

Do High-Achieving Female Students Underperform in Private? The Implications of Threatening Environments on Intellectual Processing

Michael Inzlicht
New York University

Talia Ben-Zeev
San Francisco State University

Stereotype threat research has demonstrated that stereotypes can harm student performance in the face of public evaluation by peers or an experimenter. The current study examined whether stereotypes can also threaten in private settings. Female students completed a math test in 3-person groups, which consisted of either 2 other women (same gender) or 2 men (minority). In addition, students either believed their performance would be broadcasted to their peers (public) or not (private). Results revealed that minority students performed worse than same-gender students in both public and private environments. This finding supports the concept of threatening intellectual environments and shows how far reaching the effects of stereotypes can be. The authors discuss these findings in relation to research on tokenism and to stereotype threat and its educational implications.

Despite small advances in recent years, education statistics continue to paint a discouraging portrait of underrepresentation for women in math and science. For instance, women comprise only a small minority of high school students, college students, and researchers in the physical sciences: Although they account for well over half of the student body, they comprise less than a third of all bachelor's degree recipients in the physical and computer sciences and less than a quarter of all graduate degree recipients in engineering and mathematics (National Science Foundation, 2000). Although the situation is better in high school—and has improved dramatically over the past few decades—girls comprise 44% of students in advanced math classes in the United States (12th grade) and only about 33% in many other nations (see Geary, 1996, for a review; Travers & Westbury, 1989). In 1982, this underrepresentation prompted the U.S. Congress to mandate the National Science Foundation to provide data on the participation of women in science and engineering biennially. As noted by numerous researchers and commentators, this underrepresentation of female students reflects a number of barriers to their achievement (e.g., Steele, 1997). But to what extent does this underrepresentation actually contribute to further underperformance and underrepresentation? Can a factor as subtle and seemingly benign as the gender composition of a classroom impair the academic outcomes of the underrepresented?

Michael Inzlicht, Department of Applied Psychology, New York University; Talia Ben-Zeev, Department of Psychology, San Francisco State University.

We thank Joachim Krueger, Jack Wright, Joshua Aronson, Catherine Good, Linda McKay, and Steve Fein for their invaluable insights; Melanie Fernandez, Deborah Silverman, and Praveen Basaviah for their data collection assistance; and Naomi Sarah Ball for her guidance on all things metaphysical, emotional, and downright silly.

Correspondence concerning this article should be addressed to Michael Inzlicht, Department of Applied Psychology, Steinhart School of Education, New York University, 239 Greene Street, Room 541, New York, New York 10003. E-mail: michael.inzlicht@nyu.edu

Research on solo status (e.g., Fuegen & Biernat, 2002; Sekaquptewa & Thompson, 2003) and tokenism (e.g., Kanter, 1977; Lord & Saenz, 1985) has found that people are quite sensitive to subtle variations in the group composition of a room, which often operates to the detriment of those in the minority. Being in the minority attracts a disproportionate amount of attention from others and increases feelings of responsibility for representing one's group (Saenz, 1994). For students, this reality can ultimately result in lower grades and achievement.

Not all minority situations were created equal, though. Field and laboratory studies have found that minority situations may be especially troubling to stigmatized groups, such as females. For instance, women who anticipated being tokens—the only woman in an otherwise homogeneous male group—were more likely than nontoken women to want to change groups and to expect being stereotyped by others. In contrast, male tokens didn't differ from male nontokens in these measures (Cohen & Swim, 1995). Similarly, although businesswomen (Kanter, 1977) and policewomen (Ott, 1989) suffer as a result of their low representation, male nurses, librarians, and elementary school teachers do not (Williams, 1992).

Minority situations may be particularly problematic to females because they operate under the burden of negative stereotypes. In math and science classrooms, for example, females have to contend with stereotypes alleging inferiority to males (Quinn & Spencer, 2001; Spencer, Steele, & Quinn, 1999). For females in math classrooms, then, being in the minority may constitute a threatening intellectual environment. For example, Inzlicht and Ben-Zeev (2000) argued that any factor capable of activating negative group stereotypes could affect intellectual performance. Thus, because being outnumbered by males can increase awareness of one's gender group (Cota & Dion, 1986; McGuire & Padawer-Singer, 1976) and of the negative stereotypes associated with one's gender group (Devine, 1989), they reasoned that it could also affect intellectual performance. They showed that the numerical representation of one's gender group could interact with one's stereotyped status to determine whether an environment would either

promote or frustrate academic learning, engagement, and performance. In their studies, when outnumbered by men, college women performed worse on a math test as compared with women who were tested in all-women settings.

The nature of these threatening intellectual environments became more clear by the fact that under the same conditions (i.e., being outnumbered by men), women's performance in a nonstereotyped domain (a verbal test) was unaffected. Similarly, when men took a math test while being outnumbered by women (Inzlicht & Ben-Zeev, 2000, Experiment 2), their test performance was the same as when they took the test in the presence of other men. These findings suggest that the situation was threatening to women because group composition made the negative stereotype about their math ability salient. That is, group composition may have activated the stereotype automatically—without direct presentation (Bargh, 1997). The implication is that women's underperformance in math domains may be exacerbated by their underrepresentation.

Being in the minority, then, can induce stereotype threat among stigmatized groups. The theory of *stereotype threat* (Aronson, 2002; Ben-Zeev, Fein, & Inzlicht, 2003; Spencer et al., 1999; Steele, 1997; Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002) attributes underperformance to worries about being viewed through the lens of negative stereotypes. For example, women who are confronted with negative stereotypes about their group's math ability may develop a fear that their own behaviors may substantiate the negative stereotypes about women's math abilities in the eyes of others (Aronson et al., 1999). This situational predicament can interfere with intellectual performance and lead female students to perform below their potential.¹ Thus, if teachers do not take steps to buffer the threatening climate that can arise from minority situations, classrooms may become intimidating places (Aronson, 2002).

Ironically, stereotype threat is most keenly felt by those females who care most about doing well in math—the math identified (Aronson et al., 1999). When confronted with a stereotype that challenges a valued domain, they become critically unnerved and tend to underperform (Aronson, 2002). Thus, even though high-achieving females may be expected to do poorly in the stereotyped domain, their math identification makes the idea of performing poorly unpalatable. In contrast, females who do not care about math may not be as bothered by negative stereotypes in the math domain (Aronson et al., 1999).

Classrooms: Public and Private Environments

Classrooms are not uniform settings. Rather, thousands of fleeting events combine to form the classroom environment (Jackson, 1968). These events include answering problems at the board, raising one's hand to respond to a question, or participating in group activity, among others. Although most, if not all, of the experiences a student has in the classroom are shared with others or occur in the presence of others, classroom events range in their degree of privacy. Some events, like answering a question in front of the class, are highly public—they embody immediate scrutiny from everyone in the classroom. Yet other events, like writing a term paper, are relatively private affairs because students can rest assured that their grades will be kept confidential from their peers.

Although Steele (1997; Steele & Aronson, 1995) has acknowledged that it may be possible for stereotype threat to operate in

both public and private settings, the existing literature has tended to emphasize the public source of stereotype threat over any private ones. For example, in the Inzlicht and Ben-Zeev (2000) study, female students operated under the assumption that their group members would publicly evaluate their math performance at the end of the testing session. That is, the testing environment was highly public. Other studies have shown similar underperformance but were examined in situations where students expected their scores to be publicly evaluated by either the experimenter or by other students (e.g., Aronson et al., 1999; R. P. Brown & Josephs, 1999; Steele & Aronson, 1995; Stone, Lynch, Sjomeling, & Darley, 1999). For example, in the studies by Spencer et al. (1999) and by Quinn and Spencer (2001), female students were exposed to stereotype threat and then took a math test (individually on a computer) that was to be evaluated by the experimenter. The fact that the experimenter was to evaluate the students' test raises the possibility that their underperformance was a result of desiring to disconfirm the stereotype in the eyes of the experimenter rather than in their own eyes.

Similarly, the debilitating effects of being in the minority—although theoretically possible in private situations—have only been documented with public performances. Can minority situations, then, affect private performances? In addition to the practical educational implications, this question examines the process of minority underperformance by asking whether it is caused by being outnumbered by an evaluative audience, or by the mere presence of nonevaluative others. Sekaquaptewa and Thompson (2003), for example, showed that solo status can affect performance on a public oral test and speculated that it would be ineffective on private performances. However, there were no data to support this assumption. As with stereotype threat, then, the question of minority underperformance in private domains remains largely unanswered.

In this study, we examine this very issue by asking whether it is possible for outnumbered individuals to underperform in a private setting, in which the only evaluator is oneself. We examine whether stereotyped individuals—specifically, female undergraduate students who are highly identified with success in mathematics—will continue to experience minority math situations as threatening even when they are cloaked from worries associated with being publicly evaluated. In other words, we ask whether minority situations and stereotypes are threatening because of a desire to be evaluated positively by others or by a desire to meet one's own high standards.

¹ We should note, however, that girls do not underperform in all math settings. Although there is a marked gender difference favoring boys in standardized math tests (see Hyde, Fennema, & Lamon, 1990, for a review), there is a small gender difference favoring girls in classroom grades (see Kimball, 1989, for a review). Although discrepant, we note that classroom grades and test performance measure different things. Whereas achievement tests only measure situational performance, classroom grades confound conduct and demeanor (Benbow & Stanley, 1982), teachers' shifting standards for girls and boys (Biernat & Manis, 1994), as well as actual performance. Stereotype threat, however, is most readily observed on high-stakes, timed tests, where stress or distraction can take a toll (Aronson, 2002). Girls can compensate for this stress in classroom situations by working harder on homework or doing extra credit. Thus, even though stereotypes may threaten girls' test performance, this may not always be reflected in classroom grades.

Self-Presentation and Private Self-Evaluation

In public, threats may stem from self-presentational concerns or the desire to have oneself and one's group look good to others (Steele, 1997). After all, self-presentational motives arise in situations where people seek approval and respect and avoid disapproval and disrespect (Jones & Pittman, 1982; Tetlock & Manstead, 1985)—precisely the conditions that arise in threatening intellectual environments. However, the desire to make a good impression on others and present oneself and one's group in a positive light may not be the only cause of threat. Instead, as Steele and Aronson (1995) have argued, private self-evaluation may also be capable of inducing it. That is, underperformance may also be triggered by the anxiety stemming from the desire to hold oneself in high regard. Thus, in addition to being elicited by the desire to disconfirm negative stereotype in the eyes of others, minority threat may be elicited through intrapsychic means—through the desire to disconfirm the stereotype in one's own eyes (see also Aronson et al., 1999).

The difference between public and private threat is not a trivial one. It can help us better understand the mechanisms of tokenism and stereotype threat—whether they operate using a self-presentational and/or an intrapsychic pathway, determine whether minority environments threaten because of evaluative audiences or because of the mere presence of others, and delineate the different types of learning environments that are susceptible to stereotype effects. Furthermore, there is a strong tradition of disentangling impression management versus intrapsychic explanations in psychological research (cf. Tetlock & Manstead, 1985) ranging from social facilitation (Zajonc, 1965) to self-handicapping (Berglas & Jones, 1978).

For example, researchers have asked whether students engage in self-handicapping behavior—such as staying out late the night before an important high school test—to protect their images in the eyes of others (self-presentation) or to deceive themselves (intrapsychic). Kolditz and Arkin (1982), for instance, addressed this question by varying whether the experimenter would know of the student's choice to self-handicap. Similarly researchers have asked whether social facilitation—the enhancing or debilitating effects of groups—is caused by the presence of an evaluative audience (self-presentation) or by the mere presence of nonevaluative others (intrapsychic). Markus (1978), for example, examined this question by manipulating whether undergraduate students thought the audience could evaluate them. Thus, examining whether threatening environments can undermine students' performance in private settings has the potential to shed light on a number of social phenomena related to intellectual processing.

Although the effects of threatening environments on private performances have not been explicitly investigated, some studies might be interpreted in light of our private evaluation hypothesis. In one study, Spencer et al. (1999) showed that reminding female undergraduate students of the negative stereotype about their group's math ability could undermine their math test performance. Moreover, they showed that the deficits in math performance were exhibited in women who had both high and low levels of apprehension about public evaluation. Finding that public apprehension does not affect performance suggests that it did not mediate stereotype threat and indirectly supports the hypothesis that threat can be felt in private. In another study, Schmader (2002) showed that stereotypes could threaten performance, but only to the extent

that one suspected the stereotype to be true. Finding that one's internal belief systems could moderate threat suggests that threat has elements of private self-evaluation. Although these results are consistent with the private evaluation hypothesis, we wished to conduct a more focused test of how being outnumbered could threaten people's private performances. Thus, the current study was designed to test whether outnumbered individuals would show intellectual deficits in both public and private testing conditions.

Experimental Paradigm and Hypotheses

In the present study, female undergraduate students took a difficult math test in groups of 3—either with 2 other women (same-gender condition) or with 2 men (minority condition). Before taking the math test, students were either told that at the end of the test they would publicly broadcast their scores to one another (public evaluation) or that their scores would remain anonymous and confidential by letting no one, including the experimenter, have personal access to their scores (private evaluation). If environments can only threaten public performances, then underperformance would only occur in the public minority condition but not in the private minority one. If environments can threaten both public and private performances equally, then there would be support for the idea that self-evaluation is just as important as self-presentation in eliciting stereotype threat.

Method

Participants and Design

Participants were female undergraduate students at Brown University. One-hundred fifteen potential participants completed the Mathematics Identification Questionnaire (MIQ; R. P. Brown, 2000) and indicated their SAT Math scores in an introductory psychology class at the beginning of the semester. Using a 9-point scale ranging from 1 (*strongly disagree*) to 9 (*strongly agree*), students responded to 13 questions concerning the importance they placed and the identification they had with their mathematical abilities, such as "Doing well on math-related tasks is important to me" and "Being good at math is not an important part of who I am." Brown and his colleagues (R. P. Brown, 2000; R. P. Brown & Josephs, 1999) reported that the MIQ demonstrates high interitem reliability ($\alpha = .82$). Furthermore, they reported that it has construct validity, as demonstrated by its relation to other ratings (made at test time), of the importance of performing well on similar math tests.

Research has shown that stereotype threat is most harmful to those individuals who care most about the threatened domain (Aronson et al., 1999). We thus only selected students who were highly math identified. Scores on the MIQ and SAT ranged from 1.69 to 8.15 ($M = 5.60$, $SD = 1.15$) and 490 to 800 ($M = 686.09$, $SD = 65.94$), respectively. From these, 54 students were selected on the basis of having scored above the theoretical midpoint of the MIQ ($M = 6.31$, $SD = 0.73$) and having scored 570 or above (range = 570–800) on the Math portion of the SAT ($M = 710.74$, $SD = 60.46$). These 54 highly math-identified female undergraduates were selected to participate in the study in exchange for credit toward a class requirement. Students were randomly assigned to one of four conditions in a Gender Composition (minority condition vs. same-gender condition) \times Level of Privacy (private vs. public) between-subjects factorial design, with a range of 12 to 15 subjects per group.

Materials

Math test. Students took a difficult 20-item math test culled from the GRE test guide (Educational Testing Service, 1994). Past samples reveal

that each question was answered correctly by an average of only 36.6% of test takers.

Public apprehension. This manipulation check examined whether students felt that their same-gender or mixed-gender group members would negatively evaluate them as a result of their performance. Students indicated the extent to which they agreed to a number of items on a 7-point Likert scale ranging from 1 (*completely disagree*) to 7 (*completely agree*). The four scale items, revised from Spencer et al.'s (1999) Evaluation Apprehension Scale, consisted of the following items: "I am concerned that my focus-group partners will think poorly of me if I do not do well on this test" "As a result of my math score, my focus-group partners may question my ability" "My focus group partners will look down on me if I do not do well on this test" and "I'm afraid my focus-group partners will think I'm stupid as a result of my test performance." If the manipulation of privacy works, we expected that students in the private group would be somewhat unconcerned with what their partners would think because they were led to believe that their performance would be undisclosed.

Situational anxiety. To determine whether the prospect of public evaluation creates anxiety, we asked students to complete a five-item self-report measure of situational anxiety (Mattsson, 1960). They indicated how they felt at that very moment on five 7-point semantic differentials (*uneasy* vs. *at ease*, *comfortable* vs. *uncomfortable*, *upset* vs. *peaceful*, *relaxed* vs. *tense*, and *in control* vs. *not in control*). Students completed this measure twice, both before and after hearing about the goal of the study. Situational anxiety was thus analyzed as a repeated measure. Again, if the manipulation of privacy works, we anticipated that students in the public condition would show an increase in anxiety after hearing the instructions and that private students would not.

Procedure

Female students were randomly assigned to (a) either the same-gender or minority conditions and (b) either the public or private conditions. Six men, blind to experimental hypotheses, served as confederates in the minority conditions. They were trained to act like subjects and instructed to limit interactions with the experimental students. They were also instructed to fill out questionnaires and the test at about the same rate as the real student.

The experiment began once all 3 women (same-gender condition) or 1 female and 2 male confederates (minority condition) arrived in the lab, which was a small room with three student desks. Students started by filling out the first situational anxiety questionnaire. This procedure was completed before students received the experimental instruction or the manipulation of privacy to assess baseline anxiety. The experimenter then presented the cover story to the students. They were informed that the goal of the focus-group study of effective test-taking strategies was to develop a new educational training program to improve student performance on standardized tests. Students learned that as part of the study they would take a math test, after which they would enter into discussion with their focus-group members.

Students in the public-evaluation group learned that the posttest discussion would consist of broadcasting their math test results to group members and correcting and rank ordering the test performances of their focus-group members. Thus, in the public-evaluation condition, students believed that their performance on the math test would be public and openly evaluated by their partners.

To create a truly private condition, we needed to convince students that we were limiting two sources of external evaluation—one from other group members and the one from the experimenter. Thus, we adapted the procedure used by Kolditz and Arkin (1982) and instructed students in the private-evaluation group that the posttest discussion would be based on theoretical issues related to the administration and preparation of standardized tests. The private-evaluation students read a set of instructions that emphasized the importance of their anonymity and confidentiality and were informed throughout the process that their test scores would remain completely private. To express our solemn commitment in maintaining the privacy of scores, we gave students a large manila envelope that was

unmarked and free of identifying information. Students were asked to place their test in the envelope and seal it after completing the test and prior to engaging in the discussion. Students were not asked to write their names, student numbers, gender, or other identifying marks on the envelope. Moreover, we did not include any items in the test materials relating to personal identity or group identity, such as gender. Thus, to ensure a perception of privacy and anonymity, we led students to believe that (a) their tests will be cloaked from the eyes of their partners and (b) the experimenter will be unable to determine which test was written by which student.

All students were then asked to look over three sample math problems. This procedure was done to inform students of the difficulty of the task and of the possibility of performing poorly. However, to ensure that performance on these items would not affect later measures, we did not give students enough time to actually solve these items. Students were then instructed to fill out the second situational-anxiety measure and the public-apprehension measure. Finally, the experimenter gave students 20 min to complete the math test and signaled when there were 5 and 2 min remaining. During the test, the students were seated in the small lab and could observe one another's behavior and test-taking speed. After the completion of the test, the experimenter informed students that there would be no posttest discussion. Students then filled out a demographic sheet that asked them to report their math SAT score. The experimenter probed students for suspicion and then debriefed them. In sum, students completed the first anxiety scale, heard the instructions, filled out the public-apprehension measure and the second anxiety scale, and then finally took the math test.

The experimenter was not blind to condition because (a) instructions were read aloud and (b) gender composition was visually apparent. However, to minimize any expectancy effects, we used four different experimenters and instructed them to be present in the lab only when absolutely necessary (i.e., to pass out the math test; Rosenthal, 1966). Two men and two women served as experimenters, with experimenter gender counter-balanced over students.

Results

As expected with random assignment, there was no difference in students' self-reported SAT scores across experimental conditions (all F s < 1, *ns*).² There is thus little reason to believe that the privacy manipulation might have affected the accuracy of the SAT self-report. All dependent variables were analyzed using a 2 (level of privacy: public vs. private) \times 2 (gender composition: minority vs. same gender) analysis of variance (ANOVA). Because some of the dependent variables (e.g., test performance) were affected by prior math ability, SAT scores were entered as a covariate in the analysis whenever appropriate; this ensures that any effect of SAT on the dependent variables are equated across groups.

² Self-reported SAT scores were collected from each student twice, once during the screening of high math-identified students at the beginning of the semester and a second time after the experiment. In the case of 12 students (22.2% of total students), these two self-reported SAT scores differed. As a result of these discrepant self-reports and of research indicating that a significant proportion of students may overestimate their SAT scores (Shepperd, 1993), throughout this experiment we used the lower of the two self-reported SAT scores. Because only a minority of students had discrepant SAT scores and because the two SAT scores were not significantly different from one another, $F(1, 49) < 1$, *ns*, and highly correlated, $r(52) = .97$, $p < .001$, it is not surprising that analyses using other SAT measures (e.g., averaging the two SAT scores) yielded results that were virtually identical to those reported here.

Public Apprehension

We included a measure of public apprehension to determine whether students in the private condition actually perceived their performances as being private. If they did, we anticipated that students in the private group would be less worried about what their partners would think than public students. The four individual measures of public apprehension were combined into one measure ($\alpha = .90$). Results of this measure revealed that students in the public group experienced a greater fear of being negatively evaluated by their group members ($M = 3.86$, $SD = 1.47$) than did students in the private group ($M = 2.96$, $SD = 1.48$), as expected. Quantitative SAT scores were not a significant predictor of evaluation apprehension ($F < 1$, *ns*) and were thus not entered as a covariate in the analysis. A 2×2 ANOVA revealed a main effect for the publicity manipulation, $F(1, 46) = 4.30$, $p < .05$,³ indicating that the manipulation was successful in influencing worries about public evaluation. Thus, students who thought they were going to broadcast their scores to one another were worried about this broadcasting. This finding suggests that our manipulation of privacy was successful. No other effects were significant or close to being significant—the gender-composition manipulation had no obvious effect on the measure of public apprehension.

Situational Anxiety

The measure of situational anxiety was included to determine whether students felt more anxious as a consequence of the privacy manipulation. The five individual measures of situational anxiety were administered twice, both before and after the experimenter presented the instructions. These individual measures were averaged into one baseline measure ($\alpha = .77$) and one postinstructions measure ($\alpha = .86$) of situational anxiety. Quantitative SAT scores were not a significant predictor of evaluation apprehension, $F(1, 47) = 2.60$, $p > .10$, and were thus not entered as a covariate in the analysis. The anxiety measures were analyzed as a repeated measure in a 2 (anxiety) $\times 2$ (level of publicity) $\times 2$ (gender composition) mixed ANOVA, with all factors but anxiety as between subjects. The analysis revealed a significant main effect for anxiety, $F(1, 47) = 14.12$, $p < .01$;⁴ however, this main effect was subsumed under a significant anxiety by privacy interaction, $F(1, 47) = 6.29$, $p < .02$. Students who expected to publicly share their math tests (public-evaluation group) felt significantly more anxious after hearing the instructions ($M = 4.14$, $SD = 0.95$) than at baseline ($M = 3.44$, $SD = 1.02$), $F(1, 47) = 18.55$, $p < .001$. In contrast, students who expected their scores to remain private (private-evaluation group) felt as anxious after hearing the instructions ($M = 3.51$, $SD = 1.08$) as at baseline ($M = 3.38$, $SD = 1.00$; $F < 1$, *ns*). Not surprisingly, students who believed that they were going to be the only ones to see their scores were more relaxed than those who thought their scores would be broadcast to others. No other effects were significant or close to significant.

Test Performance

This experiment investigated the possibility of situations threatening private performances. It tested whether the effects of threat are restricted to environments that are highly public or whether they extend to environments that secure privacy. If threat is restricted to public environments—if it's strictly a self-

presentational phenomenon—then minority women would only perform worse than same-gender women in the public condition. If, however, threat extends to private environments—if it involves intrapsychic elements—then minority women would perform worse than same-gender women in both public and private conditions.

Performance was measured in this study (consistent with the stereotype threat literature) with three indices: (a) scores adjusted for guessing, which utilizes the standard Educational Testing Service formula for scoring and is computed by giving 1 point for correct items, no points for items left blank, and a deduction of a quarter point for incorrect items, all divided by the total number of questions on the test (e.g., Davies, Spencer, Quinn, & Gerhardtstein, 2002; Spencer et al., 1999); (b) number correct, which is simply the raw number of items answered correctly; and (c) accuracy, which is computed by the number of items correct divided by the total number of items attempted (Steele & Aronson, 1995). Not surprisingly, past math performance, as measured by the SAT, significantly predicts math performance on our math test (using all three indexes of performance), all $F_s(1, 49) > 45.66$, $p < .001$. SAT scores were thus entered as a covariate in all analyses of performance.

The means for scores adjusted for guessing show that minority status (and stereotype threat) can affect private performances: Minority women ($M = 0.34$, $SD = 0.12$) in both public and private groups got lower adjusted scores than same-gender students ($M = 0.43$, $SD = 0.13$; see Figure 1). A 2×2 analysis of covariance (ANCOVA) on adjusted scores, with self-reported SAT scores as the covariate, resulted in a significant main effect for gender composition of the group, $F(1, 49) = 6.97$, $p < .02$. This constitutes a medium effect size (Cohen's $d = .70$). The interaction between privacy and gender composition, however, was not significant, $F(1, 49) < 1$, *ns*, showing that there was no significant difference between the public and private groups. Thus, we found little support for the view that minority environments can only threaten public performances.

The lack of a significant interaction indicates that the general main effect for gender composition was not significantly different across both levels of publicity. Nonetheless, an internal analysis of the effect was conducted to more closely examine whether the overall effect is larger and more significant for one group than for another. This analysis showed that among the public-evaluation group, minority students ($M = 0.33$, $SD = 0.13$) attained marginally lower scores than same-gender students ($M = 0.41$, $SD = 0.13$), $F(1, 49) = 3.23$, $p < .08$, $d = .62$. Similarly, among the private-evaluation group, minority students ($M = 0.35$, $SD = 0.13$) did marginally worse than same-gender students ($M = 0.45$, $SD = 0.13$), $F(1, 49) = 3.70$, $p < .06$, $d = .75$. Thus, the internal analysis reaffirms the main effect coupled with no interaction—it shows that the minority underperformance effect is the same across both levels of privacy.

³ Four students failed to complete the entire index of evaluation apprehension and were thus excluded from the analysis of this measure. The degree of freedom (46) reflects this.

⁴ Three students failed to complete both versions of the situational anxiety measure and were thus excluded from the analysis of this measure. The degree of freedom (47) reflects this.

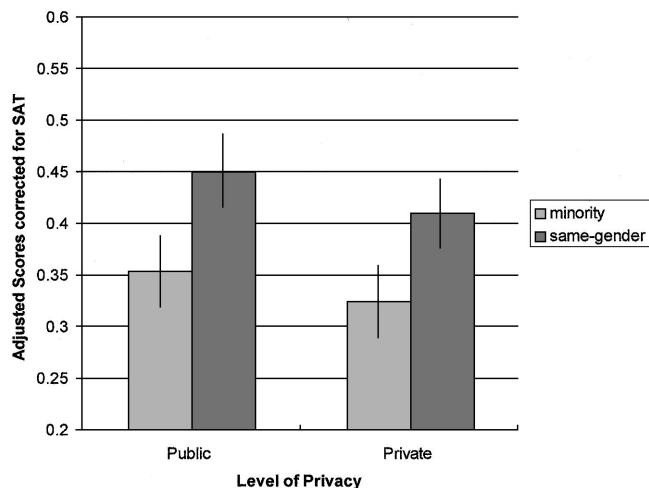


Figure 1. Adjusted scores, corrected for SAT, as a function of gender composition of group and level of privacy. Error bars represent standard error.

The analyses of number correct and accuracy scores demonstrate a similar pattern of support for the private-evaluation hypothesis. A 2×2 ANCOVA performed on the raw scores, with self-reported SAT as the covariate, revealed a main effect for gender composition such that minority women got fewer number correct ($M = 8.21$, $SD = 2.23$) than same-gender women ($M = 9.56$, $SD = 2.25$), $F(1, 49) = 4.87$, $p < .04$. This constitutes a medium effect size (Cohen's $d = .60$). A 2×2 ANCOVA on accuracy revealed the same main effect, with minority students being less accurate ($M = 0.60$, $SD = 0.15$) than same-gender students ($M = 0.71$, $SD = 0.15$), $F(1, 49) = 8.00$, $p < .01$. This is a medium effect (Cohen's $d = .73$). Thus, raw scores and accuracy scores replicated the findings, with adjusted scores showing that women in the minority group performed worse than women in the same-gender group. Most relevant to our central question, however, minority women performed poorly in both public and private settings.

Discussion

Being outnumbered in a stereotyped environment is threatening. When this environment is public—for example, when other classmates can openly evaluate a person's performance—it often leads to performance deficits. But what happens when such an environment is private—when the only eyes evaluating one's performance are one's own? Does this situation still lead to intellectual underperformance? The data from the current study answers in the affirmative. The results show that privacy cannot armor students from the assaults posed on them by stereotypes or solo status. Even when not sharing performance outcomes, women outnumbered by men got fewer items correct, were less accurate, and got lower adjusted scores than same-gender women on a difficult math test. Being in the minority may thus create a threatening intellectual environment for stigmatized groups in both public and private evaluation settings.

While acknowledging the possibility of disruptions in private performances (e.g., Steele & Aronson, 1995), the theories of stereotype threat and solo status have tended to emphasize the

public sources of threat. This study demonstrates that these processes can be engaged by more than mere self-presentational motives—they may also come about through intrapsychic processes, such as self-evaluation (Tetlock & Manstead, 1985). An intrapsychic source of threat may suggest that individuals are motivated to prove to themselves that group stereotypes are untrue; individuals may be motivated to quell any self-doubts about the veracity of group stereotypes. The data imply that stereotype threat may not be a strictly situational phenomenon, but reinforces the notion of threat as a person by situation interaction (Steele, 1997). The findings are consistent with more recent research showing that one's internal beliefs predict vulnerability to threat (Schmader, 2002) and point to the need to more fully examine how internal beliefs interact with self-evaluative factors to produce intellectual deficits.

It is important to note that this study was not meant to pit impression management versus self-evaluative explanations of threatening environments—we are not of the opinion that they are mutually exclusive oppositions. Rather, we agree with Tetlock and Manstead (1985) that both processes may be in operation some of the time. For example, it may be the case that self-presentational concerns are in operation during highly public performances, such as when making a speech, whereas self-evaluative concerns may be activated in private, such as when practicing for an SAT test at home. The main points of the current research were to extend Spencer et al.'s (1999) work by demonstrating that public evaluation neither mediated nor moderated threatening situations and to show that the mere presence of nonevaluative others is sufficient to evoke threat.

Even though stereotypes can interfere with performance in many environments (including both minority and same-gender settings), they first need to be activated in order to have an effect—and some environments are better stereotype activators than others. For instance, solo situations are more likely to activate group stereotypes than same-gender situations (Deaux & Major, 1987). We, therefore, hypothesized that minority contact, which can increase the distinctiveness of one's group (McGuire & Padawer-Singer, 1976) and the salience of stereotypes about that group (Devine, 1989), would cause performance-hindering stereotype threat. Our results supported this hypothesis by demonstrating minority underperformance in both public and private settings.

This result, however, should not come at the expense of examining individual differences. It was clear from the data that some women were not threatened by minority situations—some actually did very well, getting as many as 17 correct problems out of a maximum of 20. In other words, when confronted with the possibility of failure, students may have reacted in one of two ways: (a) they may have made threat appraisals by construing the situation as an unmanageable stress that surpassed coping resources—in this case, the possibility of failure may have had a negative effect on achievement; or (b) they may have made challenge appraisals by construing the situation as manageable and seeing the potential for gain—in this case, the possibility of failure may have strengthened motivation and thus stimulated achievement (Ben-Zeev et al., in press; Tomaka, Blascovich, Kelsey, & Leitten, 1993).

What individual differences, then, would predict how women appraised the minority situations? Although the current research cannot inform this question directly, recent research on self-monitoring may prove insightful to examining this question. Self-monitoring is the tendency to regulate one's own behavior to meet

the demands of social situations (Snyder, 1974). In a series of studies, DeMarree, Wheeler, and Petty (2003) showed that high self-monitors, those individuals who are good at managing the impressions they project, exhibited stereotype contrast effects, whereas low self-monitors, those individuals who are less concerned about the propriety of their behaviors, exhibited stereotype assimilation effects. Although speculative, these results suggest that when confronted with negative ability stereotypes, low self-monitors may have been threatened and thus may have assimilated to the stereotype (performed poorly), whereas high self-monitors—who have the social repertoire to cope with stress—were perhaps challenged and thus may have experienced contrast with the stereotype (performed well). Current and future research in our lab has been designed to investigate this conjecture (cf. Inzlicht & Aronson, 2003).

Future investigations are also aimed at exploring how being outnumbered may affect attitudes and self-esteem. Although minority situations (and stereotype threat) may lead stigmatized individuals, such as women or African Americans, from attaining high academic performances, we should not conclude that low self-esteem or low academic aspirations would necessarily follow. On the contrary, some stigmatized groups have high self-esteem (see Crocker & Major, 1989), positive abstract attitudes toward education (Mickelson, 1990), and high academic aspirations (Stevenson, Chen, & Uttal, 1990) despite suffering poor academic outcomes.

This “attitude-achievement” paradox (Mickelson, 1990) and self-esteem paradox (Crocker & Major, 1989) may be explained by a number of factors. First, although the positive abstract attitudes (which reflect dominant ideology) do not correspond to the poor outcomes the stigmatized receive, concrete attitudes do. When the concrete—or real—attitudes are examined, research finds that stigmatized individuals devalue academics following poor outcomes (Mickelson, 1990). Second, when this devaluing of academics becomes chronic, these individuals may adjust their self-concept by disengaging their self-esteem from academics, thus shielding themselves from the affective consequences of poor school outcomes and “disidentifying” from academics more generally (Major, Spencer, Schmader, Wolfe, & Crocker, 1998; Osborne, 1995, 2001). Our results suggest that minority environments threaten performance. This is a first crucial step in a research agenda that is aimed at examining whether and how threatening environments impact concrete attitudes, self-esteem, and disidentification.

Educational Implications

This study has several practical implications. First, classrooms and testing centers may become disruptive for those targeted by stereotypes—just by dint of being integrated. The mere mixing of students could at times induce stereotype threat, which is the performance-disruptive anxiety high-achieving individuals feel in situations in which their behaviors have the potential to confirm negative reputations about their group. Second, this study implies that women’s performance can be undermined even when they do not engage in overtly public activities, such as speaking up in class or participating in group work. Factors that activate negative stereotypes—even with relatively private performances—may be sufficient to disrupt performance and possibly create threatening classroom environments. Thus, stereotype threat and minority sta-

tus may be more problematic than previously thought because they may extend to a broader set of situations and intellectual tasks. For example, low-status individuals may be concerned with confirming negative stereotypes to themselves even when engaged in private tasks such as practicing for a standardized test or reading to oneself.

The present study is also highly relevant to the discussion surrounding the merits and shortcomings of single-gender education. Advocates of single-gender education argue that separating the genders can encourage female interest in the “masculine” subject areas such as math and science and reduce the effects of gender stereotypes (American Association of University Women Educational Foundation, 1992, 1998; Riordan, 1990). Stereotype threat research, including the current data, suggests that mixed-gender math classrooms may pose serious threats to high-math identified females because their math performance tends to decrease in the presence of males (e.g., Inzlicht & Ben Zeev, 2000).

It is possible that single-gender environments, such as peer groups or clubs (e.g., Women In Science and Engineering), may offer women a respite from the chronic concern of being a target of negative stereotypes about their math ability (L. M. Brown, 1998). However, our data do not necessarily advocate single-gender education because (a) our study has not been conducted in authentic classroom environments, and even though it has the potential to inform such environments, it would require further exploration to become more ecologically valid and (b) our current goals were to investigate the underlying mechanisms and effects of stereotype threat, so that it would be possible to design interventions that are aimed at mitigating the effects of stereotypes on intellectual processing by targeting these mechanisms. It may very well be that the practice of stress-reduction techniques (see Ben-Zeev et al., in press, for a discussion) and self-affirmation (Croizet & Désprès, 2003) may be very helpful to this cause without a need to resort to single-gender education.

In fact, there is already some evidence that mixed-group settings can be used to reduce the effects of threatening intellectual environments. Steele et al. (1997), for example, designed a “wise” schooling intervention for first-year students at the University of Michigan. Using mixed-race groups, students were “honorifically” recruited to the program by emphasizing that they had survived a very competitive admission process at the school and that the University recognized their strong potential and had high expectations for them—all things that signal the insignificance of negative group stereotypes. Once in the program, students were reminded of these high expectations and challenged with weekly workshops on advanced material that went beyond material presented in most freshman classes. Several years of the program demonstrate that such practices can substantially increase the school performance of African Americans. Other things teachers can do include reducing the apparent diagnosticity of tests (Steele & Aronson, 1995), presenting tests as tests of malleable skill rather than as fixed ability (Aronson, Fried, & Good, 2002; Dweck, 1999), and increasing the number of minority teachers and role models (Marx & Roman, 2002; see Aronson, 2002, for a review).

Qualifications and Alternative Explanations

Our support of the private-evaluation hypothesis comes with a number of qualifications. First, as mentioned above, we wonder whether this research generalizes to educational settings, such as

classrooms. In the interest of careful generalization, we note that we found our effects with 3-person groups. It is thus hard to say whether these effects will translate to classrooms with tens of students or to workplaces with hundreds of people. Similarly, we ask whether our work can generalize to other stigmatized groups, such as African Americans. Although other research has examined the effects of solo status (e.g., Sekaquaptewa & Thompson, 2002) or stereotype threat (e.g., Steele & Aronson, 1995) on African Americans, it would be a mistake to assume that African Americans and women are psychological equivalents and thus identically threatened by minority situations. Our research focused exclusively on women in math and thus sets limits on the generalizations to other stigmatized groups. Thus, a central challenge of future research will be to conduct field experiments where, for example, the academic performance of women and ethnic minorities can be compared across classrooms where they are and are not outnumbered by nonstigmatized others.

Second, our support for the self-evaluative route to threat rests with our success in creating a truly private condition. If we had not been successful in accomplishing this goal, it is possible that students in the private condition were actually motivated by a muted form of self-presentation, rather than self-evaluation. In other words, they may have felt the same fears about looking good to others as the students in the public group. A number of facts, however, argue against this alternative explanation. First, we took all steps necessary to assure students of their complete privacy: using procedures tried and tested by other researchers (e.g., Kolditz & Arkin, 1982; Stangor, Swim, Van Allen, & Sechrist, 2002), we told students that their test results would be anonymous and instructed them to seal their tests (free of personal and group markings) in an unmarked manila envelope. Second, data from the anxiety manipulation check indicate that private students were no more anxious after hearing the instructions than before hearing them; in contrast, the public group reacted to the instructions with increased anxiety. Finally, the items measuring public evaluation indicated that students in the private condition were less concerned with being publicly evaluated than those in the public group.

Finally, we have argued that the mere presence of men was capable of activating negative stereotypes and the desire to disconfirm them, which ultimately led to underperformance. Our data, however, may be explained another way. Namely, it is possible that women prefer different, more time-consuming problem-solving strategies than men (e.g., Quinn & Spencer, 2001). And when placed in a room where they can see male confederates perhaps racing through a test, this may make them nervous and apprehensive. In other words, it is possible that the men's speed of performance, and not their presence per se, affected the women's performance, which may still have reflected correct, yet time-consuming solutions. A number of factors lead us to believe that this interpretation is unlikely. First, male confederates were trained to take the test at about the same rate as the real students. It is therefore unlikely that they finished the test much faster—or much slower—than the real students. Second, when Inzlicht and Ben-Zeev (2000) used the identical paradigm with students and confederates openly able to perceive speed of performance, women suffered deficits in math performance, but not in verbal performance. Similarly, paradigms that involved students "interacting" with videotaped bogus partners (e.g., Sekaquaptewa & Thompson, 2003) or that involved students being unable to see one another (e.g., Ben-Zeev, Inzlicht, & Fein, 2001) resulted in

similar minority underperformance. Thus, it does not appear that women seeing male confederates possibly race through the test—which men were explicitly instructed not to do—could explain the results. Rather, something about the men's presence in the math domain seems relevant. Finally, there is little evidence that the women used correct but time-consuming strategies. As reported earlier, when we examined accuracy—which is the number correct per item attempted and reflects a score that is free from time constraints—women in the minority group did worse than those in the same-gender group.

Conclusions

Merely placing high-achieving women in a room where men outnumber them creates a threatening intellectual environment, and consequently, leads them to do worse on a math test. Threatening situations, however, are not confined to oral tasks, group testing, or public performances. Rather, they appear to extend into the realm of the private. For targets of stigmatization (Crocker & Major, 1989; Major et al., 1998; Stangor, Carr, & Kiang, 1998; Swim & Stangor, 1998), the implications of this reality are clear—women may perform and learn at levels beneath their potential. Fortunately, because the threat posed by being outnumbered and by stereotypes is partly induced by situations, there is much that concerned educators can do to shield students from these threatening classroom environments.

References

- American Association of University Women Educational Foundation. (1992). *How schools shortchange girls*. Washington, DC: Author.
- American Association of University Women Educational Foundation. (1998). *Separated by sex. A critical look at single-sex education for girls*. Washington, DC: Author.
- Aronson, J. (2002). Stereotype threat: Contending and coping with unnerving expectations. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 279–301). San Diego, CA: Academic Press.
- Aronson, J., Fried, C., & Good, C. (2002). Reducing the effects of stereotype threat on African-American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38*, 113–125.
- 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.
- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer Jr. (Ed.), *Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.
- Benbow, C. P., & Stanley, J. C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Educational Research Journal*, *19*, 598–622.
- Ben-Zeev, T., Carrasquillo, C. M., Ching, A., Kliengklom, T. J., McDonald, K. L., Newhall, D. C., et al. (in press). "Math is hard!" (Barbie, 1994): Responses of threat vs. challenge mediated arousal to stereotypes alleging intellectual inferiority. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics*. Cambridge, MA: Cambridge University Press.
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2003). *Does arousal mediate stereotype threat?* Unpublished manuscript, San Francisco State University.
- Ben-Zeev, T., Inzlicht, M., & Fein, S. (2001, November). *When the presence of males causes females to experience problem-solving deficits*.

- Paper presented at the 42nd Annual Convention of the Psychonomic Society, Orlando, FL.
- Berglas, S., & Jones, E. E. (1978). Drug choice as a self-handicapping strategy in response to noncontingent success. *Journal of Personality and Social Psychology*, *36*, 405–417.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, *66*, 5–20.
- Brown, L. M. (1998). Ethnic stigma as a contextual experience: A possible selves perspective. *Personality and Social Psychology Bulletin*, *24*, 163–172.
- Brown, R. P. (2000). *The importance of importance: Factor-analytic conclusions about domain identification*. Unpublished manuscript, University of Oklahoma.
- Brown, R. P., & Josephs, R. A. (1999). Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology*, *76*, 246–257.
- Cohen, L. L., & Swim, J. K. (1995). The differential impact of gender ratios on women and men: Tokenism, self-confidence, and expectations. *Personality and Social Psychology Bulletin*, *21*, 876–884.
- Cota, A., & Dion, K. L. (1986). Salience of gender and sex composition of ad hoc groups: An experimental test of distinctiveness theory. *Journal of Personality and Social Psychology*, *50*, 770–776.
- Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, *96*, 608–630.
- Croizet, J. C., & Désprés, G. (2003, February). How does stereotype threat undermine performance? In J. C. Croizet & S. Spencer (Chairs), *Understanding how stereotype threat impairs academic achievement: A second generation of research*. Symposium conducted at the 4th Annual Meeting of the Society for Personality and Social Psychology, Los Angeles, CA.
- 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.
- Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review*, *94*, 369–389.
- DeMarree, K. G., Wheeler, S. C., & Petty, R. E. (2003, February). *The role of self-monitoring in stereotype-to-behavior effects*. Paper presented at the 4th Annual Meeting of the Society for Personality and Social Psychology, Los Angeles, CA.
- Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.
- Dweck, C. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia: Psychology Press.
- Educational Testing Service. (1994). *GRE: Practicing to take the general test* (9th ed.). Princeton, NJ: Author.
- Fuegen, K., & Biernat, M. (2002). Reexamining the effects of solo status for women and men. *Personality and Social Psychology Bulletin*, *28*, 913–925.
- Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences*, *19*, 229–284.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta analysis. *Psychological Bulletin*, *107*, 139–155.
- Inzlicht, M., & Aronson, J. (2003). *Monitoring's moderating muscle: When being outnumbered doesn't hurt intellectual performance*. Unpublished manuscript, New York University.
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365–371.
- Jackson, P. W. (1968). *Life in classrooms*. New York: Holt, Rinehart & Winston.
- Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 1, pp. 231–262). Hillsdale, NJ: Erlbaum.
- Kanter, R. M. (1977). *Men and women of the corporation*. New York: Basic Books.
- Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin*, *105*, 198–214.
- Kolditz, T. A., & Arkin, R. M. (1982). An impression management interpretation of the self-handicapping strategy. *Journal of Personality and Social Psychology*, *43*, 492–502.
- Lord, C. G., & Saenz, D. S. (1985). Memory deficits and memory surfeits: Differential cognitive consequences of tokenism for tokens and observer. *Journal of Personality and Social Psychology*, *49*, 918–926.
- Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin*, *24*, 34–50.
- Markus, H. (1978). The effects of mere presence on social facilitation: An unobtrusive test. *Journal of Experimental Social Psychology*, *14*, 389–397.
- Marx, D., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, *28*, 1183–1193.
- Mattsson, P. O. (1960). Communicated anxiety in a two-person situation. *Journal of Consulting Psychology*, *24*, 488–495.
- McGuire, W. J., & Padawer-Singer, A. (1976). Trait salience in the spontaneous self-concept. *Journal of Personality and Social Psychology*, *33*, 743–754.
- Mickelson, R. A. (1990). The attitude-achievement paradox among Black adolescents. *Sociology of Education*, *63*, 44–61.
- National Science Foundation. (2000). *Women, minorities, and persons with disabilities in science and engineering: 2000* (NSF Publication No. 00-327). Arlington, VA: Author.
- Osborne, J. (1995). Academics, self-esteem, and race: A look at the underlying assumptions of the disidentification hypothesis. *Personality and Social Psychology Bulletin*, *21*, 449–455.
- Osborne, J. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, *26*, 291–310.
- Ott, E. M. (1989). Effects of the male-female ratio at work. *Psychology of Women Quarterly*, *13*, 41–57.
- 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.
- Riordan, C. (1990). *Girls and boys in school: Together or separate?* New York: Teachers College Press.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
- Saenz, D. S. (1994). Token status and problem-solving deficits: Detrimental effects of distinctiveness and performance monitoring. *Social Cognition*, *12*, 61–74.
- Schmader, T. (2002, February). *Stereotype threat as seen through the lens of social identity theory*. Paper presented at the 3rd Annual Meeting of the Society for Personality and Social Psychology, Savannah, GA.
- Sekaquaptewa, D., & Thompson, M. (2002). The differential effects of solo status on members of high- and low-status groups. *Personality & Social Psychology Bulletin*, *28*, 694–707.
- 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.
- Shepperd, J. A. (1993). Student derogation of the Scholastic Aptitude Test: Biases in perceptions and presentations of College Board scores. *Basic and Applied Social Psychology*, *14*, 455–473.
- Snyder, M. (1974). The self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*, 526–537.
- Spencer, S. J., Steele, C. M., & Quinn, D. (1999). Stereotype threat and

- women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology*, 75, 1191–1197.
- Stangor, C., Swim, J. K., Van Allen, K. L., & Sechrist, G. B. (2002). Reporting discrimination in public and private contexts. *Journal of Personality and Social Psychology*, 82, 69–74.
- Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.
- Steele, C. M., Spencer, S., & Aronson, J. (2002). Contending with images of one's group: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 379–440). San Diego, CA: Academic Press.
- Steele, C. M., Spencer, S. J., Hummel, M., Carter, K., Harber, K., Schoem, D., & Nisbett, R. (1997). *African-American college achievement: A "wise" intervention*. Unpublished manuscript, Stanford University.
- Stevenson, H. W., Chen, C., & Uttal, D. H. (1990). Beliefs and achievement: A study of Black, White, and Hispanic children. *Child Development*, 61, 508–523.
- 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.
- Swim, J. K., & Stangor, C. (1998). *Prejudice: The target's perspective*. San Diego, CA: Academic Press.
- Tetlock, P. E., & Manstead, A. S. R. (1985). Impression management versus intrapsychic explanations in social psychology: A useful dichotomy? *Psychological Review*, 92, 59–77.
- Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology*, 65, 248–260.
- Travers, K. J., & Westbury, I. (1989). *The IEA study of mathematics: I. Analysis of mathematics curricula*. New York: Pergamon Press.
- Williams, C. L. (1992). The glass escalator: Hidden advantages for men in the "female" professions. *Social Problems*, 39, 253–267.
- Zajonc, R. B. (1965). Social facilitation. *Science*, 149, 269–274.

Received November 14, 2002

Revision received June 24, 2003

Accepted July 1, 2003 ■

ORDER FORM

Start my 2004 subscription to the *Journal of Educational Psychology*! ISSN: 0022-0663

_____ \$65.00, APA MEMBER/AFFILIATE _____
 _____ \$133.00, INDIVIDUAL NONMEMBER _____
 _____ \$321.00, INSTITUTION _____
 In DC add 5.75% / In MD add 5% sales tax _____
TOTAL AMOUNT ENCLOSED \$ _____

Subscription orders must be prepaid. (Subscriptions are on a calendar year basis only.) Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.



AMERICAN
PSYCHOLOGICAL
ASSOCIATION

SEND THIS ORDER FORM TO:
 American Psychological Association
 Subscriptions
 750 First Street, NE
 Washington, DC 20002-4242

Or call (800) 374-2721, fax (202) 336-5568.
 TDD/TTY (202) 336-6123.
 For subscription information, e-mail:
subscriptions@apa.org

Send me a FREE Sample Issue
 Check enclosed (make payable to APA)
Charge my: VISA MasterCard American Express

Cardholder Name _____
 Card No. _____ Exp. Date _____

 Signature (Required for Charge)

BILLING ADDRESS: _____

City _____ State _____ Zip _____
 Daytime Phone _____
 E-mail _____

SHIP TO:
 Name _____
 Address _____

City _____ State _____ Zip _____
 APA Member # _____ EDUA14