percentage of the comparison problems (M = 23.49%, SD = 14.48%) than did participants in the no stereotype threat condition (M = 51.39%, SD = 11.43%), $F(1, 60) = 31.40$, $p < .001$, $d = 1.45$.

In Experiment 2, we were interested in whether describing the appropriate approach to take to each type of problem would be sufficient to improve threat participants' performance on comparison problems, and this is exactly what we found. In fact, when instructions were provided, stereotype threat participants solved the same percentage of comparison problems (M = 44.77%, SD = 5.64%) as participants not subject to threat (M = 46.71%, SD = 12.73%), $p > .20$. There was also a tendency for stereotype threat/instruction participants to correctly answer a higher percentage of solve problems (M = 59.01%, SD = 17.37%) than no threat/instruction participants (M = 51.80%, SD = 19.18%), $F(1, 60) = 2.10$, $p = .15$, $d = .37$.

The overall analysis also produced a significant instruction main effect, $F(1, 60) = 6.19$, $p < .05$, $d = .64$; a Stereotype Threat × Instruction interaction, $F(1, 60) = 6.17$, $p < .02$, $d = .64$; a problem type main effect, $F(1,60) = 20.11$, $p < .001$, $d = 1.16$; and a Stereotype Threat × Problem Type interaction, $F(1, 60) = 18.69$, $p < .001$, $d = 1.12$, each of which must be interpreted in the context of the three-way interaction.

Solution approach analysis. The percentage of trials on which participants showed that they used the prepotent, solving approach was analyzed in a 2 (condition: threat vs. no threat) × 2 (instruction: instructions given vs. no instructions given) × 2 (problem type: comparison vs. solve) ANOVA. Condition and instruction were analyzed as between-subjects factors whereas problem type was analyzed as a within-subjects factor. Pairwise contrasts were used to decompose interactions.

TABLE 3: Experiment 2: Percentage of Problems on Which the Solving Approach Was Used as a Function of Instruction, Problem Type, and Stereotype Threat Condition

	Solving Approach (%)			
	Comparison Problems		*Solve Problems*	
Condition	M	SD	M	SD
NST no instruction	48.12	14.90	68.52	15.80
ST no instruction	70.24	19.39	73.59	17.59
NST instruction	47.20	15.63	66.15	18.72
ST instruction	44.67	20.78	64.19	13.97

NOTE: ST = stereotype threat; NST = no stereotype threat.

Overall, participants used the solving approach on 65% of the problems, consistent with the argument that this approach is prepotent. The analysis produced a significant Condition × Instruction × Problem Type interaction, $F(1, 60) = 3.74$, $p = .05$, $d = .50$. The main effects for problem type, $F(1, 60) = 46.59$, $p < .001$, $d = 1.76$; instruction, $F(1, 60) = 6.84$, $p = .01$, $d = .67$; and the Condition × Instruction interaction, $F(1, 60) = 4.68$, $p = .03$, $d = .56$, must be interpreted in the context of the three-way interaction. See Table 3 for all means and standard deviations.

When not given instruction, threat participants used the solving approach significantly more often on comparison problems ($M = 70.24\%$, $SD = 19.39\%$) than did no threat participants ($M = 48.12\%$, $SD = 9.98\%$), $F(1, 60) = 31.46$, $p < .001$, $d = 1.45$. This analysis also showed that threat participants significantly reduced their use of the solving approach on comparison problems when given instruction ($M = 44.67\%$, $SD = 20.78\%$), $F(1, 60) = 31.48$, $p < .01$, $d = 1.45$. In fact, there was no difference between threat and no threat participants in the use of the solving approach on comparison problems when instructions were given ($M = 47.20\%$, $SD = 15.63\%$), $F < 1$.

Mediation. We argue that use of the solving approach on comparison problems produces the performance debilitation observed under threat. To test this argument, we conducted a mediation analysis following the procedures suggested by Kenny, Kashy, and Bolger (1998) on the number of comparison problems solved incorrectly. As shown in Figure 2, the use of the prepotent, solving approach mediated the effect of stereotype threat on comparison problem performance, Sobel $Z = 2.07$, $p = .039$.[5]

Ancillary Measures

Participants subject to stereotype threat reported that the test was more difficult ($M = 7.91$, $SD = 1.75$) than did controls ($M = 6.97$, $SD = 1.79$), $F(1, 60) = 4.46$, $p = .04$, $d = .55$. There were no significant differences in ratings of how interesting the task was, how much effort was put into the task, performance expectations, or anxiety, *p*s > .20.

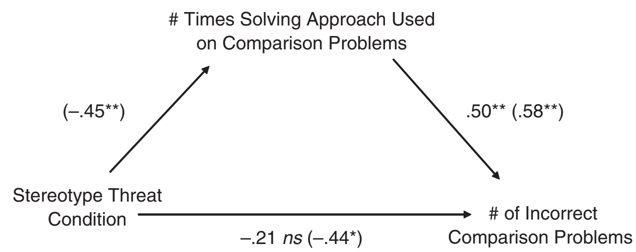

Figure 2 Number of times the prepotent, solving approach was used on comparison trials as a mediator of number of comparison problems answered incorrectly in Experiment 2.

NOTE: Coefficients in parentheses indicate zero-order correlations. Coefficients not in parentheses represent parameter estimates for a recursive path model including both predictors. Stereotype threat condition is dummy coded (1 = *stereotype threat*, 2 = *no stereotype threat*).
$^{*}p < .05$. $^{**}p < .01$.

Discussion

When not provided with instruction, the overall GRE math performance for females under threat was worse than that of females who were not subject to threat. As in Experiment 1, when not given instruction, threat participants outperformed no threat participants on solve problems by attempting more problems with no cost in accuracy, whereas they performed more poorly than no threat participants on comparison problems. Consistent with the mere effort account, the solution approach analysis showed that threat participants used the prepotent, solving approach significantly more often than no threat participants on these problems, which debilitated performance (see Figure 2). As in Experiment 1, we used the methods suggested by Meng, Rosenthal, and Rubin (1992) to compare the magnitude of the effect of stereotype threat on comparison problems with the magnitude of the effect of threat on solve problems. Replicating the previous finding, we found that the magnitude of the debilitation effect on the number of comparison problems solved was significantly greater ($r = –.70$) than the magnitude of the facilitation effect on the number of solve problems solved ($r = .35$), $Z = 3.41$, $p < .001$, accounting for the overall poorer performance on the part of females subject to stereotype threat than females not subject to threat.

We attempted to improve the performance of threat participants by providing instructions designed to reduce

TABLE 4: Experiment 2: Total Percentage Correct as a Function of Instruction and Stereotype Threat Condition

	Percentage Correct			
	No Instruction		*Instruction*	
Condition	M	SD	M	SD
NST	49.53	12.37	49.03	11.92
ST	38.23	9.25	52.07	10.48

NOTE: ST = stereotype threat; NST = no stereotype threat.

their reliance on the solving approach on comparison problems. This manipulation had a significant impact on performance, as the difference between threat and no threat participants in the percentage of comparison problems solved was eliminated. The solution approach data suggest that threat participants' improvement on comparison problems resulted from the fact that females under threat took the instructions into account and reduced the extent to which they relied on the solving approach when they tried to solve comparison problems.

In the current research, although providing instructions did improve the performance of females subject to threat on comparison problems, this improvement was only sufficient to bring their overall performance to parity with the performance of females not under threat. As shown in Table 4, when instructions were given, overall, females under threat performed as well as females not subject to threat, $F < 1$. If females subject to threat are motivated to do well and have been given instructions as to how to improve their performance, one might argue that their performance should exceed that of females not under threat, who are less motivated. However, females under threat must still cope with the potentiation of the prepotent response, which works against the instructions, and this could limit the amount of improvement that is possible.

Johns et al. (2005) argue that their teaching intervention externalizes arousal, which improves performance. In that research the experimenter made it clear that the negative stereotypes about math performance have nothing to do with the participants' ability. Under these circumstances, participants may not have felt the need to attempt to disconfirm a stereotype that the experimenter has already discounted. In contrast, the instruction manipulation used in Experiment 2 does nothing to reduce participants' motivation to disconfirm the stereotype but rather shows them how to apply their motivation more effectively. Consistent with this argument, we find that even with instructions, threat participants outperform control participants on solve problems, suggesting that they are more motivated than these participants.

GENERAL DISCUSSION

The mere effort account (Jamieson & Harkins, 2007) argues that stereotype threat motivates participants to perform well, which has direct consequences for performance. Previous support for the mere effort account of stereotype threat effects was found using the antisaccade task. However, most previous stereotype threat research on the gender/math stereotype has examined performance on problems from standardized tests such as the GRE and SAT, not on inhibition tasks such as the antisaccade task. In the current research, we extended the mere effort analysis to performance on GRE problems.

Previous research suggests that when faced with quantitative problems, participants' prepotent tendency is to take a solving approach. That is, participants attempt to employ previously learned equations, rules, and algorithms to solve the problems presented to them. Gallagher and De Lisi (1994) found that their participants used this approach on 63% of the problems as opposed to some other method, Gallagher et al. (2000, Experiment 2) found that 56% did so, and Quinn and Spencer (2001, Experiment 2) found that 57% did so. These values compare favorably with our own findings from Experiment 2, which showed that when not given instruction, participants used the solving approach on 65% of the problems.

The GRE quantitative test consists of two general types of problems: comparison and solve, which differ in the solution approach that tends to be most efficient. On comparison problems, the most efficient approach to problem solution tends to be simplifying the terms in each column and/or using logic, estimation, and/or intuition to find the correct answer. On solve problems, the most efficient approach tends to be representing the problem in an equation and solving it.

In Experiment 1 we found that female participants subject to stereotype threat performed more poorly overall than participants in the other conditions, who did not differ among themselves. However, when we looked at performance as a function of problem type, consistent with mere effort predictions, females subject to stereotype threat performed more poorly on comparison problems but better on the solve problems than control participants.

In Experiment 2, we replicated the finding that threat impaired overall performance on the GRE problems by facilitating performance on solve problems but severely debilitating it on comparison problems. Analysis of the participants' solution approaches showed that females subject to threat applied the incorrect, but prepotent, solving approach on more comparison problems than did the no threat participants, and a Sobel test showed that this use of the solving approach mediated the effect of stereotype threat on comparison problem performance. However, when instructions outlining the appropriate approach were provided, females subject to threat did not differ from controls in their use of the solving approach and performed as well as these participants.

Although the current findings are consistent with a mere effort interpretation of stereotype threat performance effects, it could be argued that other findings in the threat literature are inconsistent with this account. For example, if females are motivated by stereotype threat, one may expect to find that they spend more time working on the problems and/or attempt more problems than females not subject to threat. Yet, no consistent effects have been reported on these measures. For example, Steele and Aronson (1995, Experiment 2) found that threat participants spent more time per problem than controls, whereas Spencer et al. (1999, Experiment 1) found no differences on this measure. Steele and Aronson (1995, Experiment 4) found that threat participants attempted fewer problems than control participants, whereas Keller (2002) found no difference on this measure. However, there is no evidence to suggest that spending more time per problem or attempting more problems represents the prepotent response on this task, whereas there is evidence that the participants' prepotent response when faced with these problems is to attempt to solve the equations rather than simplifying terms or using logic, estimation, or intuition to find the correct answer (e.g., Gallagher et al., 2000).

A motivational explanation may also seem to be at odds with research that has found that stereotype threat affects performance but not self-reported effort (e.g., Brown & Pinel, 2003; Keller, 2002; Steele & Aronson, 1995). However, it is not at all unusual to find a lack of correspondence between measures of self-reported effort and measures of performance. For example, in a meta-analysis of social loafing research, Karau and Williams (1993) found that the social loafing effect was robust, but the average effect size for self-reported effort in these experiments was not significantly different from zero. Apparently, participants are unwilling or unable to acknowledge the fact that they loaf. In any event, the focus of the mere effort account is on task performance, not on self-reported effort, and we make no claims concerning the latter.

In addition, recent work by Stone and McWhinnie (2008) suggests that although motivational effects may be produced by blatant manipulations of stereotype threat (e.g., participants are told that there are gender differences on the task), subtle manipulations of threat may work through a different mechanism. That is, when the task is explicitly framed as measuring an attribute that relates to a negative ingroup stereotype, Stone and McWhinnie argue that targets attempt to "minimize mistakes and avoid failure that would confirm the negative stereotype" (p. 446). We suggest that the effects of this motivation to avoid failure can be seen in research that has shown that participants subject to threat take advantage of explanations that allow them to deflect responsibility for their performance (e.g., Keller, 2002; Steele & Aronson, 1995). In the current work, we also used a blatant manipulation of threat, but the stigmatized participants were not provided with the opportunity to avoid judgment in the context of the stereotype. It is under these conditions that we argue that threat participants are motivated to perform well, producing the effects predicted by mere effort.

On the other hand, Stone and McWhinnie (2008) argue that subtle manipulations of threat (e.g., solo status) may lead participants to expend cognitive and emotional resources on reducing uncertainty about the presence of bias, resulting in a working memory deficit and debilitated performance. Thus, the performance effects reported in research that has used subtle manipulations of threat (e.g., Croizet et al., 2004; Inzlicht & Ben-Zeev, 2003; Schmader & Johns, 2003) may be produced not by motivation but by a cognitive mechanism (i.e., working memory deficits). If there is a difference in the effects produced by these types of manipulations, a replication of the current research with a subtle manipulation of threat could produce a different pattern of performance than the one produced by blatant threat. Instead of performing better on solve problems and worse on comparison problems compared to controls, as found with a blatant threat manipulation, females subject to subtle threat may perform more poorly on both types of problems, as their ability to solve either type of problem could be diminished by deficits in working memory capacity.

Individual differences can also limit the scope of motivational explanations, such as the mere effort account. Stereotype threat is experienced when stigmatized individuals are concerned with confirming negative group stereotypes. There can be no performance concerns if an individual does not view performance in the stereotyped domain as important. Consistent with this view, research has indicated that to be threatened "a person probably needs to either care about having the ability or at least care about the social consequences of being seen as lacking the ability" (Aronson, Lustina, Good, & Keough, 1999). Thus, an individual not identified with the stereotyped domain will not exhibit motivated behavior, and the mere effort account does not make predictions regarding performance.

It must also be acknowledged that to make predictions based on the mere effort account, one must be able to identify a priori the prepotent response on the particular task under consideration. In some cases, prepotent responses will have been identified by previous research, as was the case for the quantitative GRE problems used in the current research. However, in other cases, preliminary work will be required for this purpose, as was the case for the RAT (Harkins, 2006). Even when the prepotent response for a given task has been identified and the findings for a given population of participants offer support for the mere effort account, caution must be exercised in generalizing the findings. For example, as noted previously, American students are typically taught to use the solving approach rather than logic and estimation in solving math problems (e.g., Katz et al., 2000; Stigler & Hiebert, 1999). However, students have to learn this approach; therefore, we would not expect novices to display this tendency.

In addition, as students take more advanced courses, their approaches to solving math problems may change. For example, Spencer et al. (1999, Experiment 1) examined the effect of threat on the math performance of students majoring in math and physics and found that these students exhibited no differences on quantitative GRE problems as a function of threat. However, Spencer et al. did find that stereotype threat impaired math and physics majors' performance on problems taken from the GRE Math subject test, which involves advanced calculus and knowledge of abstract algebra and real variable theory. Given these highly selected students' experience and success in solving math problems with logic and estimation, it is unlikely that they would try to use the solving approach on abstract algebra problems. Thus, to generalize the mere effort account to highly selected populations, it would be necessary to first identify what approaches these students take to solving GRE Math problems.

Despite these caveats, we would argue that the perspective provided by the mere effort account contributes to our understanding of the effects of stereotype threat by suggesting another route through which motivation can affect performance. Finally, reducing the negative effects of stereotypes on the test performance of stigmatized group members is one goal of stereotype threat research and the current research suggests another means of achieving this goal. Our work suggests that performance can be improved not only by breaking the link between stereotypes and performance (e.g., Johns et al., 2005) but also by assisting stigmatized individuals in channeling their efforts to best take advantage of their heightened motivational state.

APPENDIX

Examples of the problem types found on the quantitative GRE test. These problems appeared in the math tests used in this research.

Solve Type:

If the total surface area of a cube is 24, what is the volume of the cube?

a. 8
b. 24
c. 64
d. $48\sqrt{6}$
e. 216

For this problem, the individual must apply the formula for the volume of a cube, which is: length × width × height (all of which are the same value for a cube). To get the length of a side, the individual divides 24 by 6 (there are 6 faces on a cube) to obtain the area of one face, 4. The length of one side is 2 (area = length × width). To compute volume, the test-taker then cubes 2 to get the answer, 8. Thus, solve problems involve the application and computation of equations.

Comparison Type:

$$n = (7)(19^3)$$

Column A	Column B
The number of distinct positive factors of n	10

a. The quantity in Column A is greater
b. The quantity in Column B is greater
c. The two quantities are equal
d. The relationship cannot be determined from the information given

This problem can be solved by using intuition. First, the test taker must realize that each number presented (7, 19, 3) is a prime number. Thus, the test taker can logically deduce that the factors of the end product can only be multiples of 7 and 19. Thus, the factors of the final product are: 7, $7*19$, $7*19^2$, 19, 19^2, 19^3, plus the final product itself ($7*19^3$) and 1. Because the goal of the problem is not to compute the value of n but simply to determine whether the number of positive factors of n is greater than, less than, or equal to 10, all the test taker now needs to do is to add up the number of distinct positive factors (8) to find that Column B is greater than Column A. Thus, the correct answer choice is *b*, and only intuition and logic were used. No calculations were necessary.

NOTES

1. McFall, Jamieson, and Harkins (2009) found that the potential for evaluation produced the same pattern of results on the antisaccade task (Experiment 4).

2. In some cases, it may be possible to correctly answer solve problems by using logic and estimation and to solve comparison problems by solving the equations. In fact, on some problems, the alternative approach is even more efficient. However, in the majority of cases, it is more efficient to solve the equations on solving problems and to use logic and estimation on comparison problems. For example, of the 30 problems that were used in the current research, it would have been possible to use logic and estimation on 5 of the 15 solve problems and to solve the equations on 6 of the 15 comparison problems. However, solving the equations was the most efficient approach for 14 of the 15 solve problems, and using logic and estimation was the most efficient approach for 14 of the 15 comparison problems.

In addition to comparison and solve problems, the GRE quantitative test includes chart problems. These problems present test takers with graphs, tables, and/or figures and ask questions regarding the information presented. GRE preparation books (see Lurie, Pecsenye, Robinson, & Ragsdale, 2004) note that chart problems focus on percentages and basic arithmetic processes. Thus, these problems generally do not require the knowledge of geometry, trigonometry, or algebra that is tested by the other problem types. Moreover, only 10% to 16% of the problems on quantitative GRE tests are chart problems.

3. Number correct was used in this analysis because it represents a single measure that allows a direct comparison of the relative strength of the debilitation and facilitation effects. Percentage solved could not serve this function because the facilitation of the threat participants' performance on solve problems was reflected in the fact that they attempted more problems than no threat participants but solved them at the same rate as these participants.

4. This measure may underestimate participants' use of the solving approach. That is, participants could have used the solving approach but made the calculations in their head rather than committing their work to paper. However, this underestimation should not affect group-level comparisons. There is no reason to believe that participants would be more or less likely to compute an answer in their head as a function of either stereotype threat or instruction. We did not estimate participants' use of the comparison approach from the scratch-paper analysis because in the great majority of cases, the use of logic and estimation does not require computation.

5. Number of times participants used the solving approach, rather than percentage, was used in the mediation analysis because the former provides a more valid indicator of the role that the solving approach played in comparison problem performance. Percentage assumes that whenever participants did not provide written evidence of using the solving approach, they were not using this method, even though they may have been doing so in their heads. On the other hand, although number could underestimate the number of times the solving approach was used, it cannot misclassify problems as nonsolving when in fact they are solving.

In addition, we predicted the number of incorrect problems rather than number correct because our hypothesis is that using the solving approach on comparison problems directly leads to poor performance. However, not using the solving approach is only one requirement for success. To solve the problem correctly participants must also recognize that the comparison approach should be used and implement it successfully. Thus, the most direct test of our hypothesis requires prediction of incorrect responses.

REFERENCES

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

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

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus nonverbal anxiety. *Journal of Experimental Social Psychology, 40,* 247–255.

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

Cottrell, N. B. (1972). Social facilitation. In N. B. McClintock (Ed.), *Experimental social psychology* (pp. 185-236). New York: Holt, Rinehart, & Winston.

Croizet, J. C., Desprès, G., Gauzins, M. E., Huguet, P., Leyens, J. P., & Mèot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30,* 721-731.

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

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

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18,* 1279-1296.

Harkins, S. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91,* 436-455.

Harkins, S., & Szymanski, K. (1988). Social loafing and self-evaluation with an objective standard. *Journal of Experimental Social Psychology, 24,* 354-365.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus nonverbal anxiety. *Journal of Experimental Social Psychology, 40,* 247–255.

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

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

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16,* 175-179.

Karau, S. J., & Williams, S. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65,* 681-706

Katz, I. R., Bennett, R. E., & Berger, A. E. (2000). Effects of response format on difficulty of SAT-Mathematics items: It's not the strategy. *Journal of Educational Measurement, 37,* 39-57.

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 47,* 193-198.

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29,* 371-381.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 233-265). New York: McGraw-Hill.

Kirk, R. (1995). *Experimental design.* Pacific Grove, CA: Brooks/Cole Publishing Company.

Lurie, K., Pecsenye, M., Robinson, A., & Ragsdale, D. (2004). *Cracking the GRE.* New York: Princeton Review.

McFall, S., Jamieson, J. P., & Harkins, S. (2009). Testing the generalizability of the mere effort account of the evaluation–performance relationship. *Journal of Personality and Social Psychology, 96,* 135-154.

Meng, X. I., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111,* 172-175.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29,* 782-789.

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.
Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85,* 440-452.
Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115,* 336-356.
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.
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 *Advances in Experimental Social Psychology,* ed. MP Zanna, vol. 34, pp. 379–440. San Diego, CA: Academic.
Stigler, J. W., & Hiebert, J. (1999). *The teaching gap.* New York: Simon & Schuster.
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., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology, 44,* 445-452.
Szymanski, K., & Harkins, S. (1993). The effect of experimenter evaluation on self-evaluation within the social loafing paradigm. *Journal of Experimental Social Psychology, 29,* 268-286.
White, P., Kjelgaard, M., & Harkins, S. (1995). Testing the contribution of self-evaluation to goal-setting effects. *Journal of Personality and Social Psychology, 69,* 69-79.
Witte, K., & Freund, J. (2001). Single-letter retrieval cues for anagram solution. *Journal of General Psychology, 128,* 315-328.

Received December 8, 2008
Revision accepted March 6, 2009