EFFECTS OF STEREOTYPE THREAT ON FEMALES IN MATH AND SCIENCE FIELDS: AN INVESTIGATION OF POSSIBLE MEDIATORS AND MODERATORS OF THE THREAT-PERFORMANCE RELATIONSHIP

A Dissertation
Presented to
The Academic Faculty

by

Alice Anne Bailey

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Psychology

Georgia Institute of Technology

December, 2004

Effects of Stereotype Threat on Females in Math and Science Fields: An Investigation of Possible Mediators and Moderators of the Threat-Performance Relationship

Approved by:

Dr. Jack Feldman

Dr. Lawrence James

Dr. Ruth Kanfer

Dr. Dianne Leader

Dr. Amy S. Bruckman

November 15, 2004

For My Mother

Acknowledgements

I would like to express my sincere gratitude to the members of my committee for their time, leadership, and expertise.

I would especially like to thank Jack Feldman for his many years of support and mentorship.

Table of Contents

List of Tables

List of Figures

Summary

A mediated-moderation model of stereotype threat was tested. Domain identification and motivational orientation were treated as moderators of the threat effect on self-efficacy and cognitive interference, which were hypothesized to mediate the threat-performance relationship. Participants were primed with stereotype-consistent, stereotype inconsistent, or no information regarding sex differences in mathematical abilities. While significant performance differences were found between males and females in the control and threat conditions, no differences were found in a “female benefit” condition that described a math task as favoring females. Significant sex differences in domain identity and self-efficacy were also found. Post-hoc analyses revealed that domain identification and self-efficacy explained significant amounts of variance in sex differences in math performance. The results provide general support for Steele’s theory of stereotype threat and resulting disidentification with the task domain among targets.

Introduction

Stereotype threat theory (Steele, 1997; Steele & Aronson, 1995) predicts that performance on a challenging task will decline in the face of a negative stereotype regarding one's abilities. Steele's theory and preliminary findings have several important implications for interpreting performance on high-stakes employment and other standardized tests. Several laboratory studies have supported its predictions while extending findings to different performance domains and various stereotyped groups (Aronson, Fried, & Good, 2002; Aronson, Lustina, Keough, Steele, & Brown, 1999; Croizet & Claire, 1998; Gonzales, Blanton, & Williams, 2002; Levy 1996; McKay, 1999; Stone, Lynch, Sjomeling, & Darley, 1999; Shih, Pittanksy, & Ambady, 1999; Schmader, 2002; Spencer, Steele, & Quinn, 1999; Walsh, Hickey, & Duffy, 1999); however, other studies have failed to find any effect of stereotype threat on performance, particularly in applied settings (Foote, 2000; Harder, 1999; Lewis, 1999; McFarland, Lev-Arey, & Ziegert, 2003; Mayer & Hanges, 2003; Nguyen, O'Neal, & Ryan, 2003; Oswald & Harvey, 2000; Ployhart, Ziegart, & McFarland, 2003; Stricker & Bejar, 1999; Stricker, 1998; Stricker & Ward, 1998). Therefore, it is necessary to evaluate the basis of the phenomenon. In addition, conflict has arisen regarding the definition of a control condition in stereotype threat research. Steele (2003) contends that only a condition that removes threat by manipulating a test's description as non-diagnostic can serve as a true control, yet this manipulation may have a significant impact on both targets and non-targets. Only one study to date (Shih et al., 1999) has used three experimental conditions (positive stereotype, control, and stereotype threat) to investigate the effects of various testing conditions on minority performance. Furthermore, investigations of proposed mediators have produced conflicting results (Aronson et. al, 1999; Aronson, Quinn, & Spencer, 1998; McKay, 1999; Oswald & Harvey, 2000; Spencer et. al, 1999; Stangor, Carr, & Kiang, 1999; Steele & Aronson, 1995). Research in areas such as motivation (Kanfer & Ackerman, 1989), automaticity (Bargh, 1994), reactions to tests (Baumeister, 1984; Chan, Schmitt, Sacco, & DeShon, 1998); cognitive interference (Sarason, Sarason, Keefe, Hayes, & Shearin, 1986) and social categorization (Brewer, 1991; Tajfel & Turner, 1985) lend support to the threat phenomenon, but these theories have received little discussion in the stereotype literature. Considering the mediating and moderating factors that impact or determine its occurrence will help researchers understand how and why stereotype threat operates. An integrative model is proposed which incorporates theories of social cognition, motivation, and self-efficacy. The purpose of this study is to 1) further replicate the effect in a mathematically-gifted and highly-identified sample; 2) compare minority performance under conditions of stereotype threat, standard testing instruction, and "threat-removed" circumstances; and 3) to investigate several possible mediators and moderators in the proposed model. It is hoped that findings will expand the theory by determining its boundary conditions and help to identify the process by which stereotype threat disrupts performance. If found, knowledge of such mediators would not only further elaborate Steele and Aronson's (1995) initial theory, but may also be used to find ways of ameliorating any debilitating effect.

Implications of Negative Stereotypes for Women in Math and Science

The late twentieth century witnessed a dramatic change in the nature of work. Intense competition has driven a reliance on technology to decrease product development time, improve access to information, and increase the speed of communication. If the

U.S. is to remain competitive in an international market, it must produce more specialists in scientific and technical fields in order to stay ahead of the technology curve, yet the United States faces a dramatic shortage of technology and engineering professionals in many areas. Despite the present recession and exportation of IT jobs overseas, demand for technological skills, particularly at the most advanced levels, is still exceeding supply (ITAA, 2001). In previous years, U.S. companies could fill the demand in needed skills by luring top professionals from other countries. However, recent homeland security actions have cut off the supply of foreign nationals (Macdonald, 2004). In addition, companies experiencing a rapid growth in the IT industry in India are successfully recruiting American-educated foreign talent (who compose half of all technology-related Ph.D.s earned in the U.S.) away from the states. Corporations such as HP and Microsoft expect their hiring needs to be larger than can be met over the next decade (Macdonald, 2004.)

Greater numbers of women in the fields of engineering and computer science could fill the labor void; however, while the number of females entering the job market is increasing at a greater rate than males overall, very few women pursue engineering or information technology careers (Shashaani, 1997; Bureau of Labor Statistics, 1989). Women represent over half of the workforce (56.8%), but only 8.5% of all professional engineers (Goodman Research Group, 2002) and 24% of the total IT labor market (Johnson, 2000), where their presence is disproportionately skewed to lower-level positions such as data entry (Lorek, 2000).

One might argue that the scarcity of women in the IT field reflects an elder generation of females who were not raised with computers. However, the percentage of females earning computer science (CS) degrees has instead declined over the past twenty years, steadily decreasing from 37% in 1984 to 20% in 1999 (National Council for Research on Women, 2001). While most previously male-dominated fields such as law and medicine have achieved sex equity over the past 30 years, other academic majors such as math and engineering still have a significantly small female enrollment (Goodman Research Group, 2002). Men outnumber women 3 to 1 in postsecondary CS and engineering courses (American Association of University Women, 2000), but the gravitation away from math and science subject domains begins at an even earlier age. For example, at the secondary level, girls represent only 15% of all Advanced Placement (AP) Computer Science test takers, and less than 10% of the higher-level “AB” test takers (Johnson, 2000; Stumpf & Stanley, 1996). As a result, their percentage in the IT workforce is likewise decreasing (Johnson, 2000). This decline has steadily continued, despite the fact that women now earn a greater percentage of all college degrees than men do (Levenson, 1991; Shashaani, 1997; U.S. Census Bureau, 2000).

Even though meta-analyses of secondary school aptitude tests taken between 1947 and 1983 show that sex differences are practically disappearing, with girls increasing in spatial, mathematics, and mechanical aptitude scores and boys showing similar gains in grammar, spelling, and perceptual speed, one difference has remained consistent over time: Girls’ scores on upper-level mathematics tests continue to lag behind boys’ (Feingold, 1988). Hyde, Fennema, & Lamon (1990) found that sex differences in mathematical problem solving do exist, and that these differences increase with age and task complexity. Differences are greater, for instance, on advanced tests of problem solving ability than on those of mere computational ability. From 1977 to 1988, males were found to outperform females on the quantitative section of the Graduate Record Exam (GRE) by 80 points, which is more than half a standard deviation (Wah & Robinson, 1990). Interestingly, this sex gap in standardized test performance is found to an equal degree for both low- and high-GPA students (Strenta, Elliott, Adair, Scott, & Matier, 1993, cited in Steele, 1997).

Several theories have been proposed to account for sex differences in mathematical achievement, ranging from physiological bases, such as proportional differences in gray and white matter (Gur, Turetsky, Matsui, Yan, Bilker, Hughett, & Gur, 1999) and hormone levels (Maki, 2002; Shaywitz, Naftolin, Zelterman, Marchione, Holahan, Palter, & Shaywitz, 2003) on the one hand, to environmental bases, such as socio-cultural differences in how boys and girls are raised, on the other (Hewitt & Seymour, 1991; Hyde, Fennema, Ryan, Frost, and Hopp, 1990).

One rather recent environmental explanation proposes that females are deterred from engaging in technology-related fields because negative cultural stereotypes about their math and science abilities increase anxiety and/or reduce self-confidence. For example, a pervasive stereotype persists among college students that females are admitted to CS programs in order to increase student diversity, not because of their skills, which are often perceived to be lower than males' (Hammond, 2001; Margolis, Fisher, & Miller, 2000). Despite sex equity in course grades, women continue to drop out of math and engineering degree programs at a significantly greater rate than males do, often citing low self-confidence in their engineering ability.

In support of this argument, female "underprediction" has been demonstrated in college-level math and science courses such that women's GPAs are often higher than would be predicted by their Scholastic Aptitude Test (SAT) scores. In fact, women earn higher grades in mathematics courses than do males with corresponding SAT Mathematics (SAT-M) scores (Wainer & Steinberg, 1992); women score about 21 to 55 SAT points, one standard deviation, below men who earn the same math course grades (Sheeham & Gray, 1992). Interestingly, among adolescents, the difference between predicted and actual scores is greater for young women at the higher range of ability than at the lower ends. Kimball (1989), for example, has shown that adolescent sex differences on standardized mathematics tests increase as ability (measured by GPA), increases.

It seems likely, then, that situational factors beyond one's general intellectual ability may impact performance in high-stakes conditions, and that these factors affect mathematically talented women to a greater extent than they do the less talented. While females who distance themselves from certain subject domains in secondary school may feel it has little importance for their personal identity, females who reject social stereotypes and continue to participate in a stereotypically-masculine domain—those who elect to major in these areas or enter careers in male-dominated fields, for example—may find their self-images threatened when confronted by a stereotyped task in high-stakes situations.

Stereotype Threat Theory

In 1992, Claude Steele reasoned that the "extra burden" of negative social stereotypes creates an added stressor for minorities in academic settings. First described in 1995 by Steele and Aronson, "stereotype threat" is the fear of confirming, either to the self or others, a negative stereotype about a social group to which one belongs (Steele & Aronson, 1995); it is a "situational threat," experienced in the specific domain to which the stereotype applies (Steele, 1997, p. 614). As Steele (1999) notes, because prior experiences and socialization histories impact how events are approached and interpreted, targets of negative stereotypes may react to a seemingly “objective” situation differently than would non-targets. Targets may fear being judged by the stereotype or confirming it to be true in any area for which it applies. Because of the negative expectations about one’s ability that is associated with the stereotype, threat is thought to create an acute and debilitating anxiety that may impair task performance or altogether discourage participation in a domain, although targets may not be consciously aware of the process.

The stereotype threat concept therefore provides a straightforward explanation for at least part of observed sex differences in math performance as well as achievement differences among a number of other social groups (e.g., Steele & Aronson, 1995). In a study of sex stereotypes for math ability, Quinn and Spencer (1996) presented a difficult math test to male and female Calculus students, all of whom had roughly equivalent GRE math scores and course grades. The students were given either “diagnostic” instructions, in which they were told their mathematical abilities and limitations would be evaluated, or “non-diagnostic” instructions, in which they were told their problem-solving strategies, but not their individual scores, would be measured. Consistent with the stereotype, men outperformed women when participants were told their mathematical “limitations” would be assessed. In the less evaluative, non-diagnostic condition, however, women performed as well as men did (Quinn and Spencer, 1996). The belief that their personal ability would be evaluated may have increased female participants’ motivation, self-doubt, and anxiety to a greater degree than the same instructions did for males, though possible mediators were not tested.

Similarly, Spencer, Steele, and Quinn (1999) found that male and female participants, specifically selected for their superior mathematical skills, performed equally well on a set of math problems drawn from the basic GRE-Q exam using the standard instructions. Yet when participants were randomly assigned difficult problems from the more advanced Mathematics Subject exam, females performed significantly worse than males did. These results would seem to support the notion of a true difference in the skills of males and females at advanced mathematical levels. However, a second study using problems taken from the same advanced subject exam found that females scored as well as males when the test was first described as having been "shown not to produce sex differences" than when the standard instructions were used (Spencer et. al, 1999, p.12). Taken together, the findings from these two studies suggest that some factor other than sex-linked differences in ability may account for women's lower scores on advanced tests. In this case, the reassurance that a known negative stereotype did not apply to the task may have quelled participants' anxiety, but again, no mediators were tested.

A growing body of literature is developing around the stereotype threat construct. Several studies have replicated Steele and Aronson's (1995) findings (Aronson, et al., 1999; Inzlicht & Ben-Zeev, 2000; Spencer et al., 1999; Walsh, Hickey, & Duffy, 1999). Other studies have extended the results to additional target groups and performance domains. Stone, Lynch, Sjomeling, & Darley (1999), for example, extended the phenomenon to a psychomotor task, finding that performance was differentially impaired by stereotypes of sports abilities for whites and blacks. Croizet & Claire (1998) and Leyens, Desert, Croizet, & Darcis (2000) found support for the phenomenon in France, where stereotypes regarding intelligence and social class are prevalent. And Levy (1996) demonstrated that priming various stereotypes of aging could either depress or improve performance on memory tasks for elderly populations, depending on the valence of the stereotype. Other evidence in support of the stereotype threat explanation comes from cross-cultural research, which has found that group differences do not generalize to cultures where the tested stereotype does not exist (Levy & Langor, 1994; Schacter, Kaszniak, & Kihlstrom, 1991; Shih et al., 1999). The lack of a significant effect in these specific samples further suggests that a social stereotype, not any type of genetic predisposition, accounts for group performance differences.

This growing body of research supports the possibility that situational factors which bring to mind negative stereotypes can have a profound effect on performance. When a negative stereotype becomes accessible, any frustration experienced while performing a challenging task may make the stereotype, an allegation of one's potential inability, suddenly self-relevant. Feelings of frustration may increase the stereotype's applicability, making inherent limitation a more plausible explanation for one's struggle. Whether or not the stereotype is accepted as being self-descriptive or even generally true, almost everyone can name the particular stereotyped characteristics thought to describe a given social or ethnic group (Devine, 1989). In fact, Niemann et al's (1998) research shows how stereotypes can become internalized over time; African American participants frequently chose stereotypical terms such as "unintelligent" and "unemployed" describe their own race. Because of repeated exposure to stereotypes, these descriptions are automatic, coming to mind spontaneously (Bargh, 1994; Bulman-Fleming, Grimshaw, & Berenbaum, 2000; Devine, 1989).

Hypothesis 1: Men will outperform women on a difficult mathematical task under conditions of stereotype threat, while men and women will perform equally well in a condition that reduces stereotype threat by describing a task as sex-fair.

Implications for Employment and High-Stakes Testing

Findings in support of the stereotype threat effect have important implications for employment and educational settings because skill assessment plays a critical role in life-shaping determinations such as college acceptance, job placement, certification testing, employee selection, and job promotion. For example, cognitive ability tests are commonly used in making selection decisions because of their high predictive validity (Schmidt & Hunter, 1998); however, significant racial group differences on cognitive tests are repeatedly found (Bobko, Roth, & Potoksy, 1999; Sackett & Wilk, 1994). If these types of tests were used alone in making selection decisions, organizations would find it difficult to diversify the workforce. A solution is sought that has both high predictive validity and low adverse impact (Sackett, Schmitt, Kabin, & Ellingson, 2001). Researchers have continually sought alternate testing methods to decrease adverse impact (e.g., DeShon, Smith, Chan, & Schmitt, 1998), but have not been successful in significantly reducing subgroup differences. If stereotype threat does play a partial role in racial or sex differences in performance on standardized tests, ways of reducing threat might decrease the adverse impact of cognitive tests while maintaining their predictive validity.

Applied Studies and Conflicting Results

However, a number of applied studies and laboratory simulations of employee selection have failed to find an effect of threat-related factors on performance (McFarland, Lev-Arey, & Ziegert, 2003; Mayer & Hanges, 2003; Nguyen, O,Neal, & Ryan, 2003; Ployhart, Ziegart, & McFarland, 2003; Stricker, 1998; Stricker & Bejar, 1999; Stricker & Ward, 1998). Because of the implications that Steele and Aronson's (1995) findings have for high-stakes testing, it is important to consider the external validity of stereotype threat theory.

In the first test of stereotype threat in a true applied setting, Stricker (1998) investigated the role of demographic questionnaires (see Steele & Aronson, 1995) in implicitly inducing threat among women and African Americans taking the College Board Advanced Placement (AP) Calculus exam. As in Steele and Aronson's (1995) study, students were asked to indicate their race either prior to or after the completion of the exam. No differences in scores were found between Caucasians and African Americans or between male and female students in the two conditions. Of the 42 variables tested, only one (the number of questions left unanswered but not skipped) was significantly greater among females who were given the standard pre-test questionnaire. Given the large sample size (N = 1,652), the author interpreted this finding as not having any practical significance. Further research by Stricker and Ward (1998) examined the effect of demographic questions in inducing threat for freshmen taking computer-adapted college placement exams. No order effects of the demographic questionnaire were found for female or African American test performance.

Steele and Aronson (2002) argue that even the small effect found by Stricker (1998) is meaningful, since the manipulation was relatively weak compared to laboratory studies. Yet such strong manipulations used in experimental research do not typically occur in real-world testing situations. How much of a concern is stereotype threat to employment testing if conditions must be explicitly manipulated in order to produce the effect?

One problem that not been addressed to date is the natural level of threat that may occur under normal testing situations. Specifically, does a difficult mathematical test prime a negative stereotype in and of itself, or do stereotypes need to be explicitly called to mind in order for performance to be affected? While Stricker (1998) used only the standard testing instructions as a means of inducing threat, other researchers have stressed the evaluative nature of the experimental session (e.g., Steele and Aronson, 1995), described the experimental test as one in which men have outperformed women in the past (e.g., Spencer et al, 1999), or used a combination of several manipulations at once (e.g., Nguyen et al., 2003).

Steele & Davies (2003) believe that a difficult cognitive test alone is sufficient to prime negative stereotypes and produce threat, arguing that the applied studies involving racial differences failed to find significant differences because the experimental task was presented as a diagnostic measure in both conditions; therefore, stereotype threat was present in both the experimental and control samples. They argue that the Stricker (1998; Stricker and Ward, 1998) studies and laboratory simulations of employee selection (e.g. Ployhart et al., 2003) have a significant "design limitation… the absence of adequate no-stereotype-threat control groups" (p.312). In Stricker's (1998) research, for example, the demographic questions were presented before the test for both groups; the only difference was that they were blackened over in the experimental condition. For students accustomed to taking standardized tests, the omission of any demographic information would have been obvious, particularly given the fact that the usual place for reporting that information was masked over. In addition, participants in the experimental condition were told that the instructions for the exam were different because ETS was testing changes to the *answer sheet*. Steele and Davies (2003) note that "all of the experiments" that have compared "a control condition that presented the test as a test and an experimental condition that also presented the test as a test, but that added an extra degree of stereotype threat…. compare a condition that has aroused stereotype threat with a condition that aroused stereotype threat plus" (p. 315).

On the other hand, Sackett et al. (2001) imply that threat is not implicitly induced by normal testing instructions; the authors contend that because Steele and others have controlled SAT scores in their analyses, what is being reported in laboratory studies is not that non-diagnostic instructions "remove" pre-existing performance differences. Rather, the fact that Blacks and Whites perform equally in these conditions is what would be expected, given that their scores were statistically adjusted by using SAT as a covariate. Therefore, the correct interpretation is that manipulations used to induce threat in laboratory settings make any pre-existing differences in performance larger. Sackett et al. (2001) caution against "overinterpreting the findings to date, as they do not warrant the conclusion that subgroup differences can be explained in whole or in part by stereotype threat" (p. 307).

There are two important issues to consider regarding the experimental design used in previous research. One is the question of an appropriate control group. Steele and Davies (2003) define “control” as a condition in which threat is “removed” by describing a test as “gender fair” or “non-diagnostic” of ability. Yet, this is not a control in the true methodological sense; presenting a test as *not* having been shown to produce sex differences is a significant manipulation which may increase the performance of targets beyond normal circumstances or adversely impact the performance of non-targets. The other related issue is the method used for inducing threat. In some cases, performance under standard testing conditions is compared to performance in a condition that specifically stresses past findings of gender equity (what Steele calls “threat removed” conditions). In the other approach, the task is described as having been shown to produce sex differences or described as gender-fair. This approach, however, seems to incorporate two experimental conditions in one design with no control group. Indeed, the “no past differences” or “threat removed” condition is treated as a control condition in some studies (in which performance is compared to a “sex differences” condition), and the experimental one others (performance is compared to a true control condition in which sex is not mentioned). It is important to note that the underlying assumptions regarding stereotype threat are different in each case. In the latter methodology, stereotypes must be manipulated and distinctly called out in order to produce threat, while in the former, threat is assumed to operate on its own—only by calling attention to the finding of “no past differences” is it supposedly alleviated.

If standard instructions (true control) do not induce threat, then any findings of performance differences between males and females under these conditions cannot be accounted for by the theory. But if the outright induction of threat is required, then the stereotype threat concept may be a product of laboratory-induced manipulation and of little concern in applied settings.

While preliminary researchers have described stereotype threat as being “present” or “absent” in certain testing conditions (an “all or none” effect), it is most likely that threat is always present to some degree among targets, but certain manipulations make its effects stronger than others. The varying types of instructions used in research may increasingly induce threat and feelings of anxiety along a continuum. That is, threat for females may be greatest under the manipulated conditions of known “past differences,” lower under standard (control) conditions, and lowest under the threat-removed, “no difference” conditions. If self-efficacy and anxiety account for the effects of threat on performance, anxiety could be presumed to be debilitative under the past differences condition, in which self-efficacy is reduced, and facilitative under the “no differences condition,” in which self-efficacy is increased. An alternate hypothesis, posited by Steele and Davies (2003), is that there will be no significant difference in performance between implicit (true control) and explicit (experimental manipulation of threat) situations because stereotype threat is present and equally detrimental to performance in both conditions.

Therefore, a research design is needed that incorporates all three conditions of sex differences: one in which participants are told that men and women perform equally, that men outperform women, and a control condition in which sex differences are not mentioned. This design will allow for comparisons between a true control condition in which sex is not mentioned and the two experimental conditions in which threat is both "removed" and induced.

In their study of racial differences in verbal performance, Steele and Aronson (1995) did measure performance under diagnostic, non-diagnostic, and control conditions. However, planned contrasts were only made between the non-diagnostic and diagnostic conditions and between the diagnostic and all other conditions. The possible difference between the control and non-diagnostic conditions was not analyzed. It is possible that the non-diagnostic condition may increase performance beyond the control condition in which the standard instructions are used. In fact, using a third facet of the "sex differences" instructional set—describing a task as having been shown to favor women *over* men—may actually work to increase female performance to a greater degree than the "threat removed" conditions in which men and women supposedly perform equally well.

Only one study to date has compared performance under three differently valenced conditions within the same experiment (Shih et al., 1999), and the data supports this hypothesized trend. Asian females were presented with a seemingly unrelated pre-task which either increased the salience of their sex or Asian heritage, then given a difficult math test. Performance in the "Asian" salient condition was the greatest, followed by participants in the "neutral" (no priming) condition, while subjects in the "female" salient condition showed the lowest overall performance. Shih et al.'s (1999) research is important because it shows the benefits to performance that can be associated with a positive stereotype. Priming Asian heritage among Asian-American females prior to a high- stakes testing situation would likely be effective in combating stereotype threat and increasing their performance; however, the manipulation would only be effective for a small subpopulation of American females. A similar manipulation is needed that has a broader applicability to all women that could be implemented in such a way as to avoid any adverse impact on males.

Experience of Threat by Males

Researchers have called for utilizing threat-removed manipulations (i.e., stressing no sex differences or non-diagnosticity of the test) in applied settings as a means of reducing threat, yet presenting a test as not having been shown to produce sex differences is a significant manipulation which could adversely impact the performance of non-targets. As previously described, Spencer et al. (1999) found that males outperformed females when sex was not mentioned, yet males and females performed equally on a difficult task when the task was described as showing no previous sex differences. It is possible that the manipulations either increased female performance or decreased male performance, or that a combination of the two created the significant difference. Two studies to date have shown that a culturally dominant group can experience stereotype-related reductions in performance. Aronson et al. (1999) found that the performance of Caucasian males decreased when they were presented with a stereotyped message regarding the superior math abilities of Asian Americans; a similar experiment by Leyens et al. (2000) found that males who were presented with a stereotype about females' supposed superior social empathy skills performed significantly worse on a verbal task involving affective words than males did in the control condition (Leyens et. al, 2000). Yet no study to date has investigated the possibility that presenting information *counter* to a popular stereotype could impact performance of the dominant group.

Brown and Josephs (1999) suggest that mathematical stereotypes are experienced differently by males and females. While men focus on the positive implications of the stereotype for their ability, women focus on the negative implications. Men are concerned with displaying their stereotypical superior ability, thereby confirming themselves as an example of the positive stereotype, while women are concerned with *avoiding* a display of a *lack* of ability, thereby focusing on dispelling the stereotype as applicable to themselves. In their study, half of the participants were told that performing below a cutoff point on a set of GRE-Q problems could diagnose inherently weak math abilities, while the other half was told that performing above a cutoff could identify exceptionally superior abilities. Participants were also told if they scored above (or below, depending on the condition) that cutoff point, they would be classified into an "undetermined" category, in which they could be below average, average, or even above average. An ANOVA found that women in the "weak skill" identification condition scored significantly lower than those in the "exceptional skill" identification condition, while the reverse pattern was found for males: those in the "exceptional skill" identification condition scored significantly lower than those in the "weak skill" identification condition. Thus, it is likely that threat can also be experienced by males, but in an opposite direction. While women may fear *confirming* the negative stereotype about their ability, men may fear the possibility of *not confirming*—i.e., not "living up to"—expectations associated with the positive stereotype. It is possible that conditions which describe math tasks as having shown "no sex differences" may actually depress male performance. But because males are not direct targets, however, the experienced threat may have a smaller effect on performance than it does for women. Several studies support the general trend (Schmader, 2002; Spencer, Steele, & Quinn, 1999; Quinn & Spencer, 2001), though analyses of male performance were either not tested or not reported. For example, Quinn & Spencer (2001), found that men's performance in the "sex fair" condition was impaired to a greater degree than in the standard instruction condition; although only a slight difference, men could not effectively formulate a strategy for 9% of the items in a problem-solving measure as opposed to only 2% of the items in the standard instruction conditions. It is also interesting to note that Spencer et al. (1999) found that men in the "no sex difference" condition spent more time on the test items than women did, suggesting that the "no difference" instruction may have impaired their task self-efficacy or increased anxiety, leading to hesitation in forming their answers.

Therefore, it is expected that describing a test as favoring a particularly unique female skill will increase female performance as compared to a control condition, while it will decrease male performance as compared to a control condition. An alternate hypothesis is that the mere mention of sex may counteract the positive message of females' supposedly superior skills. Mentioning sex in the presence of a difficult math task may prime the stereotype and create threat for females regardless of the context; the schema of women's inferior math skills may be so culturally ingrained that once the stereotype is primed, it may trigger feelings of threat, lower self-efficacy, and increase cognitive interference in women, leading to reduced performance. Research on the ideomotor effect by Bargh (e.g., Chen & Bargh, 1997) supports the notion that an automated response may follow once a schema is primed, particularly when cognitive load (such as difficult task) is present.

Hypothesis 2: Females in a condition that describes a task as one that favors women will outperform females in a control condition in which sex is not mentioned, while females in a condition that describes a task as one that favors men will perform worse than females in a control condition (in which sex is not mentioned).

Hypothesis 3: Males in a condition that describes a task as one that favors men will show an increase in performance compared to males in a control condition in which sex is not mentioned, while males in a condition that describes a task as one that favors women will perform worse than males in a control condition (in which sex is not mentioned).

Proposed Model

In order to produce feelings of threat, the stereotype must first be accessible and applicable to the target. Accessibility can be directly manipulated by reminding participants of the stereotype, or stereotypes can be called to mind by more indirect routes, such as priming one's cultural/sex group or providing a stereotype-relevant task. Negative stereotypes threaten self-esteem, creating both physiological arousal and psychological stress, evidenced by negative emotions such as fear and anxiety. Feelings of threat may result from a universal motivation to maintain a positive self-image and project this image to others. Chronic exposure to stereotypes likely creates a stronger association with the stereotype among the targeted group, and those with a greater accessibility of the stereotype will experience threat to a greater degree, as evidenced by a more significant decrease in performance. Negative stereotypes may increase motivation to perform well and dispel the stereotype, but for targets who identify with the domain, threat may create an "overmotivation" effect which may impair performance.

In order for threat to impact performance, the task must be difficult enough to create a sense of challenge and frustration. It is likely that stereotypes impact self-confidence, either raising or lowering one's expectancies for task performance, depending on the valence of the stereotype. Thus, when frustration is experienced with a difficult task, a positive stereotype would boost self-efficacy and protect the self-image from threat, leading to persistence, goal-striving, and task-engaging behavior in times of stress. A negative stereotype, in contrast, may lead to self-doubt, worry, and task-irrelevant rumination when frustration with task performance creates anxiety. While women may fear confirming the negative stereotype about their math ability, men may fear the possibility of not "living up to" expectations associated with the positive stereotype. By stating that a mathematical or spatial test is favors females, the beneficial effect to state self-efficacy is removed for males, but may increase state self-efficacy for females as compared to control conditions.

Tests of Mediation: Research to Date

One of the largest difficulties facing stereotype threat theory is a lack of understanding of how certain factors that may mediate the effect of threat on performance—mediational studies have produced either non-significant results or results that appear to contradict one another (Aronson et. al, 1999; Aronson, Quinn, & Spencer, 1998; Mayer & Hanges, 2003; McKay, 1999; Spencer et. al, 1999; Stangor, Carr, & Kiang, 1999; Steele & Aronson, 1995). One limitation in several of the studies may be the measures used, most of which consisted of only one to five questions developed for the study and without mention of any investigation of their validity or reliability. Another problem is the very small sample sizes used, most less than 10 cases per cell. Such small samples make finding and interpreting experimental effects difficult, especially any interaction or mediation between constructs. But more importantly, studies of mediation have focused primarily on self-efficacy, evaluation apprehension, or anxiety as single mediators and have not investigated the relationship among the hypothesized constructs within a single integrated model. It is likely that stereotype threat may operate via multiple routes to disrupt performance.

Steele (1997) has hypothesized that that decreased performance is created by an inefficiency of information processing, but his model of the threat-attention relationship is not precise. In Steele and Aronson's (1995) study, for example, the authors report that

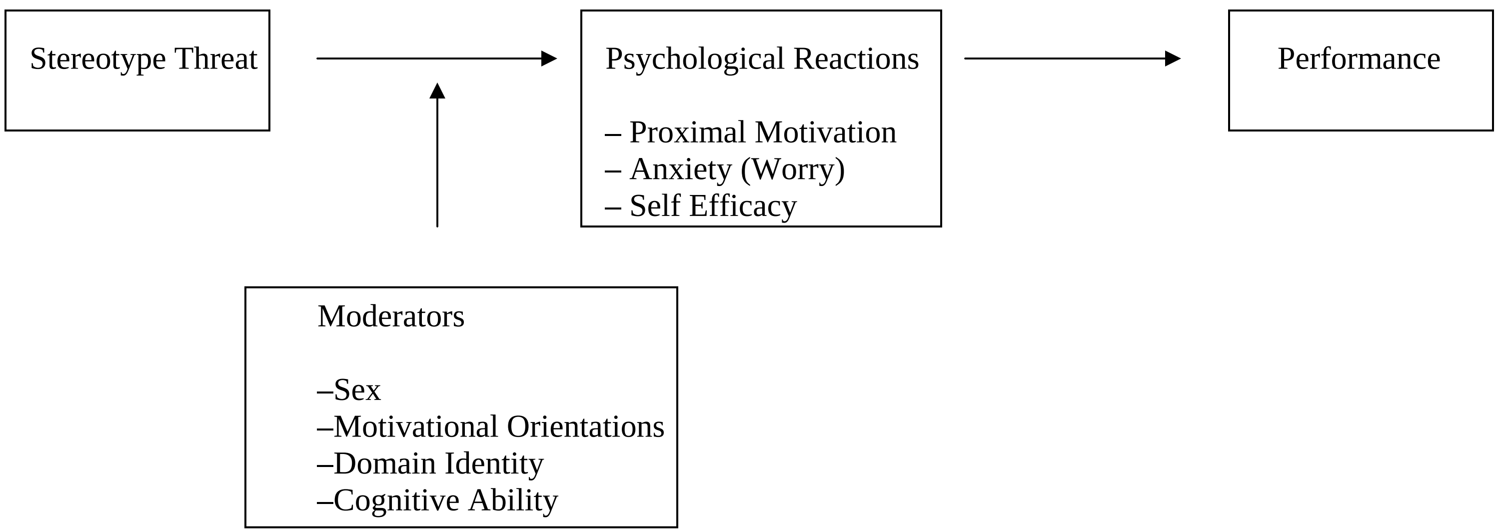
African American targets in the diagnostic condition spent more time on each math question and completed fewer items with less accuracy than Caucasian and African American participants in the control condition did. Threatened targets also showed a greater tendency to re-read the test items. What is not known is why participants answer this way—whether threat decreases task self-efficacy, decreases task-focused attention by increasing anxiety and emotion-laden thoughts, decreases task focus by increasing evaluation apprehension and performance monitoring, or increases motivation to dispel the stereotype, thereby increasing self-regulation or changing one's strategic approach to the task. In the model depicted below, the threat-performance relationship can be mediated in several different ways through cognitive, affective, and/or motivational mechanisms.

Proposed Model

Theories of individual differences have typically differentiated trait-like distal constructs that have an indirect impact on behavior from proximal states that have a more direct influence. While distal constructs such as personality traits are assumed to be relatively stable over time and across situations, state-like individual differences can be task- or situation-specific and will vary over time in reaction to environmental factors. Trait-like individual differences are thought to impact performance indirectly through their relationship with state-like constructs.

The integrative model depicted below outlines the hypothetical relationships among distal individual difference moderators and state-like mediating factors in the stereotype threat process.

Figure 1. Proposed Model

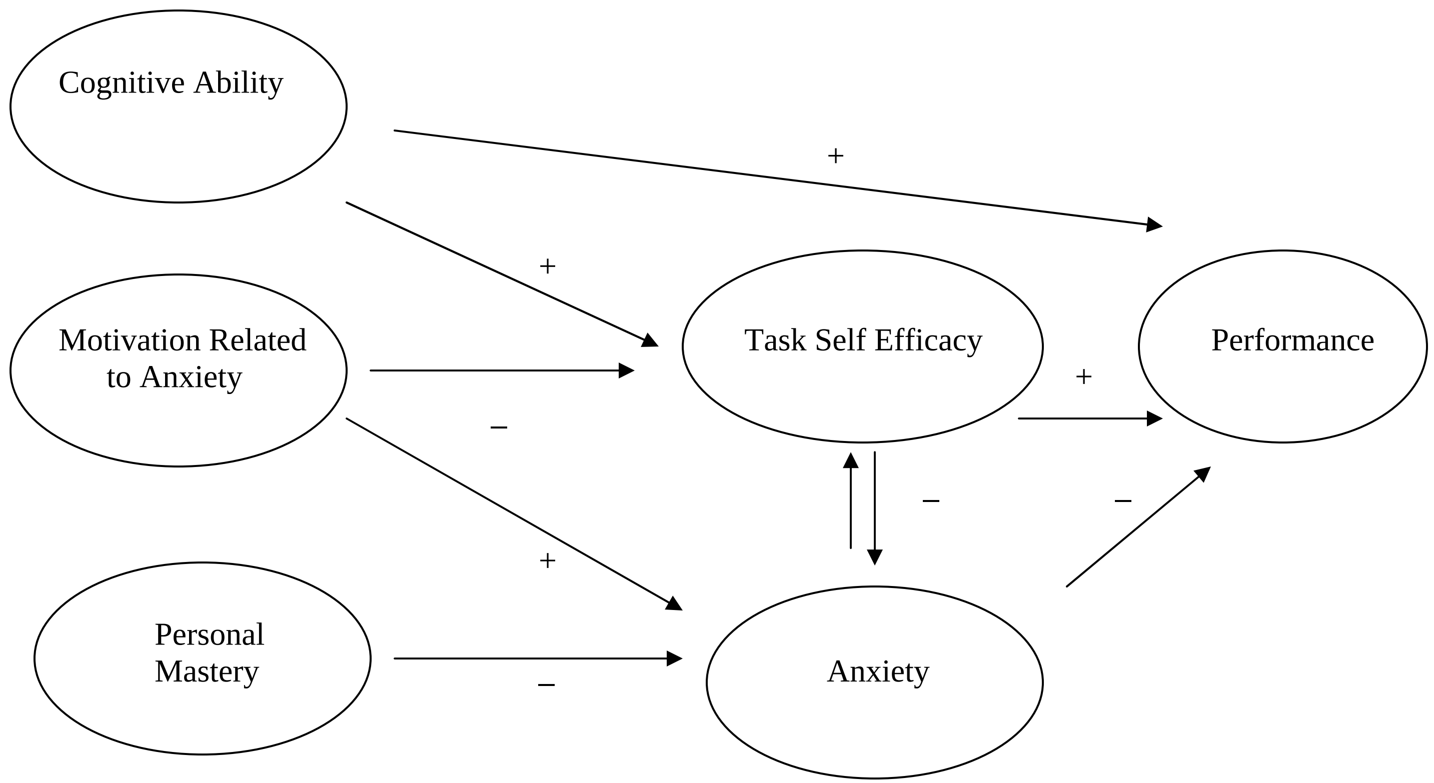


First, experimental manipulations bring the stereotype to mind. This may be the result of direct statements, subliminal priming, or any factor in the environment that increases the accessibility of the stereotype. Individual differences such as target status (see Aronson & Disko, 1998), stereotype susceptibility (e.g., Foote, 2000), or identification with the stereotyped group also impact the accessibility and applicability of the stereotype.

If the stereotype is positive, accessibility will increase state task efficacy; if it is negative, state self-efficacy for the task is threatened. Individual differences in general cognitive ability impact task self-efficacy as well, including how the stereotype is interpreted—e.g., whether or not it is identified as a threat to one's self-esteem. If the task is interpreted as difficult, state anxiety, particularly worry, will reduce the cognitive resources available for the task and impair performance.

Identification with the domain also increases task motivation. A sense that something important is at stake, in terms of evaluation apprehension and desire to display a favorable impression, may raise self-set task goals and increase regulation of one's performance. The impaired task efficacy and increased pressure to "look good" may cause targets to adopt task strategies aimed at avoiding mistakes rather than displaying skill. Speed may decrease in hopes of increasing accuracy, and targets may feel an increased need to re-read items or re-check their work. Increased motivation to dispel the stereotype and create a positive impression also increases self-regulation and tendency to evaluate one's performance in comparison to a desired standard, diverting necessary resources from the task. In addition, individual differences in motivational orientations such as fear of failure versus achievement motivation impact how the threat is interpreted and the task is approached.

Figure 2. *Hypothesized Paths of Correlations Among Variables of Interest*



State Constructs: Possible Mediators of the Threat-Performance Relationship

Motivation

Stricker (1999) points out that lowered motivation for standardized tests taken in laboratory settings is well-documented in education research (Brown & Walberg, 1993; O'Neil et al, 1996), arguing that participant motivation was higher in his applied research because of the implications of students' AP scores for college credit. Because the number of items skipped or not attempted by stereotyped targets was greater in Steele's research (12/27 items = 44.5% versus 13/40 items = 33%), and planned contrasts found that African American participants in Steele and Aronson's (1995) diagnostic condition also answered more slowly, Stricker (1999) reasons that African Americans reacted to threat by lowering motivation and psychologically removing themselves from the study. Likewise, Chan et al. (1998) found that racial differences in attitudes about cognitive tests were associated differences in motivation, which was found to mediate the relationship between attitudes and performance.

Yet most laboratory studies of stereotype threat, whether racial or sex difference studies, have found support for *increased* motivation among the stereotyped participants (Steele & Aronson, 1995; Harder, 1999; Aronson et al., 1999). Effort has been measured in several different ways: how long people persist on a task, the number of problems attempted, and self-report. None of these methods have led to support for the idea that threat reduces effort. Even Aronson & Salinas's (2001) elaborate technique, in which electrodes that could supposedly measure effort were tied to participants' wrists, failed to suppress the effect of threat on performance. Decreased performance still emerged despite participants' being told they would have to repeat the task if they gave an insufficient level of effort.

One factor that differentiates stereotype threat studies from Chan et al. (1998) and similar studies involving decreased motivation among minority participants is the level of identification with the task. Stereotype research involves high-ability participants who are strongly identified with the task domain. It is typically an important aspect of their self-concepts, and performance outcomes are linked to their self-esteem. Simply put, targets of the stereotype in threat research have something to prove.

While several studies have found support for increased motivation under conditions of threat, its role as a mediator has not been tested. Harder (1999) found that highly-identified females in an advanced mathematics reported higher levels of motivation in threat conditions, but no significant differences in performance were found. Aronson et al. (1999) found that participants in the threat condition reported expending more effort on the problems than did controls, but an ANCOVA controlling for effort did not affect the difference between mean GRE scores. One reason for the lack of consistent support may be the measures used to quantify level of effort in past studies—typically, a self-report, single-item question—that have not accurately captured the frequency or intensity of self-regulation. Another is the level of attention given to current motivation theory and research.

Motivation is thought to have both distal and proximal components. Distal processes occur prior to task engagement and refer to choice behaviors such as goal selection. Proximal processes, on the other hand, control task-striving behavior once it is engaged; they determine the distribution of effort across on- and off-task activities during performance. Naylor, Pritchard, and Ilgen (1980) propose a resource allocation model that operationalizes motivation as the proportion of one's total cognitive resources directed to goal-related activity at any one time. Most contemporary theories incorporate the basic tenets of Naylor et al.'s (1980) dynamic resource allocation process. Kanfer and Ackerman (1989), for example, propose a cognitive resource theory of attention and performance, whereby performance is a direct function of task demands, ability (defined as the total amount of one's cognitive capacity), and motivation (defined as the proportion of one's total attention that is allocated to the task). According to their model, performance will not decrease as additional cognitive resources (i.e., attention) are devoted to the task. Once complete cognitive resources are allocated, performance will plateau at a level determined by ability and task demands. However, as attention is drawn away from the task, performance will decrease proportionally.

Self Regulation

Because attention can be divided among any number of several competing stimuli and activities, proximal motivational processes ensure that effort is both directed towards the task and maintained at an appropriate level over time. Self-regulation is a system by which higher-order mental activities guide the allocation of effort during task performance. It is thought to be composed of three main activities (Kanfer, 1990): The first is o*bservation*, in which attention is given to the specific aspects of one's behavior. Observation of one's own performance provides the individual with knowledge about task progress. *Evaluation*, the second element, is the process of comparing present performance to a referential standard or desired goal. Evaluation directly influences *self*-

reaction, the third component, which is an affective response to current performance in which persons either feel satisfied or dissatisfied.

Arousal and Cognitive Interference. If the domain is important to the target of a stereotype, anticipated evaluation threatens self-esteem and energizes the individual. Arousal triggers proximal motivation to assess one's performance in order to improve work output and ensure goal achievement. But as the Yerkes-Dodson model shows, increased arousal increases performance only up to a certain optimal level. Beyond that, greater arousal is often associated with performance decrements (Teigen, 1994).

Because individuals cannot simultaneously attend to all aspects of their behavior, only goal-relevant dimensions should draw attention in order to optimize performance. Yet the engagement of proximal motivational processes requires attentional effort. If frequent or intense regulation uses too much cognitive capacity, performance may decline, as working memory necessary for processing task-relevant information is not readily available. Kanfer & Ackerman (1989), for example, show how goal setting, which involves performance monitoring and comparisons between current and desired performance levels, can detract working memory and ultimately decrease performance for novel or difficult tasks. Atkinson (1974), furthermore, found that achievement-motivated persons are more likely to suffer performance decrements than those low in motivation, as increased drive may also increase attention to one's own performance and divert needed resources from the task.

In support of this idea, Steele and Aronson (1995) found that African-American targets of a negative stereotype spent more time on each math question and completed fewer items with less accuracy than Caucasian and African American participants in the non-threat condition. African Americans in the threat condition also reported higher levels of effort and a tendency to go back to re-read the items more than participants in any other condition did, though the ratings of effort and tendency to re-read items did not correlate with performance. Quinn & Spencer (2001), likewise, found support for increased arousal and impaired problem solving ability among threatened female participants. Although the interaction between sex and stereotype threat had only marginally significant effect on performance, a significant interaction was found for self-reported ability to formulate an effective task strategy. Women in the standard threat conditions were less able to form a strategy (an average 14% of the time) than those in the sex fair condition (4%) and men in standard conditions (2%), but the possible mediation of the threat-performance relationship was not explored.

> *Hypothesis 4*: Participants under conditions of threat (sex differences favoring the opposite sex) will show more cognitive interference than non-threatened participants (control condition and sex differences favoring same sex).
>
> *Hypothesis 5*: Cognitive interference is expected to correlate positively with threat and negatively with performance and should mediate the threat-performance relationship.

Task (State) Self-Efficacy

Over the past five years, researchers at Carnegie Mellon University have undertaken a series of in-depth interviews of undergraduate computer science students, both those enrolled in the computer science program and those who began the program but changed majors. The researchers have found that, while almost all female students report entering the major with interest and enthusiasm, positive feelings quickly wane as they interact with males whom they perceive as having more advanced, self-taught skills (Margolis et al., 2000). Similar findings are reported for female engineers. The Goodman Research Group (2002) found that even though 66% of women who changed majors out of engineering had either an A or B average, they typically compared themselves negatively to male peers, whom they perceived as having a better understanding of concepts and a higher level of ability. The greatest determining factor in attrition was not course grades but level self-confidence—females with low self-confidence were more likely to leave an engineering program, while those with high self-confidence were more likely to remain.

It is likely that task self-efficacy impacts how information in the environment is interpreted; those low in self-efficacy may use situational cues to diagnose a given situation as threatening. Mischel and Shoda (1995) propose that personality systems are composed of a series of complex schema for different types of situations, including task-specific self-efficacy. Those high in self-efficacy for a situation may approach it within a positive psychological state, which guides attention to positive aspects of the situation, creating an opportunity frame. Those low in self-efficacy, on the other hand, may attend to the negative aspects of a situation, forming a threat frame. In support of this idea, Mohammed & Billings (2002) manipulated self-efficacy by having participants work a hypothetical restaurant management task and then providing fictitious feedback. Participants then rated the level of threat or opportunity for a second management scenario. Those in the high self-efficacy condition rated the scenario as higher in opportunity than those in the low self-efficacy condition, but no significant differences were found for ratings of perceived threat.

If stereotype threat manipulations impact task efficacy, it is logical that state self-efficacy would mediate the effect of threat on performance, but mediational studies present a mixed picture. Steele and Aronson (1995) found that African Americans in the diagnostic (threat) conditions reported more negative subjective assessments of their performance, measured post-task, than participants did in any other conditions. While these subjective ratings may actually reflect accurate assessments of lower comparative performance (African American students in the threat conditions did, in fact, achieve lower test scores than those in the other conditions), the students also completed a greater number of word fragments with words relating to inferiority and failure. Likewise, Stangor, Carr, & Kiang (1998) found that women in non-threat conditions predicted future performance based on positive or negative task feedback, but performance expectancies for those in the stereotype threat condition did not differ as a function of prior feedback; expectancies were low regardless. This supports the idea that stereotype threat can undermine performance expectancies, but it is not known whether decreasing stereotype threat by framing a task as showing "no past differences" can actually *increase* expectancies for women as compared to a true control condition in which sex is not mentioned. Nor does it support the mediational role of self-efficacy.

Expectancies for task performance in the Stone et al. (1999) study were significantly related to performance and were lower in the threat condition, but no interaction effects or tests of mediation were reported. A significant correlation between expectancies and performance was found, but the direction of the relationship was not clear. Because the research involved a series of task performances, expectancies could have been based on prior performance rather than the effect of stereotype threat. Brown (2000) failed to find any effect of experimental manipulation or race on performance expectancies, and Spencer et al. (1999) found that manipulations had no effect on expectancies, which had no effect on performance. Thus, there is no clear and consistent evidence of performance expectancies as a mediator in stereotype threat.

One reason for the lack of consistent support for self-efficacy as a mediator may be the measures used. Steele and Aronson (1995) found that self-reports of “academic competence” (i.e., “I feel confident about my abilities”) and “personal worth” (i.e., “I feel as smart as others”) were similar across participants in all conditions; they concluded that participants in perceived threat conditions did not suffer reduced efficacy for the task. But reported feelings of *general* ability or “worth” are not equivalent to perceptions of task-specific self-efficacy. State self-efficacy is more than just an assessment of one’s own cognitive or physical ability; it includes an orchestration component—the capability to allocate resources, regulate behavior, and integrate skills successfully in order to achieve a specific level of performance on a given task (Bandura, 1986; Bandura, 2000). A measure specific to the task of interest is needed, one that will capture participants’ belief in their ability to successfully allocate resources, regulate behavior, and integrate skills that are used specifically in the field of math and science (Bandura, 2000).

Self Efficacy and Self Regulation

If threat works to reduce task self-efficacy, it is likely that priming a negative stereotype influences whether targets evaluate their own performance as evidence of a lack of ability or sign of impending failure, while targets of a positive stereotype would attribute any suboptimal performance to other factors (such as lack of effort or poor strategy) and increase strategic thinking and goal striving. When goal discrepancies are detected by proximal processes, efficacious people are more likely to change their behavior and persist at the task rather than to give up or settle for lower performance. They are more likely to attribute failure to a lack of effort or poor strategy or unfavorable circumstances, while those low in efficacy are more likely to attribute poor performance to a lack of ability and visualize impending failure. In fact, the causal effects of attribution on achievement-related behavior are mediated almost entirely by efficacy beliefs (Bandura, 2000).

It is also likely that low self-efficacy is associated with increased proximal motivation, resulting in greater self-regulation of performance and cognitive interference due to worry. Those with a high sense of self-efficacy visualize success, which provides positive guides for performance, while those low in self-efficacy for the task visualize failure and dwell on possible scenarios in which things could go wrong. If efficacy is low, one may be motivated to observe one's own performance closely and "check and re-check" one's work. Feelings of self-doubt may increase hesitation, which impairs the initiation of cognitive control systems, and greater effort is required to ensure attention is focused efficiently (Kuhl, 1994).

Self-Efficacy and Task Strategy

Decreased self-efficacy may also impact task strategy in that dissatisfaction with current performance may cause targets to question task strategy and begin to try and apply less optimal ones. Bandura and Jourden (1991), for example, found that efficacious people exhibited greater analytic and problem-solving thinking across progressive trials than those who reported low task self-efficacy. Path analyses showed that self-efficacy enhanced future performance directly as well as indirectly through its impact on strategic thinking.

Higher self-efficacy may provide the confidence to break from convention to use short cuts and estimations, while lower self-efficacy could be linked to a tendency to follow prescribed methods. This difference in self-efficacy could explain Quinn & Spencer’s (2001) finding women in threat conditions guessed less and used an unconventional strategy less often than women in the control group. Steele and Aronson (1995) also found that Blacks guessed less when race was primed, suggesting they were less confident in their ability, while Whites guessed more when race was primed. As noted above, Blacks in the ability-diagnostic conditions reported re-reading questions, answering fewer total questions, and working at a slower rate than participants in other conditions did, which may be indicative of a “mistake avoidance” strategy.

> *Hypothesis 6*: Participants in conditions of threat will display lower task self-efficacy than participants in control conditions; participants in “benefit” conditions (in which the task is described as favoring their sex) will display greater self-efficacy than those in control conditions.

Hypothesis 7: Task self-efficacy is expected to correlate positively with performance and negatively with stereotype threat; self-efficacy should mediate the relationship between threat and performance.

Anxiety and Cognitive Interference

Anxiety is an aversive emotional state of distress. It is thought to consist of two primary components: *worry*, the cognitive reaction to one's performance, including self-criticism and concern over the implications of failure, and *emotionality*, which is the affective component (Spielberger, 1985). People who are anxious about taking a test display more cognitive interference, particularly thoughts related to worry, than those who are not. Worry has been shown to relate to decreased performance, while emotionality has not (Deffenbacher, 1980; Hong, 1999; Tryon, 1982). Worry is thought to reduce task-focused attention, interrupting efficient reading, encoding, and retrieval of information. Anxious subjects, for example, have been shown to be less able to discriminate among the subtle differences in multiple answer choices and assign category labels in a significantly less organized manner than non-anxious subjects (Sarason et al., 1986; Seta, Seta, & Wang, 1988). Anxiety may increase distracting thoughts, which taxes cognitive capacity and impairs one's ability to allocate attention effectively. Emotion control skills used to avoid ruminative thinking require additional cognitive resources, further diverting working memory.

Steele (1997) posits that stereotype threat reduces performance by increasing anxiety rather than lowering self-efficacy for the stereotyped domain, yet he does not investigate or describe the relationship between the two constructs. Emotional states have been strongly linked to state self-efficacy, with positive states resulting in positive self-efficacy, and negative states, particularly anxiety, associated with reduced self-efficacy. Anxiety has been shown to be a strong predictor of state self-efficacy, which serves a mediational role in the relationship between anxiety and performance (Bandura, 1991). Dykeman (1994), for example, found that low self-efficacy students experienced higher levels of anxiety during an in-class test than high self-efficacy students did. It is likely that stereotyped persons, reminded that people like themselves struggle with tasks in some domain, will begin to doubt their ability and perceive the task to be more difficult than those to whom the stereotype does not apply; as frustration is experienced, the stereotype becomes subjectively causal for one's struggle.

Self-efficacy is proposed to interact with the valence of the stereotype to facilitate performance if the person expects success, but to adversely affect performance if failure is expected. Robinson-Staveley and Cooper (1990), for example, found that performing a computer task in the presence of others differentially affected expectations for success and consequent performance in males and females. While there were no significant sex differences for participants performing alone, males who played a computer game in the presence of others showed *increased* performance, while females' performance *decreased* in the public, as compared to the solo, condition. The effect on performance was moderated by sex differences in expectations for success: Men in the public condition reported positive task self-efficacy, while women in the same condition reported negative performance expectations for the same game. Furthermore, women in the public condition reported being more anxious and expressed more negative attitudes about using computers than did women working alone. Men in the same public condition reported a greater sense of confidence and less negative attitudes towards computers as compared to men who worked alone. Thus, for certain sex-stereotyped tasks, social facilitation may occur when one's sex role is salient and consistent with the task. When the situation is inconsistent with one's sex role or a negative stereotype is primed, however, anxiety may be induced.

Moderators of Stereotype Threat: Individual Difference Constructs

Cognitive Ability

In order to perform a new or challenging task successfully, the necessary attention must be devoted to it. Although individual differences in working memory exist, it is presumed that people possess a finite amount. Those with greater cognitive ability, therefore, are able to devote greater attention to a given task than those with less. General cognitive ability (g) has been shown to be a significant determinant of performance on a wide variety of tasks, ranging from vocabulary, mathematical reasoning, and spatial ability tests on the one hand to training success and work sample performance on the other (e.g., Hunter, 1986). As task complexity increases, the correlation between g and performance increases. People with higher cognitive ability are able to perform a greater number of cognitive operations more quickly, hold a greater number of concepts in working memory simultaneously, are better able to recognize and apply problem-solving rules, and learn new procedures more quickly than those with low cognitive ability. Research also supports state self-efficacy as a partial mediator of the ability-performance relationship such that those with greater cognitive ability also tend to have higher levels of task self-efficacy (Bandura, 2000; Chen, Gully, Whiteman, & Kilcullen, 2000).

Domain Identification

Identification with task domain is expected to moderate the threat-performance relationship. One central tenet's of Steele's theory is that the individual most care enough about the task domain to feel his or her self-esteem threatened by a potential evaluation. This idea is well supported by research. Aronson et al. (1999), for example, found that threat effects were strongest among participants who were most identified with the task domain. Yet research to date has not clearly defined how identification relates to performance. Though contrary to their original hypothesis, Ployhart et al. (2003) found that domain identity influences performance indirectly through its impact on test-taking motivation. The authors state that this finding is contrary to stereotype threat theory. However, it is not necessarily contrary to theory; rather, theories of threat have not explicated how identification operates to disrupt performance. It is likely that strongly-identified individuals, those whose self-concept is associated with the ability to perform well in a given domain, are highly motivated to perform well under conditions of threat in order to counter the stereotype and maintain their self-esteem. This increased motivation may be detrimental, however, if too much energy is spent monitoring one's own performance. For example, Aronson et al. (1999) found that high-math identified males outperformed moderately-identified males in a control condition when not confronted with a stereotype alleging inferiority. But under conditions of threat, the opposite pattern emerged: The moderately-identified group outperformed the highly-identified group. It is likely that priming a negative stereotype heightens motivation, which can be beneficial to moderately-identified participants, engaging a functional level of self-regulation and task-focused attention. But among those who see math as an important element of the self-concept, the threat to self-esteem may be great enough to trigger a debilitating level of performance regulation.

Identification with the task domain may be key in explaining why certain studies fail to find the threat effect on performance. Oswald and Harvey (2000), for example, found that women tested in the presence of a threatening cartoon actually outperformed those in the control condition, supporting the notion of increased motivation under conditions of threat, but not of performance decrements. The lack of a significant effect may be due to the fact that Oswald and Harvey's (2000) participant sample included more average ability students than previous threat studies and did not measure identification with the task domain.

Motivational Orientations

Another distal construct that is a possible moderator of threat is motivational orientation—a traitlike difference in goal striving to either achieve success or avoid failure that is thought to impact state motivation and behavior through its impact on goal choice. Atkinson (1957) first termed two specific motivational tendencies as "achievement motivation" versus "fear or failure" to describe an individual's likelihood to take risk based on whether they were motivated to succeed at a task versus avoiding failing at it. Higgins (1987) uses the phrases "approach" versus "avoidance striving" to describe the different ways people react to discrepancies between their actual and desired goal states, while Kuhl (1994) frames similar behavioral tendencies as either "state" or "action" orientations.

Achievement Motivation. People characterized by achievement motivation have an internal drive to achieve or accomplish difficult or important tasks because of the satisfaction and sense of pride that result. Achievement-motivated individuals often commit substantial resources and persist at a task for long periods or in the face of possible failure. They are also more likely to make internal attributions for performance and believe in an internal locus of control—that individuals are responsible for their own choices and behaviors. Achievement-oriented individuals view goal-striving behavior (such as working long hours at high effort levels) as instrumental because they associate persistence with eventual success. Demanding challenges, furthermore, are often interpreted as opportunities to display talent.

Fear of Failure. The dominant response of people with a fear of failure orientation, on the other hand, is a drive to avoid failure. These individuals focus on the difficulty of the task rather than the rewards associated with it; they equate task persistence with great risk because effort can potentially be unrewarded. The sense of risk and uncertainty create a self-protective drive to avoid task engagement in order to avoid failure. Intense fear of failure inhibits enthusiasm for achievement and may become a self-fulfilling prophesy. Heightened evaluation apprehension can increase anxiety to the point that it interferes with performance on difficult tasks. People high in fear of failure adopt strategies to avoid being personally accountable for failure, such as self-handicapping, procrastination, or the adoption of a compensatory goal. Because of the risk of possible failure, goal-striving behavior is experienced as particularly stressful. Individuals with a high fear of failure are also more likely to view skills as being fixed rather than amenable to training or experience.

Learning Versus Performance Striving. Similarly, Dweck (1985) describes two distinct individual types of chronic goal striving. Those who adopt *learning* goals seek to increase competence, gain knowledge, or master a new task. This orientation is characterized by challenge-seeking behavior and persistent, high levels of effort in the face of obstacles. People who adopt a mastery goal are more likely to initiate positive self-regulatory functions; they will persist in the face of difficulty and maintain higher levels of motivation than those who adopt performance goals. Those with a learning orientation compare their current levels of performance with their own past levels or the desired end-state of one's own performance.

Individuals with *performance* orientation, on the other hand, evaluate their performance by comparing it to their peers', seeking to gain favorable judgments and avoid unfavorable assessments of their competence in comparison to others. Performance-oriented individuals may avoid challenges, adopt easy goals, and display little persistence in the face of difficulty in order to maintain self-esteem.

Integrative Models of Motivational Orientation

The different theories of goal and motivational orientations are closely related, yet distinct. While Dweck (1985) describes a learning motivation to compare performance to one's own skill level versus a performance motivation to display skill in comparison to others, Atkinson's (1957) two motivational tendencies describe a motivation to either display a skill (in comparison to any reference standard) or to avoid failing (again, in comparison to any reference). Elliot and Church (1997) integrate Dweck's theory of learning and performance goals with theories of test anxiety (e.g., Sarason, 1984) by breaking performance goals into two types: *Performance approach* (displaying competence in comparison to others) and *performance avoidance* (avoiding displaying incompetence relative to others) are distinguished from mastery, or *learning goals* in which the individual strives to improve his or her own skill level in comparison to past performance.

To test Elliott & Church's (1997) model, Elliott & McGregor (1999) assessed student's course-grade goals at the beginning of the semester and measured the influence of these goals on exam performance. State test anxiety, worry, and emotionality were measured immediately following the exam and tested for mediation. Avoidance goals were negatively related to performance, while approach goals were positively related, and mastery goals were unrelated to exam performance. Avoidance goals were positively correlated with state test anxiety, while approach and mastery goals were not. State test anxiety was found to mediate the relationship between avoidance motivation and performance.

Similarly, Kanfer and Heggestad (1997) outline a model of motivational traits in which a superordinate “achievement” trait is proposed to be associated with individual differences in mastery, competitiveness, and motivation control skills. A superordinate “anxiety” construct, on the other hand, is thought to be associated with general anxiety, fear of failure, and test anxiety traits, as well as poor emotion control skills. Kanfer and Heggestad (1997) developed a comprehensive measure of motivational traits by combining items from various different motivational orientation measures. Items were first pooled together based on theory, then factor analysis procedures were performed. Three distinct motivational traits emerged which they labeled as *Personal Mastery* (which incorporates Dweck’s mastery/learning goal orientation as well as an additional “hard work” trait), *Competitive Excellence* (similar to Atkinson’s achievement-striving and Elliot & Church’s performance-approach construct), and *Achievement Anxiety* (which captures Atkinson’s fear of failure and Elliot & Church’s performance-avoidance trait).

Motivational Orientation and Self-Efficacy

Research has consistently shown a positive and strong relationship between learning goals and performance. A recent meta-analysis (Beaubien & Payne, 1999) discovered that learning goals were more strongly related to task self-efficacy than performance goals were. Personal Mastery orientations are expected to positively correlate with task self-efficacy and negatively correlate with state anxiety, while performance goals are expected to positively relate with state anxiety and be unrelated to self-efficacy.

Motivational orientation may have an important impact on the degree to which stereotype threat is experienced. Because people with a learning orientation are more concerned with improving their own skill than performing at a given standard, they may be less vulnerable to threat and experience lower state anxiety. One key finding by Margolis et al. (2000) is that female computer science students who compare their performance to male peers feel discouraged and are more likely to disassociate themselves from the major. Women with an Achievement Anxiety orientation, therefore, are probably more vulnerable to the negative effects of stereotype threat and will show lower self-efficacy, greater cognitive interference, and reduced and performance. Women who adopt mastery goals, on the other hand, are less likely to experience reductions in self-efficacy because a different reference standard—their own past performance—is used to evaluate current performance.

> *Hypothesis 8*: Motivational orientations are expected to moderate the impact of threat manipulations. Stereotype threat effects will occur more strongly in persons with a Motivation Related to Anxiety; likewise, stereotype threat effects will occur less strongly in persons with a Personal Mastery motivational orientation.

Summary of Proposed Model and Hypotheses

In order to produce feelings of threat, the stereotype must first be accessible and applicable to the target. Accessibility can be directly manipulated by reminding participants of the stereotype, or stereotypes can be called to mind by more indirect routes, such as priming one's cultural/sex group, priming the stereotype, or providing a stereotyped-relevant task. Individuals are likely to differ in their chronic accessibility of stereotypes, with some being more susceptible to threat than others. Once the target is aware of a negative stereotype, he or she may feel a sense of threat to self-esteem, experiencing both physiological arousal and psychological stress, evidenced by negative emotions such as fear and anxiety.

Feelings of threat may result from a universal motivation to maintain a positive self-image and project this image to others; therefore it is possible that any negative information about one's social group can induce threat for any one person. However, chronic exposure to stereotypes likely creates a stronger association of the stereotype among the targeted group, and those with a greater accessibility of the stereotype will experience threat to a greater degree, as evidenced by greater anxiety and a more significant decrease in performance. Negative stereotypes may increase motivation to perform well and dispel the stereotype, but targets must also care enough about the domain for performance to be impaired; thus it is likely that threat creates an "overmotivation" effect in those who are identified with the domain.

In order for threat to impact performance, the task must be difficult enough to create a sense of challenge and frustration. It is likely that stereotypes impact self-efficacy, either raising or lowering ones confidence for the task, depending on the valence of the stereotype. Thus, when frustration with a difficult task is experienced, a positive stereotype would boost self-efficacy and protect the self-image from threat, leading to persistence and task-engaging behavior in times of stress. A negative stereotype, on the other hand, may lead to self-doubt, worry, and task-irrelevant rumination when anxiety is experienced by frustration with the task. While women may fear confirming the negative stereotype about their math ability, men may fear the possibility of not "living up to" expectations associated with the positive stereotype. By stating that a test is sex fair, the beneficial effect to self-efficacy is removed for males, but it may increase task self-efficacy for females.

- *Hypothesis 1*: Men will outperform women on a difficult mathematical task under conditions of stereotype threat, while men and women will perform equally well in a condition that reduces stereotype threat by describing a task as sex-fair.
- *Hypothesis 2*: Females in a condition that describes a task as one that favors women will outperform females in a control condition in which sex is not mentioned, while females in a condition that describes a task as one that favors men will perform worse than females in a control condition (in which sex is not mentioned).
- *Hypothesis 3*: Males in a condition that describes a task as one that favors men will show an increase in performance compared to males in a control condition in which sex is not mentioned, while males in a condition that describes a task as one that favors women will perform worse than males in a control condition (in which sex is not mentioned).

- *Hypothesis 4*: Participants under conditions of threat (sex differences favoring the opposite sex) will show more cognitive interference than non-threatened participants (control condition and sex differences favoring same sex).
- *Hypothesis 5*: Cognitive interference is expected to correlate positively with threat and negatively with performance and should mediate the threat-performance relationship.
- *Hypothesis 6*: Participants in conditions of threat will display lower task self-efficacy than participants in control conditions; participants in “benefit” conditions (in which the task is described as favoring their sex) will display greater self-efficacy than those in control conditions.
- *Hypothesis 7*: Task self-efficacy is expected to correlate positively with performance and negatively with stereotype threat; self-efficacy should mediate the relationship between threat and performance.
- *Hypothesis 8*: Motivational orientations are expected to moderate the impact of threat manipulations. Stereotype threat effects will occur more strongly in persons with a Motivation Related to Anxiety; likewise, stereotype threat effects will occur less strongly in persons with a Personal Mastery motivational orientation.

Method

Design

A 3 (experimental condition: female threat, control, or female benefit) x 2 (sex: male or female) between-subjects factorial design was used. The primary independent variable was an experimental manipulation that described a difficult mathematical reasoning task as either having been shown to favor males or females; a control condition was also included in which sex was not mentioned. Domain identification and motivational orientations were considered independent variables and used to conduct tests of moderation. The main dependent variable was performance on a set of sample problems from the Graduate Record Exam (GRE). Participant self-efficacy and cognitive interference scores were also treated as dependent variables and used to conduct tests of mediation as proposed by the theoretical model.

The three priming conditions were included in the design for two primary reasons. First, because previous research to date has not incorporated all three possible instructional sets describing sex differences (positive prime, true control, and negative prime), the true boundary conditions of the threat effect are not yet defined. It is not yet known, for example, whether true control conditions that do not mention sex actually produce stereotype threat in females on a difficult mathematical task or if additional environmental manipulations are needed to increase the accessibility of the stereotype and produce the effect. Furthermore, no study to date has attempted to manipulate a culturally-ingrained, negative stereotype by describing a task in an opposite fashion (i.e., describing a difficult math task as favoring females). It was not known whether this description would be believable—whether it could overcome a lifetime of exposure to culturally-ingrained stereotypes, whether it could reduce female performance by actually priming the stereotype, or whether it could negatively impact male performance. If male performance is significantly reduced in the “threat” condition, the likelihood that stereotype threat is merely a form of evaluation apprehension created by experimental manipulation is supported.

Participants

One hundred and forty-five undergraduates, consisting of 79 males (54%) and 66 females (46%), from an elite science and technology university were awarded extra credit for their participation in the study. Participants were tested in groups of two to ten. The data from two participants were dropped from analyses due to either knowledge of the experimental hypotheses or irregular test-taking behavior (one participant completed each of the questionnaires, including all 30 GRE problems, in only a minute or two).

Self-reported SAT scores for the student sample were high, ranging from 1000 to 1540, with a mean of 1293 (*SD* = 110.44). The sample was predominantly Caucasian (n = 95, 66%), with 10 African Americans (7%), 32 Asian or Asian Americans (22%), and 5 (3.5%) Hispanic individuals also represented. The average age was 20. Most participants were freshmen (*n* = 43; 30%), but there was a fairly equal representation of all four years of study, including 35 sophomores (24.5%), 33 juniors (23%), and 32 seniors (22%).

Measures

<u>Domain Identification</u>

Level of identification with math and science was measured by a 20-item scale developed for the study similar to the measures used by Spencer et al. (1999) and Aronson et al. (1999). Participants were asked to rate, on a 7-point Likert scale ranging from strongly agree to strongly disagree, their agreement with questions such as: 1) "The fact that I am a good at math is an important aspect of my personality," and 2) "Success in engineering courses is important to me." Agreement ranged from –3 (strongly disagree) to 3 (strongly agree). (See Appendix B.) This approach has also been validated by Markus (1977) as a measure of self-schema within a given domain. The alpha for the scale was .91.

Motivational Orientation

Heggestad & Kanfer's (2000) Motivational Trait Questionnaire (MTQ) was used as a measure of motivational orientation. The questionnaire is designed to identify individual differences in three distinct motivational traits: *Personal Mastery*, *Competitive Excellence*, and *Achievement Anxiety*. It consists of 48 statements describing different behaviors in academic achievement settings which participants rate as being generally untrue or true of themselves on a 6-point Likert scale. (See Appendix A.) Each of the orientations may be divided into two subscale factors (Kanfer and Ackerman, 2000). Personal Mastery contains two distinct goal-striving orientations: Desire to Learn, which is the need to acquire new skills or knowledge, and Mastery Goals, which is the tendency to set personal goals for continued task improvement. Competitive Excellence captures Other-Referenced Goals, the tendency to make comparisons with peers in order to establish a social reference for one's own performance, as well as Competitiveness, the tendency to make social comparisons with the goal of outperforming one's peers. Finally, Motivation Related to Anxiety, similar to Anxiety, is composed of both Worry and Emotionality goal orientations. Preliminary studies have found the MTQ to have acceptable reliability and construct validity when compared other measures of similar traits (Kanfer & Ackerman, 2000; Heggestad & Kanfer, 2000).

Filler Task

Because pre-tests determined that presenting the threat manipulation and GRE task immediately after the MTQ made participants consciously aware of the experimental hypotheses, a task unrelated to the experimental study was used between administration of the MTQ and GRE task. A sample of problems was taken from the perceptual ability section of a practice workbook for the Dental Admission Test (DAT) (Lehman, 1999). The sample problems consisted of angle discrimination, form development cubes, orthographic projections, apertures, and paper folding questions. Participants were given 30 minutes to work the DAT measure.

Experimental Manipulation of Stereotype Threat

In order to produce stereotype threat, two handouts were created from information taken from the ABC News website (www.abcnews.com) which accompanied a news special on the biological and social bases of sex differences. The information was split into two different handouts: Information favoring women's abilities was used to create the "Male Threat" condition handout, while information favoring males was compiled to create the "Female Threat" handout. No handout was used in the control condition. (See Appendices C and D.)

Task (State) Self-Efficacy

Efficacy is typically measured by asking participants to rate their expected level of performance as well as their degree of confidence in that estimate. Traditional measures are composed of at least two items and ask participants to first indicate whether or not they can perform a task at a given level (yes/no) then to rate their degree of confidence in that decision on a scale ranging from 0 – 100%. Composite measures may include a greater number of items in which scores are determined by summing the number of “yes” responses, often termed self-efficacy magnitude, and/or summing all of the confidence ratings, termed self-efficacy strength. Lee and Bobko (1994) found that the joint measures of strength and magnitude correlated with task performance better than a one-item rating of confidence.

Maurer and Pierce (1998) developed a Likert scale that could simultaneously assess magnitude and confidence while also requiring fewer responses by participants. Studies of reliability and predictive validity, as well as factor analyses, showed that their Likert scale produced results similar to traditional measures of self-efficacy. Maurer and Andrews (2000) improved this Likert measure by simplifying the number of responses in the scale as well as the wording of the questions. Scores on this revised measure were compared with satisfaction ratings to test for convergence with affect. Factor analysis revealed that the traditional, Likert, and revised Likert scales each had only one main factor that accounted for about 45% of the variance. The three scales were highly correlated, and all three showed similar reliability coefficients. Efficacy scores from each of the three measures also correlated significantly with class grades, expected grades, and satisfaction. Based on Maurer and Pierce’s (1998) Likert design, a 10-item measure of self-efficacy was developed for this study that asked participants to successively rate their degree of confidence in the number of math problems that they can complete within 20 minutes. (See Appendix E.)

Performance

Self-regulation is beneficial to performance on resource-independent tasks, when enough attention is available to attend to performance and adjust effort where necessary. For difficult tasks, however, excessive regulation can cause performance to suffer if working memory is not available for both performance regulation and task-related thinking. Decreased attention will have the greatest negative impact on performance for resource-dependent tasks, which require a greater proportion of attention than easy or well-learned ones do (Kanfer & Ackerman, 1989). Therefore, stereotype threat is expected to have a greater impact on performance as task difficulty increases (Spencer et al., 1999).

In order to produce a threat effect large enough to create differences in anxiety and cognitive interference, a practice version of the GRE-Q section of the Graduate Record Exam was used as a measure of performance. (See Appendix F.) Participants were given 20 minutes to complete the 30 test items. This task has been used successfully in several previous studies of stereotype threat (Spencer, Steele, & Quinn, 1999; Quinn & Spencer, 1996). Because the GRE is geared to advanced students who are applying to graduate programs, it should be challenging to high-ability undergraduate students.

Cognitive Interference

The Cognitive Interference Questionnaire (CIQ), was used to assess the frequency of intrusive thoughts that occurred during task performance (Sarason et al., 1986). The CIQ contains 22 items, the first 21 of which measure the frequency of particular types of intrusive thoughts. (See Appendix G.) Answers to each of these 21 questions are self-rated on a six-point scale ranging from: (1) Never to (5) Very often. Scores are derived by summing the ratings; composite CIQ scores can range from 21 (no cognitive interference) to 105 (corresponding to ratings of "5," or "very often," for each of the 21 items).

Composite CIQ items can also be broken down into three conceptually different subscales. The score for Part One, Task-Related Interference, is computed by summing responses to the first ten items, all of which refer to negative thoughts about task performance such as tension, worry, anxiety, and evaluation apprehension. For example, items six and seven state, respectively, "I thought about the difficulty of the problems," and "I thought about my level of ability." Scores on the Task-Related Interference subscale can range from 10 (each item being rated a "1" for "never") to 50 (each item rated a "5" for "very often").

Part Two consists of eleven items, labeled Task-Irrelevant Interference, which are summed to form a measure of intrusive thoughts which do not pertain to the task (i.e. "I thought about something that happened earlier today" or "I thought about members of my family.") Scores on the Task-Irrelevant scale have a possible range from 11 (each item rated a "1") to 55 (each item rated a "5"). Part Three of the CIQ is a global rating obtained from the final (22^{nd}) item. It is a single rating of the degree of overall mind wandering experienced by the subject during the task. Scores on the global rating range from 1 ("not at all") to 7 ("very much").

Manipulation Checks

Stereotype Accessibility. A lexical task similar to Steele and Aronson's (1995) word completion measure was used as a measure of stereotype accessibility. In this task, participants were given a series of letters and blanks and asked to place letters in the blanks order to form a complete word. Words either related to sex stereotypes or unrelated to stereotypes can formed for each incomplete word. (See Appendix H.) Participants were given five minutes to form as many words as possible.

In addition, a free-response question included in the post-experimental questionnaire (described below) asked subjects to "describe themselves." The content and the order of responses was examined to assess the accessibility of sex. (See Appendix I.) Because previous research has shown that participants are less likely to mention their race or sex under conditions of threat (Harder, 1999; Steele & Aronson, 1995), it was hypothesized that participants would be least likely to mention their sex under conditions of threat and most likely to mention their sex under "benefit" conditions.

Post-Experimental Questionnaire. A post-experimental questionnaire (Appendix J) was used to determine participants' awareness of the true hypotheses. Participants were asked to state the purpose of the study in an open-ended question. They were also asked to give their race, sex, enrollment status (freshman, sophomore, junior, or senior), and academic major as well GPA, and SAT scores. (See Appendix I.)

Procedure

Upon arrival, participants were asked to sign the consent form, given a brief overview of their rights as volunteers, and told that during experimental session they would be asked to complete a series of psychological measures. They were then told that that each measure represented a different line of research being conducted in the School of Psychology by different professors, and that the purpose of the study was to determine each measure's reliability and validity prior to use in later research. As part of the cover story, students were told that each of the individual measures took only five to twenty minutes to complete, and they were therefore grouped together into one experimental session in order to grant a full two hours of extra credit. Participants were then asked to complete the Domain Identification measure and MTQ, which were described as personality measures for which there were no right or wrong answers. Participants were told they should answer as accurately about themselves as possible because the researchers were interested in assuring that the measures could accurately differentiate between personality types, capturing a true “bell curve” of responses with no range restriction. Following the MTQ, participants were given thirty minutes to complete the Dental Admission Test, which was described as a sensation and perception assessment that was created by sampling items from a longer version of the same measure; researchers were supposedly interested creating an optimally-valid assessment that was not as burdensome to complete.

At this point, the participants in the male and female threat conditions were given a break to “clear their heads,” and the “Brain Game” handout was distributed supposedly as a means of providing background information for the next portion of the study. (See Appendices C and D.) The mini study was described as part of a national research project being conducted in collaboration with other colleges and universities across the United States to better understand the nature of sex differences in math abilities. Participants were told that initial researchers had males and females to complete a series of different cognitive tasks under an MRI to see what parts of the brain were utilized for the different tasks. In the female threat condition, participants were told that men and women showed a marked difference in math performance, and that this session was part of a larger, national study being conducted to determine the exact size and nature of the difference. The GRE task was then introduced as one which favored men.

In the male threat condition, participants were told the initial researchers discovered that women used their brains more efficiently; supposedly, male's brains "lit up" under the MRI in only very specific places for specific tasks, whereas female's brains "lit-up" in larger areas across several different parts of the brain at once. Participants were told that the experimental task was designed specifically to favor women's superior abilities to "multitask' and hold more information in working memory at one time. Initial research had supposedly been performed which showed women performed better than men on particular types of math problems, and this session was part of a larger, national study being conducted to determine the exact size and nature of the difference.

For the control condition, the purpose of the study was presented simply as another psychological measure in the study without any prior handout or explanation, and the normal instructions for the GRE were read aloud.

All participants were asked to complete the example problems, look over the entire GRE booklet, then given the task self-efficacy measure prior to completing the math items. Once the self-efficacy measure was completed, participants were given 20 minutes to complete the GRE task, then were asked to complete the Cognitive Interference Questionnaire.

Finally, the Stereotype Accessibility Measure was introduced as a word game that was being pre-tested for a separate cognitive psychology experiment. After the accessibility measure was collected, participants were given the post-experimental questionnaire. Participants were then thoroughly debriefed, given an overview of stereotype threat theory and the true experimental hypotheses, and excused from the study.

Results

The acceptable alpha level for all analyses was set at .05, with the exception of the alpha used in planned contrast ANOVAs which was set at .10 because of the directionality of the proposed hypotheses.

Preliminary Analyses

Demographic Variables

Although participants were randomly assigned to experimental groups, preliminary comparisons between conditions were made to determine whether there were any pre-existing differences in demographic variables that might be related to GRE performance. Analyses were performed separately for males and females. Chi Square results revealed no significant differences between conditions in terms of race, year of study, or academic major, and one-way ANOVAs found no differences in domain identification, age, or motivational traits (Personal Mastery, Competitive Excellence, or Motivation Related to Anxiety). In addition, there were no significant differences between conditions in SAT scores; therefore, SAT scores were not controlled for in statistical analyses.

Manipulation Check

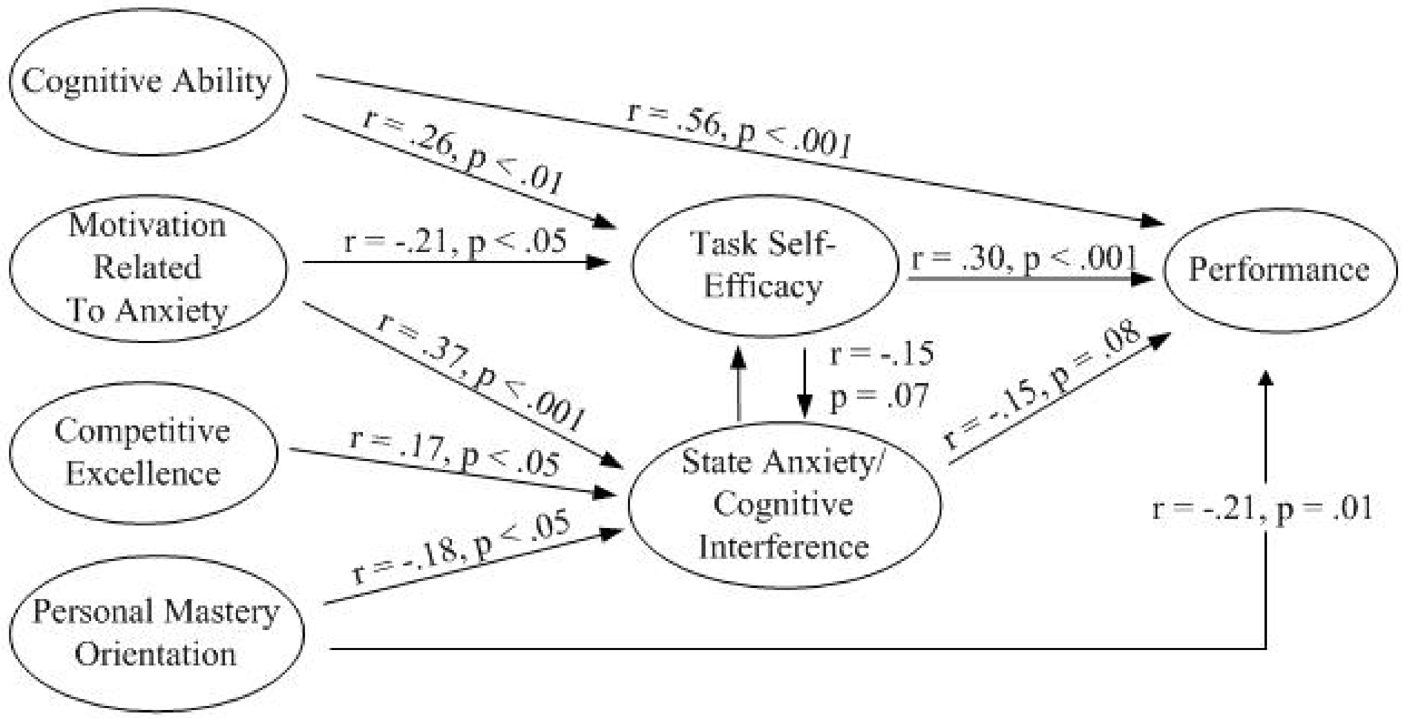
Two measures were used to determine whether or not the manipulation in the experimental conditions made the female stereotype salient to participants. The first measure was the number of stereotype-related words completed on the lexical task. A 2 (sex: male vs. female) x 3 (condition: female threat, control, and female benefit) Analysis of Variance found no main effects or interaction of sex and experimental condition on the number of stereotyped words formed. The second manipulation check assessed whether or not sex was mentioned on the post-experimental questionnaire in response to the open-ended question “Describe yourself.” Chi-square analyses performed separately for males and females revealed no significant difference among the three conditions for either sex, indicating that the manipulation did not work.

Descriptive Analyses

Table 1 displays the abbreviations used for each variable in the study. Pearson correlation coefficients between all continuous variables in the study are presented in Table 2. Descriptive statistics for all variables are provided for Males and Females in Tables 5 and 6, respectively.

Proposed Model. Correlational analyses revealed support for all proposed paths in the model, including the hypothesized sign of the relationships.

Figure 3 *Path Correlations Between Variables in Proposed Model*



In addition to the paths that were proposed, a direct negative relationship was found between Motivation Related to Anxiety and GRE performance ($r = -.21$, $p \leq .01$), while a significant positive relationship was found between Competitive Excellence and cognitive interference ($r = .17$, $p < .05$). Because the Competitive Excellence trait captures the tendency to adopt other-referenced goals (comparing one's performance with peers), it is logical that these students would have greater performance concerns and report more frequent agreement with statements on the CIQ such as "I thought about how others have done on this task," and "I thought about what the experimenter would think of me."

GRE Performance. As expected, GRE performance was significantly related to SAT scores ($r = .53$, $p < .001$). A significant positive relationship was found between GRE performance and Domain Identification such that those who were more strongly identified with the task domain completed more GRE problems successfully ($r = .38$, $p < .001$). Those who competed more GRE problems also reported higher Self-Efficacy ($r = .30$, $p < .001$) as well as higher levels of the Desire to Learn factor of Personal Mastery orientation ($r = .17$, $p < .05$). GRE scores correlated negatively with Motivation Related to Anxiety ($r = -.21$, $p < .05$), particularly the Emotionality factor ($r = -.25$, $p < .01$). Contrary to expectations, the proposed negative relationship between GRE performance and task-related cognitive interference only approached significance ($r = -.15$, $p = .08$). The Competitive Excellence motivational orientation also correlated relatively weakly with task performance ($r = .16$, $p = .06$). Interestingly, age correlated negatively with GRE such that younger students successfully completed more GRE problems than older students ($r = -.179$, $p < .05$). Older non-math-related majors may not have used their math skills in several years, while younger students were probably either currently enrolled in the required mathematics core courses or had recently completed math preparatory courses in secondary school.

Domain Identification. Those who reported greater identification with the task domain showed greater self-efficacy for the task ($r = .24, p < .01$) and lower ratings of overall cognitive interference ($r = .38, p < .001$). As noted above, identification also correlated significantly with GRE performance ($r = .38, p < .001$) such that those who were strongly identified with the domain completed more GRE problems correctly.

Motivational Orientations. In general, the Personal Mastery motivational orientation and its related factors correlated positively with domain identification and negatively with cognitive interference, while Motivation Related to Anxiety and its factors correlated with positively cognitive interference and negatively with self-efficacy. A significant positive correlation was also found between the Desire to Learn factor of Personal Mastery and GRE performance, while Motivation Related to Anxiety and its related Emotionality factor were negatively correlated with GRE performance. (See Table 2.)

Self-Efficacy. Self-efficacy was significantly related to performance ($r = .30, p < .001$) such that self-efficacious persons completed more GRE problems correctly. Participants who reported greater self-efficacy also completed fewer stereotyped-related words on the lexical task ($r = -.18, p < .05$).

Cognitive Interference. The proposed negative relationship between task-related cognitive interference and self-efficacy only approached significance, $r = -.15, p = .07$.

Likewise, its proposed negative relationship with GRE performance only approached significance, $r = -.15$, $p = .08$, failing to meet the .05 alpha level criterion.

Male and Female Samples. Because the pattern of relationships may differ for each sex, correlations among all variables in the study are provided separately for males and females in Tables 3 and 4, respectively.

It is interesting to note that among females, the Emotionality factor of Motivation Related to Anxiety correlated significantly with GRE performance ($r = -.26$, $p < .05$), but the same relationship was not found within the male sample. Likewise, the number of stereotype-related words completed on the lexical task correlated negatively with self-efficacy ($r = -.26$, $p < .05$) as well as with Competitive Excellence ($r = -.25$, $p < .05$), specifically the tendency to use Other-Referenced Goals ($r = -.24$, $p < .05$), within the female sample only. These results may be interpreted as an indication that the threat manipulation did work with some of the women; participants whose task self-efficacy was chronically low (or negatively impacted by the manipulation) and who were concerned with outperforming their peers—those most likely to feel their self-esteem threatened by possible poor performance—also reported thinking about words related to the stereotype more than participants with high self-efficacy and low competitive motivations.

Hypotheses Tests

GRE Performance

Hypothesis one predicted that men would successfully complete more GRE problems than females in conditions of stereotype threat, but that males and females would perform equally under "female benefit" conditions. One-way analyses of variance conducted separately for each condition (female threat, control, and female benefit) found partial support for the hypothesis. While no significant differences between males and females were found in either the control or benefit conditions, a significant effect of sex on GRE performance was found in the female threat condition, $F(1, 47) = 7.55$, $p < .01$, with men ($M = 18.73$) outperforming women ($M = 15.26$). (See Table 7.)

Hypothesis two predicted that women in the benefit condition would outperform those in the control and threat conditions. Because of the a priori theoretical rationale, planned contrasts were performed; two degrees of freedom associated with the three experimental groups (threat, control, and benefit) allowed for two planned contrasts to be conducted. One-tailed contrasts were expected to reveal significantly lower GRE performance for females in the threat as compared to control condition, while female performance in the counter-stereotype "benefit" condition would be higher than in control condition. The trend was supported by the direction of mean GRE scores for women (Threat = 15.26, Control = 17.23, and Benefit = 18.12). However, neither mean contrast (threat versus control group and control versus benefit condition) was statistically significant in the female sample at the .10 alpha level; thus, hypothesis two was not supported.

Hypothesis three predicted that one-tailed, planned contrasts would reveal that males in the female benefit condition (in which the task is described as favoring females) would have significantly lower scores on the experimental task as compared to those in the control condition, while those in the female threat condition would outperform those in the control group. This hypothesis was not supported; no significant differences were found for GRE performance between either of the experimental conditions and the control group. It would appear, then, that male performance was not adversely impacted by the "female benefit" manipulations.

Cognitive Interference

Hypothesis four predicted that participants under conditions of threat (in which the test was described as favoring the opposite sex) would report greater cognitive interference than those in the control condition, while those in the benefit conditions (in which the test was described as favoring one's own sex) would report significantly lower cognitive interference than the control group. Planned contrasts revealed no significant difference in task-related or total cognitive interference between experimental conditions and the control group for either males or females. Thus, hypothesis four was not supported.

State Self-Efficacy

Hypothesis six predicted that participants under conditions of stereotype threat would show significantly lower self-efficacy than those in the control conditions, while those in "benefit" conditions would show significantly greater self-efficacy. One-tailed, planned contrasts were performed separately for males and females. This hypothesis was partially supported for females, as females' self-efficacy in the benefit condition was significantly higher than in the control condition, $t\ (63) = -1.78$, $p < .10$. (See Table 8.) As predicted, self-efficacy was greatest in the female benefit condition ($M = 15.36$); however, females in the threat condition reported greater task self-efficacy ($M = 14.32$) than those in the control group ($M = 12.95$), which was counter to the hypothesis.

Among men, one-tailed planned contrasts were expected to show that male self-efficacy scores in the female benefit condition would be significantly lower than self-efficacy scores in the control and female threat condition. This hypothesis was not fully supported, as levels of reported self-efficacy in the control and female benefit conditions were not significantly different. However, planned contrasts revealed that males' self-efficacy in the female threat condition (M = 17.8) was significantly greater than the control condition (M = 15.60), t (74) = 2.06, $p < .05$. (See Table 9.) Therefore, it seems that males' self-efficacy was increased by the "female threat" manipulations which primed a positive stereotype for males' superior abilities, but their self-efficacy was not adversely impacted by the "female benefit" manipulation which described a task as favoring females.

Tests of Mediation

Tests of mediation were conducted using the regression procedure established by James and Brett (1984). In this process, the mediator is first regressed on the independent variable. If this relationship is significant, a second equation is tested in which the dependent variable is regressed onto the mediator. The independent variable is then added to this regression equation, and the change in R^2 is tested for significance. If the increment is not significant, then it can be stated that the mediator completely mediates the relationship between the independent and dependent variables. If the increase in R^2 is significant, however, the independent variable has a direct effect on the dependent variable. And if the change in R^2 is less than the original R^2 accounted for by the independent variable, then an argument can be made for partial mediation. Finally, a change in R^2 equal to the original R^2 between the independent and dependent variables indicates that the proposed mediator does not mediate the IV/DV relationship.

Cognitive Interference. Hypothesis five predicted that cognitive interference would correlate positively with threat and negatively with performance and mediate the threat-performance relationship. No significant relationship was found between cognitive interference and experimental condition for either the male or female sample; therefore no argument could be made for cognitive interference as a mediator.

Self-efficacy. Hypothesis seven predicted that task self-efficacy would mediate the relationship between threat and performance. In the female sample, no relationship between experimental condition and self-efficacy was found; therefore, mediation was not supported. For males, the proposed relationship between experimental condition and self-efficacy was found to be significant, $\beta = -.24$, $p < .05$. Therefore, a second equation was tested in which GRE performance was regressed on self-efficacy. This relationship was also significant, $\beta = .29$, $p = .01$. Finally, experimental condition was added to the equation, and the change in R^2 was tested. The addition of experimental condition to the equation increased the Beta from .29 to .32, but the change in R^2 was not significant ($p =.32$), indicating the complete mediation by self-efficacy. (See Table 10).

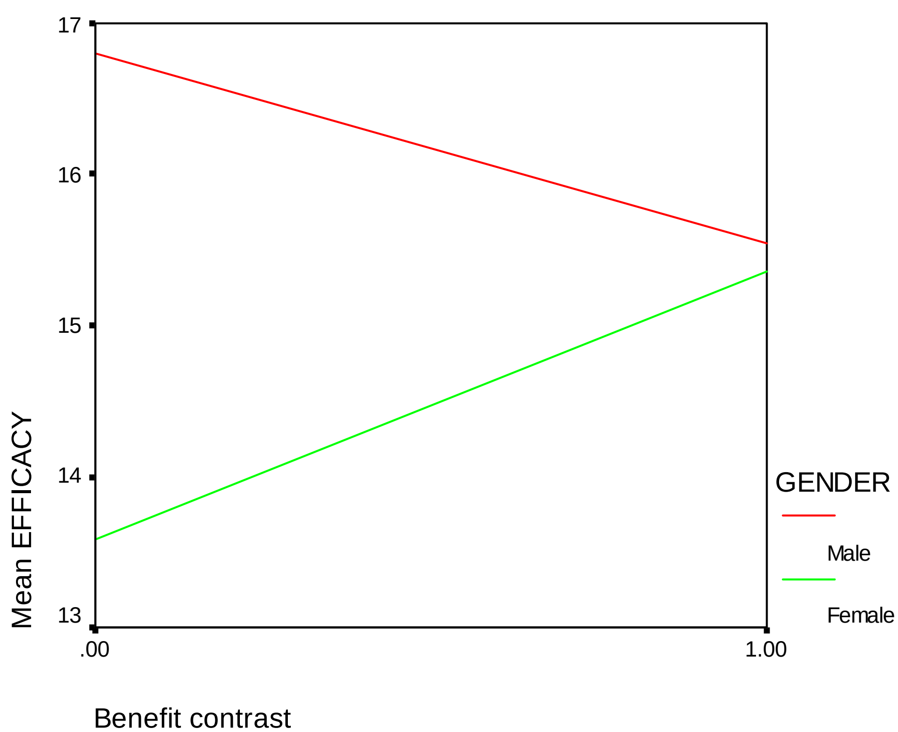
Tests of Moderation

To test for moderation, the continuous predictor variables (domain identification and motivational traits) were first mean centered by subtracting each value from the appropriate sample mean for each sex. Because the focal independent variable, experimental condition, was a qualitative predictor, contrast terms were computed using the control condition as a reference group. A dummy-coded “threat” term was created by assigning those in the female threat condition a “1” and all others a “0;” likewise, a dummy-coded “benefit” term was created by assigning those in the benefit condition a “1” and all other groups a “0.” Product terms for each proposed interaction were then created by multiplying the moderator with the threat contrast term and the benefit contrast term. Hierarchical regression was used in which the main effects (threat term, benefit term, and proposed moderator) were entered in the first step and the product terms (threat term x moderator and benefit term x moderator) were entered in the second step. Finally, the change in R^2 between the two steps was tested for significance.

Sex. Sex was proposed to moderate the impact of experimental manipulations on self-efficacy. “Female benefit” manipulations which described a task as favoring females were predicted to increase self-efficacy and decrease cognitive interference for women but not for men, while the “female threat” manipulation which primed the stereotype of males’ superior math skills was predicted to decrease self-efficacy and cognitive interference among women but not among men. To test for moderation by sex, contrast terms for experimental condition were computed using the control condition as a reference group. Product terms for each proposed interaction were then created by multiplying sex with the dummy-coded threat and benefit contrast terms. Hierarchical regression was used in which the main effects (threat term, benefit term, and sex) were entered in the first step and the product terms (threat x sex term and benefit x sex term) were entered in the second step. Finally, the change in R^2 between the two steps was tested for significance.

A simple effect for the Benefit term approached significance in predicting Self-Efficacy, $\beta = -.453$, $p = .08$; in addition, the beta weight for the Benefit x Sex product was significant, $\beta = -.528$, $p = .05$. (See Table 11.) The change in R^2 associated with entering the Benefit x Sex product term in the second step of the regression was also significant, $F(1, 139) = 3.86$, $p = .05$. Thus, the female benefit manipulation may have impacted self-efficacy differentially for males and females. As Figure 4 shows, the female benefit manipulation increased female self-efficacy and decreased male self-efficacy as compared to the other two conditions.

Figure 4. *Differences in Efficacy by Experimental Contrast*



Motivational Orientations. Hypothesis eight predicted that motivational orientations would moderate the impact of threat manipulations on cognitive interference and self-efficacy. Stereotype threat effects (increased cognitive interference and reduced self-efficacy) were expected to occur more strongly in persons with a higher Motivation Related to Anxiety orientation than those with less anxiety. Likewise, stereotype threat effects were expected to occur less strongly in persons with a Personal Mastery motivational orientation such that they would show less cognitive interference and higher self-efficacy than those with a low Personal Mastery orientation.

Because the effects of the experimental manipulations were theoretically different for males and females, tests of moderation were computed separately for each sex. Within each sex, Personal Mastery and Motivation Related to Anxiety were tested in separate regression procedures. Hierarchical regression equations were computed by entering the threat term, benefit term, and mean-centered motivational orientation of interest in the first step and product terms (threat x mean-centered motivational orientation and benefit x mean-centered motivational orientation) in the second step. For cognitive interference, no significant changes in R^2 were associated with adding the interaction terms to equations involving Personal Mastery or Motivation Related to Anxiety in either the male or female sample; therefore, no indication of the moderation of experimental condition by motivational orientations on GRE performance was found.

The same hierarchical regression procedure was repeated for predicting self-efficacy. No evidence of moderation by Self-Efficacy was found for males or females.

Post-Hoc Analyses

<u>Mediation of Sex Differences</u>

Based on preliminary results, two post-hoc tests of mediation were conducted. One of the most striking findings was the significant difference between males' (M= 23.53, *SD* = 18.7) and females' (M = 6.28, *SD* = 23.2) reported domain identification as well as the significant relationship between domain identification and GRE performance ($r = .38$, $p <.001$). Therefore, the possibility that domain identification may explain sex differences in performance was explored. First, possible sex differences in GRE performance and domain identification were established by a post-hoc ANOVA, with sex as the independent variable and domain identification and GRE performance as the dependent variables. The anticipated effects were observed, with males (M = 19.05) successfully completing significantly more GRE problems than females (M = 17.00), $F (1, 141) = 6.76$, $p < .01$. Females also reported significantly lower identity with the task domain (M = 6.11) than males did (M = 23.53), $F (1, 138) = 23.7$, $p < .001$. (See Table 12.)

A test of the mediation of sex difference by domain identification was then conducted using the steps outlined by James and Brett (1984). First, domain identification was regressed on a dummy-coded sex variable, and a significant relationship was found, $\beta = -17.25$, $p < .01$. (See Table 13.) A second equation was then tested in which GRE performance was treated as the outcome and regressed onto domain identification. The beta weight for identification was significant, $\beta = .382$, $p < .001$. One final analysis examined whether the introduction of sex would significantly reduce the relationship between domain identification and GRE performance. Domain identification was entered as a predictor in the first step, then sex was added in the second step of a stepwise regression predicting GRE performance. The beta weights of each were then tested for significance. The relationship between domain identification and GRE performance, controlling for sex, remained significant, $\beta = .074$, $p < .001$, but the beta for sex was no longer significant, $\beta = -.95$, $p > .05$. The change in R^2 was associated with adding sex to the equation was non-significant, $F (2, 137) = 12.47$, $p > .05$. Therefore, support was found for complete mediation of sex differences in math performance by domain identification.

Another interesting finding was the significant sex difference in self-efficacy. (See Table 11.) Because self-efficacy was also positively related to GRE performance, the same regression procedure was conducted to test for mediation of sex differences in performance by self-efficacy. First, self-efficacy was regressed onto sex, and a significant relationship was found, $\beta = -.245$, $p < .01$. A significant relationship between self-efficacy and GRE performance was then established, $\beta =.300$, $p < .01$. A final analysis examined whether the introduction of the proposed mediator would significantly reduce the relationship between sex and GRE performance. The results shows that when self-efficacy was covaried, the beta for sex was reduced, $\beta = -.149$, $p = .07$. (See Table 14.) Therefore, support was found for mediation of sex differences in math performance by self-efficacy.

A related question was the differential impact of experimental manipulations on participants' self-efficacy. While women's self-efficacy appeared to be equal to men's in the benefit condition, women's efficacy was lower in the control and threat conditions. Therefore, an ANOVA was conducted to test for gender differences in efficacy for each condition. Results show that while no significant differences in efficacy existed in the Female Benefit condition, men reported significantly greater task self-efficacy ($M = 17.8$) than women did ($M = 14.32$) in the Female Threat condition, $F(1, 47) = 11.5$, $p < .01$. A main effect for sex also approached significance in the control condition, $F(1, 45) = 3.8$, $p = .058$. (See Table 15.)

Discussion

When a negative stereotype about one's ability is cognitively accessible, targets of the stereotype may fear being judged by it or confirming it to be true in an applicable domain. Targets may be more likely to attribute frustration with a difficult task to their own lack of ability and ruminate over negative thoughts and feelings. This phenomenon, labeled stereotype threat, has been shown to reduce performance. Past research has found that when targets of a negative stereotype are primed with information about the stereotype, they perform poorly. The current study replicates previous research, finding that males performed significantly better than females did when a negative stereotype about women's math abilities was explicitly made salient. When the task was described as one designed to capture females' superior ability to "multi-task" and hold more information in their working memory, however, no significant gender differences in performance were found.

One unanswered question is whether threat must be explicitly primed via experimental manipulations or whether it is always cognitively accessible within any relevant situation, such as when facing a difficult math test. Some scientists argue that standard testing instructions alone are sufficient to create a level of threat that is as detrimental to performance as a threat that is explicitly-induced. No study to date has compared explicit manipulation of threat to a true control group to determine if one form of threat is more detrimental to performance than the other.[1] It was proposed that threat

[1] In past research, two methodologies have been used to measure threat: 1) Performance under standard testing conditions (a true control) has been compared to performance in a condition that supposedly "removes" threat by specifically stressing past findings of gender equity; or 2) the task is described as having been shown to produce sex differences (an explicit manipulation of the effect) or described as being gender-fair.

does not occur in an “all-or-none” fashion. Rather, threat effects (such as reduced performance, cognitive interference, and increased anxiety) may be more acute in certain situations (such as laboratory-induced manipulations) and weaker in others, depending on the surrounding context. Threat may exist, for example, in standardized test settings, but it may not be strong enough to cause significant performance decrements unless explicit statements consciously remind test candidates of the stereotype. The current study attempted to clarify the issue by comparing performance in all three scenarios: 1) participants who received an explicit threat induction, 2) a true control group in which participants received the standard instructions only, and 3) those who heard a counter-stereotype message. Threat to females was expected to be greatest in the explicit threat condition (“past differences”), lower in the standard (control) condition, and lowest in the counter-stereotypic (“no difference”) condition. Mean performance scores support this trend, with women in the threat condition completing fewer GRE problems (M = 15.26) than those in the control condition (M = 17.23), who completed less than those in the counter-stereotype condition (M= 18.12). However, planned contrasts found no significant difference in scores between the control condition and either experimental contrast, although the contrast between the threat and control condition did approach significance. Had the participant sample been larger, it is possible that a significant difference could have been detected, indicating that more threat was experienced in the threat than the control condition.

Given the present results, an argument can be made for the existence of threat in the control condition. Gender difference contrasts show that, while female students performed as well as the males did when the stereotype was “counteracted,” women performed worse than men in both the threat and control settings. Furthermore, while females reported lower levels of self-efficacy than males did in both the control and "female threat" conditions, no sex difference in efficacy was found when the task was described as one in which women typically performed as well as or better than men. In this condition, females reported their highest level of efficacy (M=15.4), equivalent to the level of efficacy reported by males in the control (M = 15.6) and female threat (M = 15.6) conditions. The significant difference between male and female self-efficacy in the threat condition was partially accounted for by an increase in self-efficacy among men, whose efficacy (M = 17.8) was increased to its highest level by manipulations which explicitly primed a positive stereotype about their abilities; but the effect is also due to a reduction in self-efficacy among women. (See Figure 4.) The analysis comparing women only found no difference in efficacy between the control and threat conditions, while those in the benefit condition reported significantly higher levels.

Taken together, these findings support Steele and Davies's (2003) argument that feelings of threat were present among women in both the control and threat conditions; only when threat was removed by presenting information counter to the stereotype did women perform as well as men. It appears that threat can be produced by high-pressure testing situations alone and that simply being in a stereotype-relevant situation, whether the stereotype is explicitly mentioned or primed through more subtle means, is sufficient to induce threat among women.

Converging evidence for this argument is also found in research published since the current study began. Smith and White (2003) investigated the impact of math stereotypes on women's performance using comparable experimental conditions. As in the present study, when the stereotype was “nullified” by telling participants that men and women performed equally on the experimental task, females performed better than a control group. Those in the control group did not perform any better than women who were explicitly reminded of the negative stereotype.

However, unlike Smith and White (2003), the current study failed to find a significant difference in women’s performance between the control and threat-removed groups. That is, presenting information counter to the stereotype did not result in significantly improved performance among women. One important difference is the possibility that the students in the current experiment may not have felt as threatened in the control condition as students at other universities because of their high ability level. The current sample was composed of students attending an elite science and technology institution; the average participant SAT score in the current study was 1293, while the national average is 1000 (College Board, 2004). Furthermore, the institution is consistently ranked in the top ten public universities in the country by *U.S. News & World Report* and is well-known for its nationally-ranked computer science and engineering programs. Unlike liberal arts colleges and universities, courses at this institution focus heavily on mathematics, and students are typically attracted to the school because of their interest and skill in math and science.

Therefore, it is possible that the females in this study were threatened only slightly by a standardized math test in the control condition—not to the degree that the general population of women would be. Rather, only by explicitly reminding them of the stereotype was the effect strong enough to reduce performance. Although their self-efficacy was initially decreased prior to the task, once the women in the control condition began to work the math problems, they might have been able to put aside their fears and remain task focused as they realized that the problems were not exceptionally difficult. Because the general GRE test was used (and not the more advanced GRE math subject test), the task might not have been sensitive enough to performance decrements caused by off-task thoughts; difficult and novel tasks require a greater allocation of working memory than easy or well-learned ones (Kanfer & Ackerman, 1989; Sarason et al., 1986). Spencer et al. (1999), for example, found that female participants who were assigned more difficult problems from the GRE Mathematics subject exam displayed the typical pattern of underperformance relative to men, while those assigned the easier General GRE-Q problems did not. Future research investigating math stereotypes should be certain to use a task that is challenging enough to create a sense of frustration in the sample population. Because threat likely increases arousal level and motivation, it may also enhance performance on tasks that are considered easy to the participant sample; only when the task is difficult enough such that students cannot attend to both the task and to self-regulation simultaneously will reduced performance occur. Had the task been harder, it is possible that female performance in the control condition would have been reduced.

It is also interesting to note that high-ability students may be less susceptible to implicitly-induced than explicitly-induced threat. Because they have a history of success in the domain, and because the math tasks are not as challenging to them, high-ability students might be less likely to spontaneously dwell on the implications of negative performance unless specifically told to do so. Moderate-ability students, on the other hand, likely experience a greater sense of frustration with a difficult task and thus call the stereotype to mind more easily. Future research should investigate whether more explicit means are required to induce threat as ability increases.

Stereotype Threat Effects on Non-Targets

If threat is implicitly induced in the general population by normal testing situations, measures to counteract or ameliorate its negative impact in real-world settings should be explored. Promising results are beginning to emerge. Spencer, Steele, and Quinn (1999), for example, found no gender differences in math performance when men and women were first told that a test had been shown not to produce gender differences. However, such "gender-fair" manipulations can potentially impact non-targets. Prior research, for example, shows that men can feel pressure to confirm the stereotype of male's superior math ability (Brown and Josephs, 1999) and that Caucasian males can feel threatened by the stereotype of Asian's advanced math skills (Aronson et al., 1999). Yet no study to date has investigated whether the "gender-fair" manipulation negatively impacts male performance.

The current study found that that males outperformed females in the control and threat conditions, but women performed as well as men in the threat-removed condition. An unanswered question of most stereotype threat studies is the source of the difference in the latter condition—whether male performance was negatively impacted by the counter-stereotypic female "benefit" message, whether the benefit information removed feelings of threat for females, or whether both occurred simultaneously.

It was hypothesized that men in the "female benefit" condition would have significantly lower scores on the experimental task as compared to those in the control condition, while those in the "female threat" condition would outperform those in the control group. This hypothesis was not supported. No performance differences among males were found. While males' self-efficacy was increased by the "female threat" manipulations (which described a math task as favoring males), their self-efficacy was not reduced by the "female benefit" manipulation. Therefore, in the current study, men were not negatively impacted by the manipulation to remove threat among women.

Targets may have to face repeated exposure to the stereotype in order for threat to have any significant impact. Because direct targets are often physically distinctive (as are ethnic minorities, for example), they may be more identified with their particular group (Brewer, 1991). Chronic exposure to stereotypes by direct targets likely engenders performance concerns that are chronically accessible in certain situations. Dijksterhuis, Aarts, Bargh, & van Knippenberg (2000), for example, have shown that frequent elicitation of stereotypes in the presence of members of a stereotyped group can lead to stronger associations between the stereotype and target group. Thus, because direct targets come into contact with a negative stereotype about their group relatively frequently, the stereotype can become automatically linked to certain domains.

Moderators of the Threat Effect

<u>Domain Identification</u>

According to Steele's (1997) theory, targets must care enough about the task domain to fear being viewed negatively by others and experience threat. Numerous studies have found that the threat effect is strongest among those who are highly identified with the task domain (e.g., Aronson et al., 1999). However, in the current study the opposite occurred: Participants who reported greater identification with the task domain completed more GRE problems correctly and experienced less cognitive interference. This may be due to the fact that SAT scores correlated significantly and positively with domain identification ($r = .243$, $p < .01$). Among the most highly-identified participants (those in the top 30% on the domain identification measure), the average SAT score was 1321, whereas the lowest 30% in domain identification had an average SAT of 1293 (see Table 16). It is likely that among the most highly-identified students, the task was simply not difficult enough to produce the threat effect. Threat might have increased arousal among the high-identified participants; however, the task was not likely difficult enough to show performance decrements associated with anxiety and self-regulation.

Motivational Orientations

Personal Mastery and Motivation Related to Anxiety were expected to moderate the impact of threat on two proximal predictors of performance: cognitive interference and self-efficacy. As predicted, Personal Mastery was negatively related to cognitive interference, yet no evidence of moderation of cognitive interference by Personal Mastery was found. This is likely attributed to the fact that the method used to assess cognitive interference was imprecise, as discussed below. Personal Mastery, furthermore, was unrelated to task self-efficacy, and no moderation of the threat-efficacy relationship by personal mastery was supported.

The lack of any support for moderation by motivational orientations may indicate that students' proximal goal striving was impacted by the threat manipulation. That is, stereotype threat may operate to reduce performance by priming performance goals in targets, independent of their chronic motivational orientation. Research by Carver & Scheier (1981) shows that participants who perform a task in the presence of an audience or evaluator report greater self-focus and adoption of performance goals more often than those who perform the task alone. Likewise, Nicholls (1975) found that under ability-diagnostic conditions, individuals are more likely to adopt performance-type goals than when the task is framed as non-diagnostic. Targets of a negative stereotype may fear impending evaluation, increasing self-focus and the salience of performance goals despite their typical motivational orientation. Future research should investigate this possibility by using a state (rather than trait) measure of proximal motivation and comparing the task goals of targets versus non-targets.

Although it did not moderate the impact of threat on the two proposed proximal predictors, Motivation Related to Anxiety was negatively related to task self-efficacy and positively related to cognitive interference. This is significant because gender differences in motivational orientations have been demonstrated in prior research (Kanfer & Ackerman, 2000). The tendency for women to adopt an avoidance rather than a mastery goal puts them at a much greater risk for reduced performance. While performance-avoidance goals can be beneficial on easy tasks, they are also linked to lower performance as task difficulty increases.

Task self-efficacy likely moderates the impact of performance-avoidance goals on actual performance. Efficacy must be high and remain high in order for an avoidance orientation to enhance goal setting and performance. If not, anxiety-related goals can decrease performance. People who adopt avoidance goals are more likely to interpret task failure as an indication of low ability (Ames, 1984; Elliot & Dweck, 1985), leading to withdrawal of effort in the face of obstacles (Elliot & Dweck, 1985) and increased worry, which reduces task-focused attention. Therefore, because women have both lower task efficacy and a greater tendency to experience anxiety than men, they are at a significant disadvantage in the mathematics domain.

Mediators of the Threat Effect

<u>Cognitive Interference</u>

When confronted by a difficult task, those for whom a negative stereotype applies may ruminate over negative thoughts and feelings, decreasing available working memory for the task; however, no impact of threat on cognitive interference was found in the current study, and the proposed relationship between cognitive interference and GRE performance was relatively weak. This result is surprising, given the strong theoretical rationale in support of cognitive interference as a mediator. One possibility that could account for the lack of significant results is that the high ability sample did not experience cognitive interference to create a strong effect on performance.

Another possibility may be the measure used, which might not have accurately captured distracting thoughts as they occurred. Because the measure was completed post-task, it relied on participants' level of self-awareness and ability to recall prior thoughts. A more accurate method would be to assess cognitive interference as it occurs by measuring reductions in working memory capacity. Schmader and Johns (2003), for example, recently modified Turner & Engle's (1998) operation-span task for use in stereotype threat research. The task requires participants to solve a series of equations while also memorizing a set of words. Each equation is followed by a word the participant is asked to memorize, and working memory is estimated by the number of words participants can recall at the conclusion of the math task. Future research using such behavioral measures of cognitive load may better estimate cognitive interference.

Assessment of working memory is a promising method that has had relatively little use to date, even though working memory is perhaps the strongest theoretical mediator of the threat-performance relationship.

Future research should explore possible routes by which attention is diverted. One method, for example, would be to present a stereotyped task to a set of targets, and then periodically ask them to self-report their thoughts at that moment. Females working advanced math problems, for example, may expend working memory by trying to suppress thoughts about the stereotype itself; alternately, they may suffer reductions in capacity by ruminating over task failure, or an increase in self-doubt may compel them to constantly stop to check and recheck their own work. Threat may operate differently for different individuals, but if one or two types of cognitive interference are consistently identified, ways to address these particular problems could be developed.

Self-Efficacy

Self-efficacy played a significant role in participants' performance in this study. Overall, efficacious persons completed more GRE problems correctly, and women completed fewer stereotyped-related words on the lexical task than those low in self-efficacy. Self-efficacy was found to mediate the relationship between experimental manipulations and performance within the male sample; female threat manipulations were related to increased self-efficacy, which increased men's performance, while the female benefit manipulation decreased self-efficacy, which decreased men's performance. Self-efficacy was similarly expected to mediate the threat-performance relationship for women; however, the manipulation had no impact on women's self-efficacy scores. In fact, though it was a non-significant difference, females in the threat condition reported (although non-significantly) greater task self-efficacy than those in the control group. Since participants in this sample were highly identified with math and science and the efficacy measure was administered prior to the task, the women in the threat condition may have reacted to the threat manipulation with a desire to disprove the stereotype. However, once the task was administered, the women in the threat condition fell victim to performance decrements anyway, despite their high motivation. Most laboratory studies of stereotype threat have found support for increased motivation among the threatened participants (Steele & Aronson, 1995; Harder, 2000; Aronson et al., 1999). It is likely that increased motivation may have led to increased proximal regulation, reducing working memory, though the cognitive interference measure failed to accurately capture this effect. Future research should investigate whether threat increases proximal motivation among high-identified participants using a more appropriate measure.

Reducing Stereotype Threat

Perhaps the most valuable direction for future research would be to identify and evaluate possible ways to reduce stereotype threat without adversely affecting non-targets. While simply taking measures to ensure that stereotypes are not explicitly mentioned would seem to be a fair approach, results from this study and others (e.g., Smith and White, 2003) show that testing in general, because it is diagnostic of ability, may nevertheless activate threat. Though the effect is likely to be smaller than when threat is explicitly-induced, women may still underperform on standardized tests. It is important to note that in this study, demographic information was recorded only after all measures were completed, so any priming of gender identity by demographic survey (see

Steele & Aronson, 1995) cannot account for the results; rather, the standardized conditions alone depressed women's performance and self-efficacy. Removing the diagnosticity from a testing situation by telling participants the experimental task is not a "test" of abilities has been shown to be effective in reducing threat (Steele & Aronson, 1995), but it is unimaginable that test candidates sitting for the SAT or a selection measure, for example, would believe their skills and abilities are not being evaluated.

Results from the current study show that increasing task self-efficacy is a promising method for reducing threat in women. Self-efficacy manipulations had a significant impact on both male and female participants. Those who heard that their sex typically performed better on the GRE task reported much higher expectations for performance than those in the control condition. Participants (both male and female) who received the "benefit" instructions reported greater self-efficacy, which was associated with higher GRE performance as well as reduced stereotype activation. While past research has shown that stereotype threat can undermine target's task self-efficacy (Stangor, Carr, & Kiang, 1998), this is the first study to show that framing a task as showing "no past differences" can actually increase the performance expectancies of women.

How does increasing targets' self-efficacy improve performance? Self-efficacy impacts how information in the environment is interpreted, and targets with low self-efficacy may use situational cues to diagnose a given situation as threatening. Thus, people with low task efficacy may be more likely to call the stereotype to mind in ambiguous circumstances. When a task is difficult enough to produce a sense of frustration, targets low in task efficacy may be more likely to attribute their struggle to stereotypically low ability. Those with high efficacy, on the other hand, may attribute difficulty to other factors (such as lack of effort or poor strategy) and are more likely to remain task-focused. Low self-efficacy is also associated with higher self-regulation of performance and cognitive interference due to worry. Those with a high sense of self-efficacy, on the other hand, are more likely to visualize success, which provides a positive guide for performance. Efficacious people, furthermore, typically exhibit more analytic and problem-solving behavior than those with low task self-efficacy (Bandura and Jourden, 1991).

Future research should concentrate on ways in which threat can be reduced by increasing females' task self-efficacy. McIntyre, Paulson, & Lord (2003), for example, found that females performed significantly better than a control group when they read case studies about successful women in various careers prior to the experimental task. However, despite the fact that males' self-efficacy and performance were not adversely impacted by the "female benefit" manipulation, further research is needed before implementing a similar intervention. Actions designed to increase female self-efficacy, if administered to all test candidates, could be perceived as unfair to men. Future studies should investigate how to best design and administer interventions that are fair to all.

Organizational interventions to help employees effectively regulate emotions is one such area of research that deserves attention. Richards and Gross (2000) have shown that cognitive reappraisals of potentially threatening situations can reduce cognitive interference by preventing negative emotions from occurring. (Emotion suppression techniques—or actively trying not feel or show emotion once it emerges—on the other hand, were found to reduce working memory compared to control conditions in which participants simply experienced emotions as they occurred.) Training employees to regulate work-related emotions by reframing is therefore a promising direction for future research. Employees could be trained to view high-stakes evaluations as opportunities to gain feedback and improve performance compared to their own past performance standards, for example. Likewise, teaching students that intelligence is malleable rather than fixed—that skills improve with practice—is a potentially effective means of reducing threat.

Except for a few studies of college and high-school students, very little research to date has been conducted in the work place, and no research has been undertaken in employee selection or certification contexts. Future studies should concentrate on designing a methodology that is both fair and unobtrusive to participants in organizational settings. The only applied study of stereotype threat conducted in an organization-related setting, Roberson Deitch, Brief, & Block (2003) found that threat could be measured unobtrusively via a large-scale mail survey. Results show that threat is a concern for human resource professionals. African Americans who held token or solo status within their work group reported significantly greater stereotype threat (measured by degree of agreement with statements such as “Some people feel I have less ability because of my race”) than those who held relatively equal numerical status with other ethnic groups. Presumably, solo status increases the salience of ethnicity, thereby increasing the accessibility of the African American stereotype of poor ability. Importantly, stereotype threat impacted feedback seeking methods of targets, who were more likely to use indirect monitoring strategies (observing others’ reactions to their own behavior) rather than direct strategies such as asking a manager for his/her feedback. This puts targets at a disadvantage because indirect strategies provide less useful information about how to improve performance; indirect strategies can also serve to reduce performance by detracting attention and resources from one's current task. Thus, the research findings show two important ways in which targets' job performance can be negatively impacted by stereotype threat, a concept that deserves the attention of academic and applied psychologists alike.

Conclusion

Controversy has arisen over whether sex equality in math achievement is possible. It has been argued that the rapid increase in testosterone that emerges during adolescence may provide males with a biological advantage for math and spatial tasks. Meta-analyses of high-school students support this notion, as young men and women perform equally on all tests of academic achievement with the exception of upper-level mathematics (Feingold, 1988). In fact, males have been shown outperform females on the quantitative section of the Graduate Record Exam by more than half a standard deviation (Wah & Robinson, 1990). Yet women in this study performed as well as men did when a negative stereotype was counteracted, which shows that they do have the ability to succeed on an advanced standardized test.

However, measures of competence do not accurately predict measures of confidence in ability, particularly among the most talented women. Despite their equal level of ability (there was no preexisting difference in male and female SAT scores), a significant gender difference in task efficacy was found in this study overall, with males reporting much greater confidence in their math abilities than females. Task self-efficacy completely mediated any gender differences in performance.

Domain identification was also found to be an important predictor of performance on a difficult mathematical task. Those who were most identified with the task showed the highest levels of performance, yet significant sex differences in identification were found, with women showing significantly lower identification with math and science. If women are to overcome stereotype threat and succeed in mathematics, they must be able to reject the stereotype and remain identified with the domain. However, this study supports previous findings that the avoidance of mathematics and engineering domains occurs prior to and during college. Significant sex differences were found in types of academic major, showing that the women in the study had already distanced themselves from computer science and engineering domains. (See Table 17.)

One of the central propositions in Steele's theory is that the anxiety created by a negative stereotype, if consistently experienced, may lower interest in and liking for the domain over time (Steele, 1997). Targets exposed to negative stereotypes may come to accept them as applicable to the self, reducing self-efficacy in a particular domain, or they may lose interest in the domain and limit their experience with it to protect their self-esteem. Steele (1997) refers to this process as "disidentification." Harder (1999), for instance, found that women who were presented with a difficult set of math problems and told their performance would be used to evaluate their math ability showed greater anxiety and frustration with the task than those who were told that their individual performance would not be assessed. Whereas women in the "non-diagnostic" condition rated their interest in computer magazines and pursuing a career in math much higher after participation in the laboratory experiment, women in the evaluative condition (who experienced stereotype threat) rated these interests much lower after the experiment than they did prior to it.

Disidentification may explain why so few females enroll in advanced science and mathematics courses at the secondary and postsecondary level. The well-known example of Mattel's "Talking Barbie," who moaned "math class is tough" (Miller, 1992) illustrates how the pervasive stereotype of women's inferior mathematical abilities develops early in childhood; they are evident in children's stories, games, and attributions, and continue into secondary school and college (Phillips & Imhoff, 1997). Such negative stereotypes may be internalized, affecting girls' motivation, efficacy, and interest in these subject areas to the point that any skill in math or science is viewed as a suspiciously "unfeminine" trait, and young women therefore avoid displaying it.

Teachers and parents, furthermore, may encourage this withdrawal pattern by discouraging girls from enrolling in math and science courses for fear of failure or loss of self-esteem. For example, girls report less support from parents and teachers for their math interests than do boys (Hyde, Fennema, Ryan, Frost, and Hopp, 1990). Hewitt and Seymour (1991) found that teachers had lower expectations for the girl's achievement in math and science courses, calling on them in class less often than they did than boys. And parents believe that math is more difficult for their daughters than their sons (Frome & Eccles, 1998; Yee & Eccles, 1988); they are more likely to attribute a son's academic success to talent but a daughter's success to hard work and effort (Eccles & Jacobs, 1992). This is important because mothers' attitudes towards their children's abilities have been shown to have a greater influence on children's own perceptions of their abilities than do grades earned in math classes (Jacob & Eccles, 1992). Once internalized, such stereotypes may lead to a "learned helplessness" in which expectations for success are lowered.

If stereotype threat is going to be eradicated, changes in how girls and boys are raised to think about both academic domains and their own skills must be addressed. Previous studies show that chronic sex differences in motivational orientations exist—women and young girls are more likely to display anxiety-related motivation than males. These differences emerge early and are already evident in young children, with girls being more likely to avoid challenges (Licht et al, 1984; Legget, 1985), attribute failure to a lack of ability (Licht & Shapiro, 1982; Nicholls, 1979), and believe that intelligence is “fixed” as opposed to malleable (Leggett, 1985). And, consistent with Steele’s (1997) theory that threat is a problem experienced by the most talented women, researchers have found that high-ability girls are more likely to adopt performance avoidance orientations than low-ability students (Licht et al, 1984; Licht & Dweck, 1984; Licht & Shapiro, 1982; Stipek & Hoffman, 1980). Teachers might be trained, for example, to emphasize how intelligence (and thus classroom performance) is a malleable trait that is a function of practice and effort. As Osborne (2001) notes, teachers could remind students of the learning curve and how much they have improved over time prior to administering a test in order to increase self-efficacy and prime mastery goals.

While increasing task efficacy prior to task performance may prove effective as a means of temporarily combating the negative effects of stereotype threat, interventions should be employed at a younger age to engender a mastery orientation in young girls and prevent them from disidentifying with mathematics or other academic domains. Research that identifies the age and developmental stage at which stereotype threat begins to emerge in children would be helpful. It is likely that earlier interventions are more effective, but research with children of various ages is necessary to determine exactly when such an intervention would be most effective. No research investigating the origins of stereotype threat have been conducted.

Appendices

Appendix A

Motivational Trait Questionnaire

INSTRUCTIONS: This questionnaire asks you to respond to statements about your attitudes, opinions, and behaviors. Read each statement carefully, and decide whether or not the statement describes you. Using the scale at the top of each page indicate the degree to which the **ENTIRE** statement is true of you. Give only one answer for each statement.

Some of the statements may refer to experiences you may not have had. Respond to these statements in terms of how true you think it **WOULD BE** of you. Look at the sample statement below.

SAMPLE STATEMENT:

1 ▼	2 ▼	3 ▼	4 ▼	5 ▼	6 ▼
Very **UNTRUE** of Me	**UNTRUE** of Me	Somewhat **UNTRUE** of Me	Somewhat **TRUE** of Me	TRUE of Me	**Very TRUE** of Me

I like to go to parties.

MARK **1** ➡ *if you really dislike parties and you try to avoid them.*
2 ➡ *if you generally dislike parties and only go when you have to.*
3 ➡ *if you think parties are okay but generally prefer not to go.*
4 ➡ *if you think parties are okay and generally prefer to go.*
5 ➡ *if you generally like parties and go to most of the time.*
6 ➡ *if you really like parties and only miss one if you absolutely must.*

PLEASE NOTE:

- There are no right or wrong answers. Simply describe yourself honestly and state your opinions accurately.
- In deciding on your answer, consider your life in general and not only the last few weeks or months.
- Deciding on an answer may be difficult for some of the statements. If you have a hard time deciding, choose the answer that is **MOST** true of you.
- Some of the items will seem repetitive. These are not meant to be trick questions. Do not look back at your previous answers, simply answer each question honestly.

DO NOT TURN TO THE NEXT PAGE UNTIL YOU ARE TOLD TO DO SO

1 ▼	2 ▼	3 ▼	4 ▼	5 ▼	6 ▼
Very **UNTRUE** of Me	**UNTRUE** of Me	Somewhat **UNTRUE** of Me	Somewhat **TRUE** of Me	TRUE of Me	**Very TRUE** of Me

1. ____ When I become interested in something, I try to learn as much about it as I can.
2. ____ I set goals as a way to improve my performance.
3. ____ It really upsets me when someone does something better than I do.
4. ____ I perform best when I compete with others.
5. ____ When working on important projects, I am constantly fearful that I will make a mistake.
6. ____ If I know someone is judging me, I get so focused on how I am doing that I have difficulty concentrating on the task.
7. ____ When I am learning something new, I try to understand it completely.
8. ____ If I already do something well, I don't see the need to challenge myself to do better.
9. ____ I tend to put extra effort into tasks that involve competition with others.
10. ____ I am not a competitive person.
11. ____ I do not get nervous in achievement settings.
12. ____ My heart beats fast before I begin difficult tasks.
13. ____ Even when I have studied hard enough to get a good grade, I study more because I want to completely understand the material.
14. ____ When learning something new, I focus on improving my performance.
15. ____ It is important for me to outperform my co-workers.
16. ____ I try to avoid competitive situations.
17. ____ I am unconcerned even if I know that other people are forming an unfavorable impression of me.
18. ____ I have trouble relaxing because I worry about things at work.
19. ____ I like to take classes that challenge me.
20. ____ I compete with myself -- challenging myself to do things better than I have done before.
21. ____ Whether or not I feel good about my performance depends on how it compares to the performance of others.
22. ____ I would rather cooperate than compete.

1 ▼	2 ▼	3 ▼	4 ▼	5 ▼	6 ▼
Very **UNTRUE** of Me	**UNTRUE** of Me	Somewhat **UNTRUE** of Me	Somewhat **TRUE** of Me	TRUE of Me	**Very TRUE** of Me

23. ____ Before beginning an important project, I think of the consequences of failing.
24. ____ I am unable to concentrate fully in stressful situations.
25. ____ I am an intellectually curious person.
26. ____ I set high standards for myself and work toward achieving them.
27. ____ I am motivated to do things better than others.
28. ____ I like to turn things into a competition.
29. ____ I am afraid of other people noticing my shortcomings.
30. ____ I get headaches when I have a lot of important things to do.
31. ____ I prefer activities that provide me the opportunity to learn something new.
32. ____ I work hard at everything I undertake until I am satisfied with the result.
33. ____ I strive to do my job better than the people I work with.
34. ____ Even in non-competitive situations, I find ways to compete with others.
35. ____ I get nervous just thinking about having an important project evaluated.
36. ____ I am able to remain calm and relaxed in stressful situations.
37. ____ I am naturally motivated to learn.
38. ____ I do not set difficult goals for myself.
39. ____ I compare my performance to that of others.
40. ____ I worry about the possibility of failure.
41. ____ I am able to remain calm and relaxed before I take a test.
42. ____ I thirst for knowledge.
43. ____ My personal standards often exceed those required for the successful completion of a project.
44. ____ I get tense when other people assess my progress.
45. ____ I get an uneasy feeling in my stomach when working toward something I really want to accomplish.
46. ____ I worry about how others will view my work performance.
47. ____ I sleep because I am troubled by thoughts of failure.
48. ____ I am cautious about trying to do something that could lead to embarrassment.

Appendix B

Domain Identification Measure

Instructions: The following statements can be used by people to describe themselves. Please select the number to the right of the statement that you think appropriately describes you. There are no right or wrong answers.

	Strongly disagree						Strongly agree
1. I am good at math.	-3	-2	-1	0	1	2	3
2. I like computer programming.	-3	-2	-1	0	1	2	3
3. Success in engineering courses is important to me.	-3	-2	-1	0	1	2	3
4. I want to seek a career in science, math, or engineering when I graduate.	-3	-2	-1	0	1	2	3
5. Success in math courses is NOT important to me.	-3	-2	-1	0	1	2	3
6. I would describe myself as science- oriented.	-3	-2	-1	0	1	2	3
7. I consider myself above average in computer knowledge and experience.	-3	-2	-1	0	1	2	3
8. Getting good grades in my engineering courses is important to me.	-3	-2	-1	0	1	2	3
9. I would identify myself as mathematically-gifted.	-3	-2	-1	0	1	2	3
10. The fact that I *enjoy* math and science is an important aspect of my personality.	-3	-2	-1	0	1	2	3
11. The fact that I *am good at* math and science is an important aspect of my personality.	-3	-2	-1	0	1	2	3

	Strongly disagree						Strongly agree
12. I take pride in the fact that I excel at math tasks.	-3	-2	-1	0	1	2	3
13. I do NOT plan to pursue a career in anything related to math, science, or engineering.	-3	-2	-1	0	1	2	3
14. I would be embarrassed if I did not do well on a computer-related task.	-3	-2	-1	0	1	2	3
15. I would describe myself as a computer-oriented person.	-3	-2	-1	0	1	2	3
16. I consider myself above average in my mathematical ability.	-3	-2	-1	0	1	2	3
17. I feel a sense of pride in my computer knowledge and ability.	-3	-2	-1	0	1	2	3
18. I would be embarrassed if I were to perform poorly on an engineering-type task.	-3	-2	-1	0	1	2	3
19. The fact that I am intelligent is an important aspect of my self-concept.	-3	-2	-1	0	1	2	3
20. My favorite courses to study are in the fields of engineering, math, or science.	-3	-2	-1	0	1	2	3

Appendix C

Female Threat Condition Handout

The Brain Game

What's Sex Got to Do With It?

abcNEWS.com

July 31, 2002

— If you want a clear illustration of one of the many differences in the ways men and women think, a simple car ride will paint a pretty clear picture.

When Lori and Rich Boulware of Kendall Park, N.J., hit the road recently, their navigational radars were tuned into different frequencies. Rich used a mental map, while Lori used landmarks to get around. As the couple tried to get around a tricky area of town, Rich said, "Turn left on Webster," while Lori said, "You have to turn before the ice cream cone."

Dr. Helen Fisher, an expert in sex differences, says the Boulwares are not unusual in their navigational skills. Many scientists say it's all in our heads, or, more precisely, in the way men's and women's brains are designed and the way they function.

There are significant, differences that distinguish male and female brains. Male brains are wired to move information quickly within each side — or hemisphere — of the brain. This gives them better spatial abilities. They can see an object in space, and react quickly.

Hard-Wired in the Womb?

Scientists are developing new ways of looking inside the working brain — to see just how it's wired. Diagnostic tests such as Functional MRIs, which can measure blood flow, electrical activity and energy use, are being used to give researchers pictures of our brains in action.

Drs. Ruben and Raquel Gur, a husband and wife neuroscience team at the University of Pennsylvania, put men and women inside an MRI and studied how their brains responded to various verbal and spatial tasks.

In each case, the men's brains "lit up" in a few specific areas, while the women's brains showed activity in many areas — for both spatial and verbal tasks. Ruben Gur said the men's brain activity became completely focused, while women did exactly the opposite, activating other parts of their brain.

Researchers have found that the male brain's ability to focus on one area works better for spatial tests, while the female brain's approach is better for verbal tests. Scientists are still trying to figure out why that's the case.

The differences, researchers say, begin in the womb. At first, all fetuses' brains are virtually the same. At about nine weeks, however, testosterone surges through the male fetus, not only creating a boy's body but actually hard-wiring the brain to be male. Without testosterone to spur those changes, girls develop "female" brains.

Many researchers say the perception that men excel in motor and spatial skills while women are stronger in the verbal department is *not* just a stereotype.

Evolutionary scientists claim it all began with our ancient ancestors. Fisher said it all goes back to the hunter-gatherer days. Women needed verbal and emotional skills to cajole, educate and discipline their babies, while men needed spatial skills out on the hunt. "We've got an old brain in a very modern culture."

From the Classroom to Career Choices

Michael Lewis, director of the Institute for the Study of Child Development at the Robert Wood Johnson Medical School in New Jersey, has documented behavioral differences in children and adults. He notes that "early math really isn't math. It's really more language problems."

Once puberty hits, boys get a second surge of testosterone and their math and spatial abilities climb dramatically — but some researchers don't exactly know what the connection is. By the time high school kids take their SATs, boys outscore girls in the math section by 7 percent.

Fisher said, "It's quite remarkable how much better boys become at all kinds of spatial skills, mechanical skills, engineering skills, when that surge of testosterone comes on them."

Some researchers say these physiological differences may predispose men and women to gravitate toward certain careers. Fisher notes that despite the move toward equal employment opportunity in the U.S. job market, some 85 percent of the architects in America are still men, and 90 percent of the mechanics are still men. She said she's not at all surprised that men gravitate to those jobs that need and require mechanical spatial skills. Meanwhile, 94 percent of all speech therapists are women, and 99 percent of all pre-school and kindergarten teachers are female.

Appendix D

Female Benefit Condition Handout

The Brain Game

What's Sex Got to Do With It?

abcNEWS.com

July 31, 2002

— Are there really differences in the ways men and women think? And can men ever catch up?

When Lori and Rich Boulware of Kendall Park, N.J., hit the road recently, their navigational radars were tuned into different frequencies. Rich used a mental map, while Lori used landmarks to get around. As the couple tried to get around a tricky area of town, Rich said, "Turn left on Webster," while Lori said, "You have to turn before the ice cream cone."

Dr. Helen Fisher, an expert in sex differences, says the Boulwares are not unusual in their navigational skills. "Women go from one object to another. … A man will say, go two miles down the road and then head east."

Size Isn't Everything

A century ago, it was discovered that female brains were about 10 percent smaller than male brains, but we now know that size isn't everything when it comes to brainpower. In fact, the highest recorded I.Q. belongs to a woman, a writer named Marilyn vos Savant. Today, scientists are learning more about the human brain than ever before—including surprising new information that dispels traditional stereotypes about male and female abilities.

Hard-Wired in the Womb?

Scientists are developing new ways of looking inside the working brain — to see just how it's wired. Diagnostic tests such as Functional MRIs, which can measure blood flow, electrical activity and energy use, are being used to give researchers pictures of our brains in action.

Drs. Ruben and Raquel Gur, a husband and wife neuroscience team at the University of Pennsylvania, put men and women inside an MRI and studied how their brains responded to various verbal and spatial tasks. In each case, the men's brains "lit up" in only a few specific areas and for only spatial tasks, while the women's brains showed activity in many areas — for *both* spatial and verbal tasks.

The differences, researchers say, begin in the womb, where different hormones actually hard-wire the brain to be male or female. Once puberty hits, estrogen starts flooding the girls' bodies, and experts think that boosts helps them develop stronger memory skills. According to Fisher, a woman's verbal ability climbs rapidly during the middle of the monthly menstrual cycle, when estrogen levels peak.

Furthermore, in women's brains, areas of the cerebral cortex — linked to memory and judgment — are more densely packed with nerve cells than men's brains. This allows them to process that information more effectively.

Fisher explains that the corpus callosum, which she describes as a "big highway between the two sides of the brain," is larger in women toward the rear than it is in men. "Hence," she said, "the two sides of the brain are better interconnected" in women. This means that women can absorb and analyze all sorts of information from the environment simultaneously. This makes women more adept at multitasking and processing complex information.

Many researchers are finding the perception that men excel in motor and spatial skills is just an over-parodied stereotype. Many women, because they use different parts of their brain and because they use them more efficiently, have been shown to excel—not only verbal skills, but also in specific spatial and mathematical skills—to a greater degree than men.

Appendix E

Self-Efficacy Measure

1. I can complete 100% of the problems in 20 minutes.
   - ❑ Strongly disagree
   - ❑ Disagree
   - ❑ Neither agree nor disagree
   - ❑ Agree
   - ❑ Strongly agree

2. I can complete 90% of the problems in 20 minutes.
   - ❑ Strongly disagree
   - ❑ Disagree
   - ❑ Neither agree nor disagree
   - ❑ Agree
   - ❑ Strongly agree

3. I can complete 80% of the problems in 20 minutes.
   - ❑ Strongly disagree
   - ❑ Disagree
   - ❑ Neither agree nor disagree
   - ❑ Agree
   - ❑ Strongly agree

4. I can complete 70% of the problems in 20 minutes.
   - ❑ Strongly disagree
   - ❑ Disagree
   - ❑ Neither agree nor disagree
   - ❑ Agree
   - ❑ Strongly agree

5. I can complete 60% of the problems in 20 minutes.
   - ❑ Strongly disagree
   - ❑ Disagree
   - ❑ Neither agree nor disagree
   - ❑ Agree
   - ❑ Strongly agree

6. I can complete 50% of the problems in 20 minutes.
   - ❑ Strongly disagree
   - ❑ Disagree
   - ❑ Neither agree nor disagree
   - ❑ Agree
   - ❑ Strongly agree

7. I can complete 40% of the problems in 20 minutes.

❑ Strongly disagree ❑ Disagree ❑ Neither agree nor disagree ❑ Agree ❑ Strongly agree

8. I can complete 30% of the problems in 20 minutes.

❑ Strongly disagree ❑ Disagree ❑ Neither agree nor disagree ❑ Agree ❑ Strongly agree

9. I can complete 20% of the problems in 20 minutes.

❑ Strongly disagree ❑ Disagree ❑ Neither agree nor disagree ❑ Agree ❑ Strongly agree

10. I can complete 10% of the problems in 20 minutes.

❑ Strongly disagree ❑ Disagree ❑ Neither agree nor disagree ❑ Agree ❑ Strongly agree

Appendix F

GRE Problems

Numbers: All numbers used are real numbers.

Figures: Position of points, angles, regions, etc. can be assumed to be in the order shown; and angle measures can be assumed to be positive.

Lines shown as straight can be assumed to be straight.

Figures can be assumed to lie in a plane unless otherwise indicated.

Figures that accompany questions are intended to provide information useful in answering the questions. However, unless a note states that a figure is drawn to scale, you should solve these problems NOT by estimating sizes by sight or by measurement, but by using your knowledge of mathematics (see Example 2 below).

Directions: Each of the Questions 1-15 consists of two quantities, one in Column A and one in Column B. You are to compare the two quantities and choose

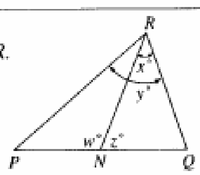
A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

Note: Since there are only four choices, NEVER MARK (E).

Common Information: In a question, there may be additional information, centered above the two columns, that concerns one or both of the quantities to be compared. A symbol that appears in both columns represents the same thing in Column A as it does in Column B.

	Column A	Column B	Sample Answers
Example 1:	2×6	$2 + 6$	● Ⓑ Ⓒ Ⓓ Ⓔ

Examples 2-4 refer to ΔPQR.



	Column A	Column B	Sample Answers
Example 2:	PN	NQ	Ⓐ Ⓑ Ⓒ ● Ⓔ (since equal measures cannot be assumed, even though PN and NQ appear equal)
Example 3:	x	y	Ⓐ ● Ⓒ Ⓓ Ⓔ (since N is between P and Q)

	Column A	Column B
1.	3^4	4^3

$$x = 2y + 3$$
$$y = -2$$

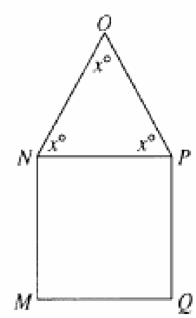
	Column A	Column B
2.	x	-1

$d = 5.03894$ and $\boxed{d}$ is the decimal expression d rounded to the nearest thousandth.

	Column A	Column B
3.	The number of decimal places where d and $\boxed{d}$ differ	4

$$x + 2y > 8$$

	Column A	Column B
4.	$2x + 4y$	20



Square $MNPQ$ has area 36.

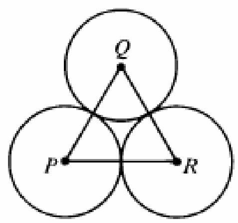
	Column A	Column B
5.	The perimeter of pentagon $MNOPQ$	30

p and q are different prime numbers. r is the le prime number greater than p, and s is the least prime number greater than q.

	Column A	Column B
6.	$r - p$	$s - q$

A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

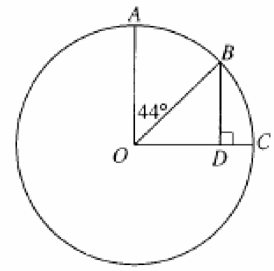
	Column A	Column B
	$\lvert -3 \rvert = -m$	
7.	m	3
	n is an even integer and a multiple of 3.	
8.	The remainder when n is divided by 12	6



Equilateral triangle PQR is formed by joining centers P, Q, and R of the circles. Each pair of circles has exactly one point in common.

	Column A	Column B
9.	The perimeter of triangle PQR	The circumference of the circle with center Q
10.	The volume of a cylindrical tank that has a radius of 2 meters and a height of 10 meters	The volume of a cylindrical tank that has a radius of 1 meter and a height of 20 meters
	$ds \neq 0$	
11.	The time required to travel d miles at s miles per hour	The time required to travel $\frac{d}{2}$ miles at $2s$ miles per hour
	$\triangle RST$ is isosceles and $\angle RST = 40°$.	
12.	The sum of the measures of the two angles of $\triangle RST$ that have equal measure	120°
13.	$\sqrt{x^4 + 6x^2 + 9}$	$x^2 + 3$

Column A Column B



O is the center of the circle and $\angle AOC$ is a right angle.

	Column A	Column B
14.	OD	BD

Before Maria changed jobs, her salary was 24 percent more than Julio's salary. After Maria changed jobs, her new salary was 24 percent less than her old salary.

	Column A	Column B
15.	Julio's salary	Maria's new salary

Directions: Each of the Questions 16-30 has five answer choices. For each of these questions, select the best of the answer choices given.

16. $(19 - 18 - 17 - 16) - (20 - 19 - 18 - 17) =$

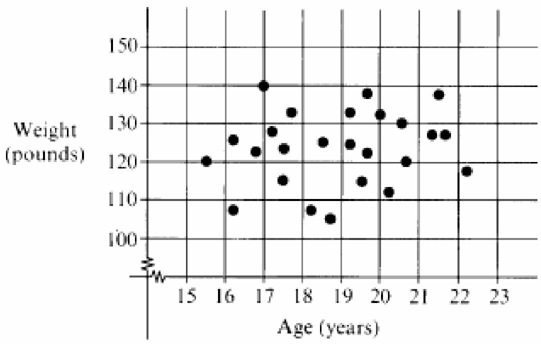
(A) -36
(B) -6
(C) -4
(D) 1
(E) 2

17. If $3x - 2 = 7$, then $4x =$

(A) 3
(B) 5
(C) $\frac{20}{3}$
(D) 9
(E) 12

18. Of the following, which is closest to $\sqrt[3]{30}$?

(A) 6
(B) 5
(C) 4
(D) 3
(E) 2



19. The dots on the graph above indicate age and weight for a sample of 25 students. What percent of these students are less than 19 years old and weigh more than 110 pounds?

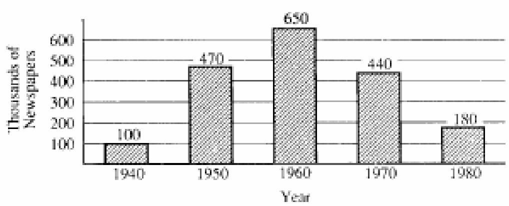
(A) 36% (B) 40% (C) 44% (D) 48% (E) 52%

20. The greatest number of diagonals that can be drawn from one vertex of a regular 6-sided polygon is

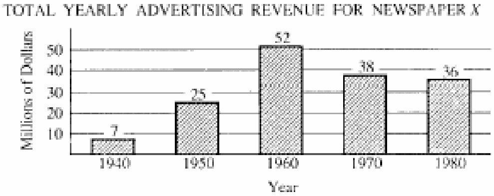
(A) 2 (B) 3 (C) 4 (D) 5 (E) 6

<u>Questions 21-25</u> refer to the following graphs.

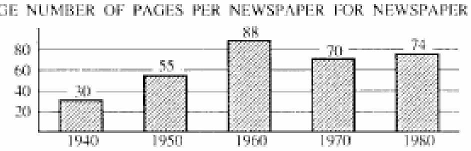
AVERAGE DAILY CIRCULATION FOR NEWSPAPER *X*



TOTAL YEARLY ADVERTISING REVENUE FOR NEWSPAPER *X*



AVERAGE NUMBER OF PAGES PER NEWSPAPER FOR NEWSPAPER *X*



21. In how many of the years shown was the average number of pages per newspaper at least twice as much as the average in 1940 ?

(A) Four
(B) Three
(C) Two
(D) One
(E) None

22. In 1950, if the printing cost per newspaper was \$0.05, what would have been the total cost of printing the average daily circulation?

(A) \$32,500
(B) \$26,000
(C) \$23,500
(D) \$22,000
(E) \$2,600

23. In 1980 the number of dollars of advertising revenue was how many times as great as the average daily circulation?

(A) 500
(B) 200
(C) 100
(D) 50
(E) 20

24. The percent decrease in average daily circulation from 1960 to 1970 was approximately

(A) 10%
(B) 12%
(C) 20%
(D) 26%
(E) 32%

25. Which of the following statements can be inferred from the data?

I. The greatest increase in total yearly advertising revenue over any 10-year period shown was \$27 million.
II. In each of the 10-year periods shown in which yearly advertising revenue decreased, average daily circulation also decreased.
III. From 1970 to 1980 the average number of pages per newspaper increased by 10.

(A) I only
(B) II only
(C) III only
(D) I and II
(E) II and III

26. If $0 < st < 1$, then which of the following can be true?

(A) $s < -1$ and $t > 0$
(B) $s < -1$ and $t < -1$
(C) $s > -1$ and $t < -1$
(D) $s > 1$ and $t < -1$
(E) $s > 1$ and $t > 1$

27. On segment WZ above, if $WY = 21$, $XZ = 26$, and YZ is twice WX, what is the value of XY?

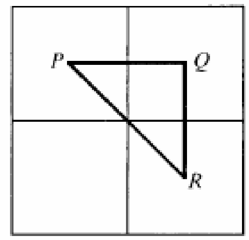
(A) 5
(B) 10
(C) 11
(D) 16
(E) It cannot be determined from the information given.

28. To reproduce an old photograph, a photographer charges x dollars to make a negative, $\frac{3x}{5}$ dollars for each of the first 10 prints, and $\frac{x}{5}$ dollars for each print in excess of 10 prints. If \$45 is the total charge to make a negative and 20 prints from an old photograph, what is the value of x?

(A) 3
(B) 3.5
(C) 4
(D) 4.5
(E) 5

29. Which of the following is equal to $\frac{1}{4}$ of 0.01 percent?

(A) 0.000025
(B) 0.00025
(C) 0.0025
(D) 0.025
(E) 0.25



30. In the figure above, each of the four squares has sides of length x. If $\triangle PQR$ is formed by joining the centers of three of the squares, what is the perimeter of $\triangle PQR$ in terms of x?

(A) $2x\sqrt{2}$

(B) $\frac{x\sqrt{2}}{2} + x$

(C) $2x + \sqrt{2}$

(D) $x\sqrt{2} + 2$

(E) $2x + x\sqrt{2}$

Appendix G

Cognitive Interference Questionnaire

Instructions: This questionnaire concerns the kinds of thoughts that go through people's heads at particular time, for example, while they are working on a task. The following is a list of thoughts, some of which you might have had *while working the task you have just completed*. Please indicate approximately how often each thought occurred to you while working on it by placing the appropriate number in the blank provided to the left of each question.

1 = Never 2 = Once 3 = A few times 4 = Often 5 = Very Often

______ 1. I thought about how poorly I was doing.

______ 2. I thought about what the experimenter would think of me.

______ 3. I thought about how I should work more carefully.

______ 4. I thought about how much time I had left.

______ 5. I thought about how others have done on this task.

______ 6. I thought about the difficulty of the problems.

______ 7. I thought about my level of ability.

______ 8. I thought about the purpose of the experiment.

______ 9. I thought about how I would feel if I were told how I performed.

______ 10. I thought about how often I got confused.

______ 11. I thought about other activities (for example, assignments, work).

______ 12. I thought about members of my family.

______ 13. I thought about friends.

______ 14. I thought about something that made me feel guilty.

______ 15. I thought about personal worries.

______ 16. I thought about something that made me feel tense.

______ 17. I thought about something that made me feel angry.

______ 18. I thought about something that happened earlier today.

______ 19. I thought about something that happened in the recent past (last few days, but not today).

______ 20. I thought about something that happened in the distant past.

______ 21. I thought about something that might happen in the future.

Please circle the number on the following scale which best represents the degree to which you felt your mind wandered during the task you have just completed.

Not at all 1 2 3 4 5 6 7 Very much

Appendix H

Stereotype Accessibility Measure

Instructions: Using any letter of the alphabet, place a letter in each of the blanks below to create a word. Please work quickly and write the letter that completes the first word that comes to mind. Feel free to skip a word if you find yourself spending too much time on it. Try to complete as many words as you can.

_ A R D	Hard (card; bard; ward)
D U _ _	Dumb (duck; dunk)
W_ _ K	Weak (work; week)
G _ R _	Girl (gore; guru; germ)
G_ _DE _	Sex (golden; graded; goaded; gilder)
F L _ _ _	Flake/Flaky (flank, flirt, fling, flung, flour)
_ _ _ H _ _ D	Airhead (offhand)
S W_ _ _	Sweet (sweat; swing; swish)
_ _ M A _ _	Female (tomato; remake; demand)
T O _ _ _	Token (touch; total; toast)
S O C _ _ _	Social (soccer; socket)
D _ _ _ Y	Ditzy (daisy)
_ _ N T L E	Gentle (mantle)
_ _ _ T U R E	Nurture (picture; torture)
_ _ _ _ S T	Sexist (assist; breast; rarest)
_ _ R I N G	Caring (string; daring)
DE_ _ _ A T E	Delicate (definite; delegate; decorate)
I N F _ _ _ _ _	Inferior (infinite; informal; infected)

C O _ _ _ _ A T E	Cooperate (correlate; coagulate)
S H _ _ _ I N G	Shopping (shipping; shouting; shelling)
_ _ _ T I O N	Emotion (mention; bastion)
I N D_ _ _ _ _ _ _ E	Indecisive (indefinite; indelicate)
_ _ _ _ _ T I V E	Sensitive (attentive; talkative; combative; plaintive; summative)
S U B M _ _ _ _ _ _	Submissive (submarines)

Appendix I

Post-Experimental Questionnaire

1. Please describe yourself.

____

____

____

____

____

2. The purpose of this study was:

____

____

____

____

____

Student Number: ____________________

Race (Circle one):
Caucasian African American Asian Hispanic Native American

Sex: Male Female

SAT:_____ Verbal:_____Quantitative: _____

GPA: _____

Major: ___________________________ Age: _____

Academic Year (Circle one): Freshman Sophomore Junior Senior

Appendix J

Experimental Results

Table 1

Abbreviation Key For all Variables of Study

Abbreviation	Variable
CONDTN	Condition
SEX	Participant Sex
STEREOTY	Number of Stereotype-Related Words Completed on Lexical Task
GRE	GRE task performance
IDENT	Domain Identification
EFF	Self-Efficacy
CI_REL	Task-Related Cognitive Interference
CI_IRR	Task-Irrelevant Cognitive Interference
CI_TTL	Total Cognitive Interference score
RACE	Participant Race
SAT	Participant SAT score
GPA	Participant GPA
MAJOR	Academic Major
AGE	Participant Age
YEAR	Class Year (freshman, sophomore, junior, or senior)
DESLERN	Desire to Learn
MASTGLS	Mastery Goals
PERSMAS	Personal Mastery Orientation
OTHEREF	Other-Referenced Goals
CMPSEEK	Competition Seeking
COMEXEL	Competitive Excellence
WORRY	Worry
EMOTION	Emotionality
MOTANXI	Motivation Related to Anxiety

Table 2

Intercorrelations Between All Continuous Variables of Study

	STEREOTY	GRE	IDENT	EFF	CI_REL	CI_IRR	CI_TTL	SAT	GPA	AGE
STEREOTY	1	-.041	-.129	-.180(*)	-.073	-.068	-.098	.079	.063	-.086
GRE	-.041	1	.382(**)	.300(**)	-.147	-.121	-.158	.529(**)	.158	-.179(*)
IDENT	-.129	.382(**)	1	.240(**)	.021	-.093	-.036	.243(**)	.099	.115
EFF	-.180(*)	.300(**)	.240(**)	1	-.151	.058	-.059	.269(**)	-.038	.032
CI_REL	-.073	-.147	.021	-.151	1	.287(**)	.851(**)	-.094	.118	.096
CI_IRR	-.068	-.121	-.093	.058	.287(**)	1	.747(**)	-.001	-.083	.116
CI_TTL	-.098	-.158	-.036	-.059	.851(**)	.747(**)	1	-.055	.043	.134
SAT	.079	.529(**)	.243(**)	.269(**)	-.094	-.001	-.055	1	-.009	-.027
GPA	.063	.158	.099	-.038	.118	-.083	.043	-.009	1	-.180(*)
AGE	-.086	-.179(*)	.115	.032	.096	.116	.134	-.027	-.180(*)	1
YEAR	.056	.016	.063	.061	-.066	-.018	-.049	.037	-.099	.749(**)
DESLERN	.030	.173(*)	.392(**)	.082	-.205(*)	-.199(*)	-.242(**)	.055	.133	.063
MASTGLS	-.028	.084	.186(*)	-.009	-.115	-.065	-.113	-.145	.249(**)	-.003
PERSMAS	.001	.142	.320(**)	.040	-.178(*)	-.145	-.196(*)	-.052	.213(*)	.033
OTHEREF	-.142	.146	.186(*)	.063	.257(**)	-.086	.142	-.022	.190(*)	.065
CMPSEEK	-.105	.142	.193(*)	.100	.062	-.006	.044	.056	.145	.054
COMEXEL	-.134	.157	.206(*)	.089	.172(*)	-.049	.100	.019	.182(*)	.064
WORRY	-.108	-.139	-.063	-.174(*)	.294(**)	-.021	.198(*)	-.159	.055	-.055
EMOTION	-.101	-.247(**)	-.079	-.204(*)	.396(**)	-.021	.262(**)	-.241(**)	-.014	-.031
MOTANXI	-.111	-.210(*)	-.076	-.205(*)	.370(**)	-.019	.249(**)	-.220(*)	.027	-.052

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 2, cont.

Intercorrelations Between All Continuous Variables of Study

	YEAR	DESLERN	MASTGLS	PERSMAS	OTHEREF	CMPSEEK	COMEXEL	WORRY	EMOTION	MOTANXI
STEREOTY	.056	.030	-.028	.001	-.142	-.105	-.134	-.108	-.101	-.111
GRE	.016	.173(*)	.084	.142	.146	.142	.157	-.139	-.247(**)	-.210(*)
IDENT	.063	.392(**)	.186(*)	.320(**)	.186(*)	.193(*)	.206(*)	-.063	-.079	-.076
EFF	.061	.082	-.009	.040	.063	.100	.089	-.174(*)	-.204(*)	-.205(*)
CI_REL	-.066	-.205(*)	-.115	-.178(*)	.257(**)	.062	.172(*)	.294(**)	.396(**)	.370(**)
CI_IRR	-.018	-.199(*)	-.065	-.145	-.086	-.006	-.049	-.021	-.021	-.019
CI_TTL	-.049	-.242(**)	-.113	-.196(*)	.142	.044	.100	.198(*)	.262(**)	.249(**)
SAT	.037	.055	-.145	-.052	-.022	.056	.019	-.159	-.241(**)	-.220(*)
GPA	-.099	.133	.249(**)	.213(*)	.190(*)	.145	.182(*)	.055	-.014	.027
AGE	.749(**)	.063	-.003	.033	.065	.054	.064	-.055	-.031	-.052
YEAR	1	.121	.042	.090	.051	.038	.048	-.105	-.129	-.133
DESLERN	.121	1	.613(**)	.895(**)	.027	-.021	.003	-.115	-.116	-.119
MASTGLS	.042	.613(**)	1	.902(**)	.196(*)	.249(**)	.242(**)	-.122	-.083	-.104
PERSMAS	.090	.895(**)	.902(**)	1	.126	.129	.139	-.132	-.111	-.124
OTHEREF	.051	.027	.196(*)	.126	1	.691(**)	.917(**)	.292(**)	.147	.244(**)
CMPSEEK	.038	-.021	.249(**)	.129	.691(**)	1	.922(**)	.013	-.080	-.034
COMEXEL	.048	.003	.242(**)	.139	.917(**)	.922(**)	1	.163	.034	.111
WORRY	-.105	-.115	-.122	-.132	.292(**)	.013	.163	1	.725(**)	.935(**)
EMOTION	-.129	-.116	-.083	-.111	.147	-.080	.034	.725(**)	1	.923(**)
MOTANXI	-.133	-.119	-.104	-.124	.244(**)	-.034	.111	.935(**)	.923(**)	1

Table 3

Intercorrelations Between All Continuous Variables of Study: Male Sample

	STEREOTY	GRE	EFF	CI_REL	CI_IRR	CI_TTL	SAT	GPA	AGE	YEAR
STEREOTY	1	.053	-.099	-.136	-.035	-.139	.148	.047	-.062	.094
GRE	.053	1	.291(*)	-.126	-.187	-.169	.348(**)	.352(**)	-.272(*)	.000
IDENT	-.113	.272(*)	.110	.082	-.150	-.017	.116	.268(*)	-.060	.008
EFF	-.099	.291(*)	1	-.060	.013	.010	.180	.030	-.049	.026
CI_REL	-.136	-.126	-.060	1	.229(*)	.846(**)	.024	.083	.145	-.061
CI_IRR	-.035	-.187	.013	.229(*)	1	.712(**)	.119	-.145	.296(**)	.076
CI_TTL	-.139	-.169	.010	.846(**)	.712(**)	1	.117	-.009	.277(*)	.013
SAT	.148	.348(**)	.180	.024	.119	.117	1	.020	-.120	-.023
GPA	.047	.352(**)	.030	.083	-.145	-.009	.020	1	-.150	.042
AGE	-.062	-.272(*)	-.049	.145	.296(**)	.277(*)	-.120	-.150	1	.739(**)
YEAR	.094	.000	.026	-.061	.076	.013	-.023	.042	.739(**)	1
DESLERN	.034	.149	.143	-.220	-.290(*)	-.301(**)	-.029	.239(*)	.029	.166
MASTGLS	.013	.175	.025	-.192	-.199	-.244(*)	-.038	.328(**)	.029	.129
PERSMAS	.026	.174	.095	-.224	-.267(*)	-.296(**)	-.036	.303(**)	.031	.161
OTHEREF	-.059	.234(*)	.094	.307(**)	-.067	.210	-.028	.157	.061	.104
CMPSEEK	.004	.264(*)	.121	.074	.061	.097	.191	.183	.011	.011
COMEXEL	-.030	.276(*)	.119	.210	-.002	.169	.092	.190	.040	.063
WORRY	-.124	-.086	-.217	.416(**)	.155	.396(**)	-.170	-.054	-.004	-.069
EMOTION	-.042	-.184	-.270(*)	.509(**)	.101	.419(**)	-.159	-.069	.027	-.116
MOTANXI	-.090	-.139	-.255(*)	.486(**)	.136	.428(**)	-.174	-.064	.011	-.096

Table 3, cont.

Intercorrelations Between All Continuous Variables of Study: Male Sample

	DESLERN	MASTGLS	PERSMAS	OTHEREF	CMPSEEK	COMEXEL	WORRY	EMOTION	MOTANXI
STEREOTY	.034	.013	.026	-.059	.004	-.030	-.124	-.042	-.090
GRE	.149	.175	.174	.234(*)	.264(*)	.276(*)	-.086	-.184	-.139
IDENT	.513(**)	.295(**)	.444(**)	.189	.234(*)	.234(*)	-.019	-.048	-.035
EFF	.143	.025	.095	.094	.121	.119	-.217	-.270(*)	-.255(*)
CI_REL	-.220	-.192	-.224	.307(**)	.074	.210	.416(**)	.509(**)	.486(**)
CI_IRR	-.290(*)	-.199	-.267(*)	-.067	.061	-.002	.155	.101	.136
CI_TTL	-.301(**)	-.244(*)	-.296(**)	.210	.097	.169	.396(**)	.419(**)	.428(**)
SAT	-.029	-.038	-.036	-.028	.191	.092	-.170	-.159	-.174
GPA	.239(*)	.328(**)	.303(**)	.157	.183	.190	-.054	-.069	-.064
AGE	.029	.029	.031	.061	.011	.040	-.004	.027	.011
YEAR	.166	.129	.161	.104	.011	.063	-.069	-.116	-.096
DESLERN	1	.707(**)	.933(**)	.023	-.056	-.018	-.169	-.114	-.151
MASTGLS	.707(**)	1	.914(**)	.187	.202	.215	-.235(*)	-.157	-.210
PERSMAS	.933(**)	.914(**)	1	.108	.071	.099	-.217	-.145	-.193
OTHEREF	.023	.187	.108	1	.634(**)	.902(**)	.345(**)	.243(*)	.314(**)
CMPSEEK	-.056	.202	.071	.634(**)	1	.906(**)	.005	-.057	-.025
COMEXEL	-.018	.215	.099	.902(**)	.906(**)	1	.192	.102	.158
WORRY	-.169	-.235(*)	-.217	.345(**)	.005	.192	1	.791(**)	.953(**)
EMOTION	-.114	-.157	-.145	.243(*)	-.057	.102	.791(**)	1	.939(**)
MOTANXI	-.151	-.210	-.193	.314(**)	-.025	.158	.953(**)	.939(**)	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 4

Intercorrelations Between All Continuous Variables of Study: Female Sample

	STEREOTY	GRE	IDENT	EFF	CI_REL	CI_IRR	CI_TTL	SAT	GPA	AGE	YEAR
STEREOTY	1	-.151	-.128	-.265(*)	.001	-.109	-.059	.011	.083	-.108	.018
GRE	-.151	1	.387(**)	.232	-.141	-.041	-.116	.706(**)	-.033	-.126	.000
IDENT	-.128	.387(**)	1	.207	.022	-.027	.000	.261(*)	.009	.218	.059
EFF	-.265(*)	.232	.207	1	-.218	.117	-.080	.281(*)	-.081	.065	.061
CI_REL	.001	-.141	.022	-.218	1	.351(**)	.856(**)	-.203	.150	.055	-.059
CI_IRR	-.109	-.041	-.027	.117	.351(**)	1	.785(**)	-.095	-.021	-.083	-.107
CI_TTL	-.059	-.116	.000	-.080	.856(**)	.785(**)	1	-.184	.088	-.010	-.098
SAT	.011	.706(**)	.261(*)	.281(*)	-.203	-.095	-.184	1	-.022	.024	.075
GPA	.083	-.033	.009	-.081	.150	-.021	.088	-.022	1	-.208	-.229
AGE	-.108	-.126	.218	.065	.055	-.083	-.010	.024	-.208	1	.762(**)
YEAR	.018	.000	.059	.061	-.059	-.107	-.098	.075	-.229	.762(**)	1
DESLERN	.028	.193	.291(*)	-.039	-.173	-.061	-.148	.160	-.022	.107	.050
MASTGLS	-.081	.021	.187	-.002	-.040	.075	.015	-.235	.157	-.025	-.035
PERSMAS	-.040	.109	.265(*)	-.020	-.111	.019	-.063	-.076	.092	.037	.002
OTHEREF	-.243(*)	-.012	.101	-.026	.224	-.100	.093	-.063	.249(*)	.042	-.024
CMPSEEK	-.232	-.056	.079	.022	.074	-.072	.009	-.147	.126	.078	.048
COMEXEL	-.253(*)	-.037	.096	-.002	.157	-.092	.053	-.114	.198	.064	.014
WORRY	-.099	-.144	.002	-.070	.124	-.227	-.044	-.086	.164	-.086	-.125
EMOTION	-.182	-.258(*)	.014	-.074	.251(*)	-.162	.077	-.271(*)	.019	-.057	-.117
MOTANXI	-.152	-.232	.004	-.085	.208	-.209	.022	-.209	.108	-.091	-.149

Table 4, cont.

Intercorrelations Between All Continuous Variables of Study: Female Sample

	DESLERN	MASTGLS	PERSMAS	OTHEREF	CMPSEEK	COMEXEL	WORRY	EMOTION	MOTANXI
STEREOTY	.028	-.081	-.040	-.243(*)	-.232	-.253(*)	-.099	-.182	-.152
GRE	.193	.021	.109	-.012	-.056	-.037	-.144	-.258(*)	-.232
IDENT	.291(*)	.187	.265(*)	.101	.079	.096	.002	.014	.004
EFF	-.039	-.002	-.020	-.026	.022	-.002	-.070	-.074	-.085
CI_REL	-.173	-.040	-.111	.224	.074	.157	.124	.251(*)	.208
CI_IRR	-.061	.075	.019	-.100	-.072	-.092	-.227	-.162	-.209
CI_TTL	-.148	.015	-.063	.093	.009	.053	-.044	.077	.022
SAT	.160	-.235	-.076	-.063	-.147	-.114	-.086	-.271(*)	-.209
GPA	-.022	.157	.092	.249(*)	.126	.198	.164	.019	.108
AGE	.107	-.025	.037	.042	.078	.064	-.086	-.057	-.091
YEAR	.050	-.035	.002	-.024	.048	.014	-.125	-.117	-.149
DESLERN	1	.511(**)	.827(**)	.014	.012	.014	-.003	-.107	-.044
MASTGLS	.511(**)	1	.906(**)	.233	.331(**)	.303(*)	-.021	-.040	-.015
PERSMAS	.827(**)	.906(**)	1	.159	.223	.205	-.016	-.079	-.032
OTHEREF	.014	.233	.159	1	.748(**)	.931(**)	.278(*)	.088	.212
CMPSEEK	.012	.331(**)	.223	.748(**)	1	.938(**)	.063	-.063	.001
COMEXEL	.014	.303(*)	.205	.931(**)	.938(**)	1	.180	.011	.111
WORRY	-.003	-.021	-.016	.278(*)	.063	.180	1	.639(**)	.910(**)
EMOTION	-.107	-.040	-.079	.088	-.063	.011	.639(**)	1	.901(**)
MOTANXI	-.044	-.015	-.032	.212	.001	.111	.910(**)	.901(**)	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 5

Descriptive Statistics for Male Sample

	Mean				Std. Deviation			
	Total	Threat (N = 30)	Control (N = 25)	Benefit (N = 22)	Total	Threat (N = 30)	Control (N = 25)	Benefit (N = 22)
Dependent Variables								
STEREOTY	5.64	5.43	6.08	5.41	2.16	2.13	2.45	1.84
GRE	19.05	18.73	19.36	19.14	4.79	4.54	5.11	4.94
EFF	16.44	17.80	15.6	15.55	4.04	2.58	4.86	4.33
CI_REL	23.22	22.73	23.08	24.05	6.31	7.80	5.75	4.62
CI_IRR	14.13	15.27	13.75	13.00	4.77	5.59	3.95	4.21
CI_TTL	37.25	38.00	36.50	37.05	8.72	11.54	5.83	7.01
Independent Variables								
SAT	1313.29				110.04			
GPA	2.95				.59			
AGE	20.34				1.81			
IDENT	23.53				18.72			
DESLERN	34.73				6.35			
MASTGLS	35.32				5.63			
PERSMAS	70.05				11.07			
OTHEREF	33.55				7.44			
CMPSEEK	27.75				7.56			
COMEXEL	61.30				13.56			
WORRY	37.29				8.15			
EMOTION	28.48				7.19			
MOTANXI	65.77				14.51			

Table 6

Descriptive Statistics for Female Sample

	Mean				Std. Deviation			
	Total	Threat (N = 19)	Control (N = 22)	Benefit (N = 25)	Total	Threat (N = 19)	Control (N = 22)	Benefit (N = 25)
Dependent Variables								
STEREOTY	5.74	5.68	5.95	5.60	2.06	2.31	2.28	1.71
GRE	17.00	15.26	17.23	18.12	4.61	3.90	3.80	5.45
EFF	14.26	14.32	12.95	15.36	4.66	4.62	4.40	4.79
CI_REL	24.24	24.37	25.78	22.8	6.01	6.54	6.36	5.11
CI_IRR	14.41	14.79	13.05	15.32	5.02	4.71	3.67	6.09
CI_TTL	38.65	39.16	38.82	38.12	9.08	10.14	9.03	8.63
Independent Variables								
SAT	1270.48				107.23			
GPA	3.02				.61			
AGE	19.95				1.76			
IDENT	6.28				23.21			
DESLERN	34.00				4.32			
MASTGLS	36.26				5.75			
PERSMAS	70.26				8.78			
OTHEREF	31.76				7.26			
CMPSEEK	25.80				7.64			
COMEXEL	57.56				13.93			
WORRY	39.65				8.018			
EMOTION	31.03				7.66			
MOTANXI	70.55				14.19			

Table 7

Analyses of Variance of Sex Differences in GRE Performance by Condition

Source	SS	*df*	MS	*F*	*p*
Female Threat					
Between	140.08	1	140.08	7.55	.008
Within	871.55	47	18.54		
Total	1011.63	48			
Control					
Between	53.23	1	53.23	2.578	.115
Within	929.624	45	20.66		
Total	982.85	46			
Female Benefit					
Between	12.09	1	12.09	.444	.509
Within	1225.23	45	27.23		
Total	1237.32	46			

Table 8

Analyses of Variance for Females' Self-Efficacy by Experimental Condition

Source			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		67.801	2	33.901	1.593	.211
	Linear Term	Unweighted	11.771	1	11.771	.553	.460
		Weighted	16.105	1	16.105	.757	.388
		Deviation	51.696	1	51.696	2.429	.124
Within Groups			1340.820	63	21.283		
Total			1408.621	65			

Contrast Test

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
EFFICACY	Assume equal variances	Benefit vs. Control	2.4055	1.34860	1.784	63	.079
	Does not assume equal variances	Benefit vs. Control	2.4055	1.34057	1.794	44.901	.079

Table 9

Analyses of Variance for Males' Self-Efficacy by Experimental Condition

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		90.732	2	45.366	2.914	.061
	Linear Term	Unweighted	64.515	1	64.515	4.143	.045
		Weighted	71.456	1	71.456	4.589	.035
		Deviation	19.276	1	19.276	1.238	.269
Within Groups			1152.255	74	15.571		
Total			1242.987	76			

Contrast Test

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
EFFICACY	Assume equal variances	Threat vs. Control	-2.2000	1.06858	-2.059	74	.043
	Does not assume equal variances	Threat vs. Control	-2.2000	1.07933	-2.038	35.002	.049

Table 10

Summary of Hierarchical Regression Analyses for Variables Predicting Males' Self-Efficacy (N = 76)

Variable	*B*	*Std. Error*	β	*t*	*Sig.*
Exp. Condition	-1.182	.553	-.240	-2.139	.036

Summary of Hierarchical Regression Analyses for Variables Predicting Males' GRE Performance (N = 77)

Variable		*B*	*Std. Error*	β	*t*	*Sig.*
Step 1	Self-Efficacy	.345	.131	.291	2.635	.010
Step 2	Self-Efficacy	.377	.135	.319	2.799	.007
	Exp. Condition	.669	.664	.115	1.007	.317

Mediation of the Threat-Performance Relationship by Self-Efficacy for Males

Original Relationship:	Exp. Condition	→	GRE Performance	$R^2 = .038$	p = .74
Step 1:	Exp. Condition	→	Self-Efficacy	$R^2 = .057$	p = .036
Step 2:	Self efficacy	→	GRE performance	$R^2 = .085$	p = .010
Step 3:	Self efficacy & Exp. condition	→	GRE performance	$R^2 = .097$	p = .023
Step 4:	Test the significance of $\triangle R^2$			$F(2, 74) = 1.015, p > .05$	

Table 11

Summary of Hierarchical Regression Analysis for Variables Predicting Self-Efficacy (N = 142)

							Change Statistics			
		B	*Std. Error*	β	*t*	*Sig.*	R^2 Change	F Change	*df*	*Sig.* F Change
Model 1							.061	4.530	2,140	.012
	Benefit contrast	.251	.778	.027	.323	.747				
	SEX	-2.207	.733	-.248	-3.010	.003				
Model 2							.025	3.864	1, 139	.051
	Benefit contrast	-4.284	2.432	-.453	-1.761	.080				
	SEX	-3.215	.889	-.361	-3.618	.000				
	Benefit X SEX	3.029	1.541	.528	1.966	.051				

Table 12

Analysis of Variance for Sex Differences in Domain Identification and GRE Performance

Source	Dependent Variable	SS	*df*	MS	*F*	*p*
Sex	GRE	170.021	1	170.021	7.704	.006
	IDENT	10332.258	1	10332.258	23.679	.000
Error	GRE	3045.372	138	22.068		
	IDENT	60215.885	138	436.347		
Total	GRE	49117.000	140			
	IDENT	104806.000	140			

Table 13

Summary of Hierarchical Regression Analyses for Variables Predicting Domain Identification (N = 140)

Variable	*B*	*Std. Error*	β	*t*	*Sig.*
Sex	-17.245	3.544	-.383	-4.866	.000

Summary of Hierarchical Regression Analyses for Variables Predicting GRE Performance (N = 139)

Variable		*B*	*Std. Error*	β	*t*	*Sig.*
Step 1	Domain Identification	.082	.017	.382	4.854	.000
Step 2	Domain Identification	.074	.018	.344	4.048	.000
	Sex	-.945	.818	-.098	-1.154	.250

Mediation of the Sex-Performance Relationship by Domain Identification

Step 1:	Sex	→	Domain Identification	R^2 = .146	p = .000
Step 2:	Domain Identification	→	GRE performance	R^2 = .146	p = .000
Step 3:	Domain Identification & Sex	→	GRE performance	R^2 = .154	p = .000
Step 4:	Test the significance of $\triangle R^2$			F(2, 137) = 12.47, p > .05	

Table 14

Summary of Hierarchical Regression Analyses for Variables Predicting Self-Efficacy (N = 142)

Variable	*B*	*Std. Error*	β	*t*	*Sig.*
Sex	-2.184	.727	-.245	-3.002	.003

Summary of Hierarchical Regression Analyses for Variables Predicting GRE Performance (N = 139)

Variable		*B*	*Std. Error*	β	*t*	*Sig.*
Step 1	Self-Efficacy	.323	.087	.300	3.730	.000
Step 2	Self-Efficacy	.283	.089	.263	3.200	.002
	Sex	-1.433	.789	-.149	-1.817	.071

Mediation of the Sex-Performance Relationship by Self-Efficacy

Step 1:	Sex	→	Self-Efficacy	$R^2 = .060$	p = .003
Step 2:	Self-Efficacy	→	GRE performance	$R^2 = .090$	p = .000
Step 3:	Self-Efficacy & Sex	→	GRE performance	$R^2 = .111$	p = .000
Step 4:	Test the significance of $\triangle R^2$			F (1, 140) = 3.302, p > .05	

Table 15

Analyses of Variance for Sex Differences in Task Self-Efficacy by Experimental Condition

Source	SS	*df*	MS	*F*	*p*
Female Threat					
Between	141.217	1	141.217	11.505	.001
Within	576.905	47	12.275		
Total	718.122	48			
Control					
Between	81.897	1	81.897	3.788	.058
Within	972.955	45	21.621		
Total	1054.851	46			
Female Benefit					
Between	.402	1	.402	.019	.890
Within	943.215	45	20.960		
Total	943.617	46			

Table 16

Differences in Domain Identification as a Function of Ability

N Tiles of IDENT	Mean SAT	N	Std. Deviation
Low	1269.11	45	108.917
Medium	1290.89	45	113.735
High	1321.40	43	99.369
Total	1293.38	133	108.933

Table 17

Sex Differences in Academic Major

		MAJOR								
		Computer Science	Engineering	Psychology	Science	Liberal Arts	Management	Architecture	Undecided	Total
SEX	Male	15 (19.5%)	44 (57.1%)	2 (2.6%)	7 (9.1%)	2 (2.6%)	3 (3.9%)	3 (3.9%)	1 (1.3%)	77 (100%)
	Female	4 (6.1%)	21 (31.8%)	5 (7.6%)	9 (13.6%)	7 (10.6%)	10 (15.2%)	6 (9.1%)	4 (6.1%)	66 (100%)
Total		19 (13.3%)	65 (45.5%)	7 (4.9%)	16 (11.2%)	9 (6.3%)	13 (9.1%)	9 (6.3%)	5 (3.5%)	143 (100%)

References

American Association of University Women. (2000). *Tech Savvy: Educating Girls in the New Computer Age*.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing effects of stereotype threat on American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38* (2), 113-125.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C., and 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.

Aronson, Quinn, & Spencer. (1998). Stereotype threat and the academic underperformance of minorities and women. In Swim, J. K., & Stangor, C. (Eds.), *Prejudice: The target's perspective* (pp. 83-103). New York: Academic Press.

Arvey, R. D., & Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, *43* (4), 695-716.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64,* 359-372.

Bandura, A. (2000). Self-efficacy: Foundation of agency. In W. J. Perrig and A. Grob (Eds.), *Control of human behavior, mental processes, and consciousness* (pp.17-33). New Jersey: Erlbaum Associates.

Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision-making. *Journal of Personality and Social Psychology*, *60* (6), 941-951.

Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361-382). New York: Guilford Press.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1-40). New Jersey: Erlbaum.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71* (2), 230-244.

Bargh, J. A., & Barndollar, K. (1996). Automaticity in action: The unconscious as a repository of chronic goals and motives. In P. M. Golliwitzer and J. A. Bargh (Eds.), *The Psychology of action: Linking motivation to behavior*. New York: Guilford Press.

Baron, R. S. (1986). Distraction-conflict theory: Progress and problems. *Advanced Experimental Social Psychology*, *19*, 1-40.

Baumeister, R. F. (1984). Choking under pressure: Self-conscious and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46* (3), 610-620.

Beyer, S. (1999). The accuracy of academic sex stereotypes. *Sex Roles*, *40*, 787-813.

Blair, D.V., O'Neil, H.F., & Price, D.J. (1999). Effects of expertise on state self-efficacy and state worry during a computer-based test. *Computers in Human Behavior*, *15*, 511-530.

Blascovich, J., Quinn, D. M., Spencer, S. J., & Steele, C. M. (1998). *The effect of stereotype threat on African Americans' blood pressure*. Unpublished manuscript, University of California at Santa Barbara.

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, *52* (3), 561-589.

Brewer, M. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, *17* (5), 475-482.

Broadnax, S., Crocker, J., & Spencer, S. J. (1997). *African Americans and academic underperformance: The mediational role of stereotype vulnerability*. Unpublished manuscript, State University of New York at Buffalo.

Broderick, P. C. (1998). Early adolescent sex differences in the use of ruminative and distracting coping strategies. *Journal of Early Adolescence*, *18* (2), 173-191.

Brown, J. L. (2000). Performance expectations are not a necessary mediator of stereotype threat in African American verbal test performance. (Doctoral dissertation, Stanford University, 2000). *UMI Proquest Digital Dissertations*, 9995186.

Brown, R.Y., & Josephs, R.A. (1999). A burden of proof: Stereotype relevance and sex differences in math performance. *Journal of Personality and Social Psychology*, *76*, 246-257.

Bulman-Fleming, B. M., Grimshaw, G. M., & Berenbaum, S. A. (2000). Achieving convergent evidence through divergent approaches. *Brain & Cognition Special Issue: Millennium Issue, 42* (1), 85-88.

Butler, J. L., & Baumeister (1998). The trouble with friendly faces: Skilled performance with a supportive audience. *Journal of Personality and Social Psychology, 75* (5), 1213-1230.

Camp, T. (2001). Women in Computer Science: Reversing the trend. *Syllabus, 15* (1), 24-28.

Camp, T., Miller, K., & Davies, V. (2000). *The incredible shrinking pipeline unlikely to reverse*. Unpublished manuscript, Colorado School of Mines.

Cascio, W. F. (1995). Whither an Industrial and Organizational Psychologist in the changing world of work? *American Psychologist*, *50* (11), 928-939.

Catsambis, S. (1994). The path to math: Sex and racial-ethnic differences in mathematics participation from middle school to high school. *Sociology of Education*, 67, 199-215.

Chan, D. (1997) Racial subgroup differences in predictive validity perceptions on personality and cognitive ability tests. *Journal of Applied Psychology*, *82* (2), 311-320.

Chan, D., Schmitt, N., DeShon, R.P., Clause, C.S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationship between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, *82* (2), 300-310.

Chan, D., Schmitt, N.S., Sacco, J.M., & DeShon, R.P. (1998). Understanding pretest and posttest reactions to cognitive ability and personality tests. *Journal of Applied Psychology, 83* (3), 471-485.

Chen, G., Gully, S. M., Whiteman, J.A., & Kilcullen, R.N. (2000). Examination of relationships among trait-like individual differences, state-like individual differences, and learning performance.

Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology*, *33*, 541-560.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Erlbaum.

Cortina, J. M., & DeShon, R. P. (1998). Determining relative importance of predictors with the observational design. Journal of Applied Psychology, *83* (5), 798-804.

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

Davies, P. G., Spencer, S. J., & Quinn, D.M. (2002). All consuming images: How demeaning commercials that elicit stereotype threat can restrain women academically and professionally. *Journal of Personality and Social Psychology*, *83* (6), 1281-1297.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56* (1), 5-18.

Dijksterhuis, A., Aarts, H., Bargh, J. A., & van Knippenberg, A. (2000). On the relation between associative strength and automatic behavior. *Journal of Experimental Social Psychology, 36*, 531–544.

Dweck, C. S. (1985). Motivational processes affecting learning. *American Psychologist*, *41* (10), 1040-1048.

Dykeman, B. F. (1994). The effects of motivational orientation, self-efficacy, and feedback condition on test anxiety. *Journal of Instructional Psychology*, *21* (2), 114-119.

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

Eccles, J.S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109-132.

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality & Social Psychology*, *72*(1), 218-232.

Elliot, A. J., & McGregor, H. A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *76* (4), 628-644.

Elliott, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: Approach and avoidance temperaments and goals. Journal of Personality and Social Psychology, 82 (5), 804-818.

Ernest, J. (1976). *Mathematics and Sex*. Berkeley: University of California Press.

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

Fisher, A., Margolis, J., & Miller, F. (1997). Undergraduate women in Computer *Science*: *Experience, motivation, and culture*.

Foote, W. E. (2000). The effect of stereotype threat on women's math performance: An examination of mediating and moderating variables. (Doctoral dissertation, California State University, 2000). *Digital Dissertations*, AAT 1398785.

Forster, J., Grant, H., Idson, C., & Higgins, E.T. (2001). Success/failure feedback, expectancies, and approach/avoidance motivation: How regulatory focus moderates classic relations. *Journal of Experimental Social Psychology, 37* (3), 253-260.

Forster, J., Higgins, E. T., & Idson, L. C. (1998). Approach and avoidance strength during goal attainment: Regulatory focus and the "goal looms larger" effect. *Journal of Personality and Social Psychology, 75* (5), 1115-1131.

Gallagher, A., Bridgeman, B., & Cahalan, C. (2000). The effect of computer-based tests on racial-ethnic and sex groups. *Journal of Educational Measurement*, *39*(2), 133-147.

Geen, R. G. (1991). Social Motivation. *Annual Review of Psychology, 42*, 377-399.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, *28* (5), 659-670.

Goodman Research Group, Inc. (2002, April). *Final report of the women's experiences in college engineering project*. Cambridge, MA: Goodman, I. F., Cunningham, C.M., Lachapelle, C., Thompson, M., Bittinger, K., Brennan, R. T., & Delci, M.

Gur, R.C., Turetsky, B.I. Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. (1999). Sex differences in brain gray and white matter in healthy young adults. *Journal of Nueroscience, 19*, 4065-4072.

Hammond, R. (2001). Computer Science opens its doors to women. *Carnegie Mellon Magazine*, *19* (3), 13-17.

Harder, J. A. (1999). The effect of private versus public evaluation on stereotype threat for women in mathematics. (Doctoral dissertation, The University of Texas, 1999). *Dissertation Abstracts International*, *60* (9-A), 3270.

Heggestad, E. D., & Kanfer, R. (2000). Individual differences in trait motivation: Development of the Motivational Trait Questionnaire. *International Journal of Educational Research, 33*, 751-776.

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

Hilton, J., & Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, *47*, 237-271.

Hong, E. (1999). Test anxiety, perceived test difficulty, and test performance: Temporal patterns of their effects. *Learning & Individual Differences, 11*(4), 431-447.
Howard, J. & Hammond, R. (1985). Rumors of inferiority. *New Republic*, *72*, 18-23.

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

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

Jaccard, J., & Turrisi, R. (2003). Interaction effects in multiple regression. Thousand Oaks, CA: Sage.

Jackson, S. E., & Zedeck, S. (1982). Explaining performance variability: Contributions of goal setting, task characteristics, and evaluative contexts. *Journal of Applied Psychology*, *67* (6), 759-768.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, *69* (2), 307-321.

Johnson, W. (2000, November 16). Women key to skills gap. *Computer Weekly*.

Information Technology Association of America (2001). When Can You Start? Building Better Information Technology Skills and Careers. April, 2001.

Information Technology Association of America. (2000). Bridging the gap: Information technology skills in the new millennium. April, 2000.

Information Technology Association of America (1999). Building the 21st Century Information Technology Work Force: Underrepresented Groups in the Information Technology Workforce.

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* (5), 365-371.

Kaiser, C. R., & Miller, C. T. (2001). Reacting to impending discrimination: Compensation for prejudice and attributions to discrimination. *Personality and Social Psychology Bulletin*, *27* (10), 1357-1367.

Kanfer, R. & Ackerman, P. L. (2000) Individual differences in work motivation: Further explorations of a trait framework. *Applied Psychology: An International Review, 49* (3), 469-482.

Kanfer, R., & Ackerman, P. L. (1994). A self-regulatory skills perspective to reducing cognitive interference. In I. G. Sarason (Ed.), *Cognitive interference: Theories, methods, and findings*. Hillsdale, NJ: Erlbaum.

Kanfer, R. & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74* (4), 657-690.

Kanfer, R. & Heggestad, E. (1997). Individual differences in motivation: Traits and self-regulatory skills. In P. L. Ackerman, P. C. Kyllonen, and R. D. Roberts (Eds.), *Learning and Individual Differences. Process, Trait, and Content Determinants* (pp. 293-309). Washington, D.C.: American Psychological Association.

Kiesler, S., Sproull, L., & Eccles, J. S. (1985). Poll halls, chips, and war games: Women in the culture of computing. *Psychology of Women Quarterly, 9*, 451-462.

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

Krueger, N., & Dickson, P. R. (1994). How believing in ourselves increases risk taking: Perceived self-efficacy and opportunity recognition. *Decision Sciences*, *25*, 385-400.

Kuhl, J., & Baumann, N. (2000). Self regulation and rumination: Negative affect and impaired self-accessibility. In W. J. Perrig and A. Grob (Eds.), *Control of human behavior, mental processes, and consciousness* (pp.283-305). Mahwah, New Jersey: Erlbaum.

Lee, C., & Bobko, P. (1994). Self-efficacy beliefs: Comparison of five measures. *Journal of Applied Psychology, 79* (3), 364-369.

Lehman, R. A. (1999). *How to prepare for the Dental Admission Test*. New York: Barron's.

Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology, 71* (6), 1092-1107.

Lewis, P. B. (1999). Stereotype threat, implicit theories of intelligence, and racial differences in standardized test performance. . (Doctoral dissertation, Kent State University, 1999). *Dissertation Abstracts International, 59* (8-B), 4539.

Leyens, J.P., Desert, M., Croizet, J.C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin*, *26* (10), 1189-1199.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57* (9), 705-717.

Lorak, L. A. (2000, April 14). Male-dominated tech industry launching programs to attract women. South Florida Sun-Sentinel.

Mayer, D. M., & Hanges, P. J. (2003). Understanding stereotype threat effect with 'culture-free' tests: An examination of its mediators and measurement. *Human Performance*, *16* (3), 207-230.

McCullough, W. F. (1997). Differential effects of expected evaluation and task requirements on performance and cognitions. . (Doctoral dissertation, New York University, 1996). *UMI Dissertation Services*, 9702038.

McDonald, Marci. (2004, August 16). A few good women: Tech firms want more female computer whizzes. *U.S. News & World Report*.

McFarland, L. A., Lev-Arey, D. M., & Ziegert. (2003). An examination of stereotype threat in a motivational context. *Human Performance*, *16* (3), 181-205.

McIntyre, R. B., Paulson, R.M., & Lord, C.G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology*, *39*, 83-90.

McKay, P. F., (1999). Stereotype threat and its effect on the cognitive ability test performance of African Americans: The development of a theoretical model. (Doctoral dissertation, The University of Akron, 1999). *Digital Dissertations*, AAT 9925150.

Miller, C. T., & Kaiser, C. R. (2001). A theoretical perspective on coping with stigma. *Journal of Social Issues*, *57* (1), 73-92.

Miller, J. K. (1992, October 20). Pressure adds up: Mattel to change talking Barbie doll. The Atlanta Journal and Constitution, D2.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system of theory and personality: Re-conceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246-268.

Mitchell, T.R., Hopper, H., Daniels, D., & George-Falvy, J. (1994). Predicting self-efficacy and performance during skill acquisition, *Journal of Applied Psychology*, *79* (4), 506-517.

Margolis, J., & Fisher, A. (1997) Geek Mythology and attracting undergraduate women to computer science. *Impacting Change Through Collaboration*, proceedings of the Joint National Conference of the Women in Engineering Program Advocates Network and the National Association of Minority Engineering Program Administrators.

Margolis, J., Fisher, A., & Miller, F. (2000). The anatomy of interest: Women in undergraduate computer science. *Women's Studies Quarterly*, Spring/Summer.

Margolis, J., Fisher, A., & Miller, F. (1997). Caring about connections: Sex and computing. *IEEE Technology & Society*, December.

Margolis, J., Fisher, A., Miller, F. (2000). Anatomy of interest: Women in undergraduate Computer Science. *Women's Studies Quarterly*, 104-126.

Margolis, J., Fisher, A., Miller, F. (2000) Caring About connections: Sex and computing. *Technology and Society*, 13-20.

Margolis, J. & Fisher, A. (2000). Women in Computer Sciences: Closing the Sex Gap in Higher Education. Final Report to the Alfred B. Sloan Foundation.

Maurer, T. J., & Andrews, K. (2000). Traditional, Likert, and simplified measures of self-efficacy. *Educational and Psychological Measurement*, *60* (6), 965-973.

McGregor, H.A., & Elliot, A.J. (2002). Achievement goals as predictors of achievement-related processes prior to task engagement. *Journal of Education Psychology*, *94* (2), 381-395.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Re-conceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246-268.

Mohammed, S., & Billings, R. (2002). The effect of self-efficacy and issue categorization on threat and opportunity categorization. *Journal of Applied Psychology, 32*, (6), 1253-1275.

Mulkey, J.R., & O'Neill, H.F. (1999). *Computers in Human Behavior*, *15*, 494-509.

National Council for Research on Women. *Balancing the Equation: Where Are Women and Girls in Science, Engineering, and Technology?*

Nguyen, H. D., O'Neal, A., & Ryan, A. M. (2003). Relating test-taking attitudes and skills to stereotype threat effects to the racial gap in cognitive ability test performance. *Human Performance*, *16* (3), 261-293.

Niemann, Y. F., O'Connor, E., & McClorie, R. (1998). Intergroup stereotypes of working class blacks and whites: Implications for stereotype threat. *Western Journal of Black Studies*, *22* (2), 103-108.

Osborn, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, *26* (3), 291-310.

Oswald, D. L., & Harvey, R. D. (2000). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology: Developmental, Learning, Personality, Social*, *19* (4), 338-356.

Ployhart, R. E., Ziegert, J. C., & McFarland, L. A. (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactions research. *Human Performance, 16* (3), 231-259.

Quinn, D. M., & Spencer, S. J. (1996). *Stereotype threat and the effect of test diagnosticity on women's math performance*. Paper presented at the annual American Psychological Association Conference, Toronto.

Richards, J. M., & Gross, J. J. (2000). Emotion regulation and memory: The cognitive costs of keeping one's cool. *Journal of Personality and Social Psychology*, *79* (3), 410-424.

Roberson, L., Deitch, E. A, Brief, A. P., Block, C. J. (2003). Stereotype threat and feedback seeking in the workplace. *Journal of Vocational Behavior*, *62*, 176-188.

Robinson-Staveley, K., & Cooper, J. (1990). Mere presence, sex, and reactions to computers: Studying human-interaction in the social context. *Journal of Experimental Social Psychology*, *26*, 168-183.

Sackett, P. R. (2003). Stereotype threat in applied selection settings: A commentary. *Human Performance*, *16* (3), 295-309.

Sackett, P. R., Schmitt, N., Kabin, M. B., & Ellingson, J. E. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist*, *56* (4), 302-318.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in pre-employment testing. *American Psychologist*, *49*, 929-954.

Sanna, L. J., & Shotland, R. L. (1990). Valence of anticipated evaluation and social facilitation. *Journal of Experimental Social Psychology*, *26*, 82-92.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology*, *46*, 929-938.

Sarason, I. G., Sarason, B. R., Keefe, D. E., Hayes, B. E., & Shearin, E. N. (1986). Cognitive Interference: Situational determinants and traitlike characteristics. *Journal of Personality and Social Psychology, 51* (1), 215-226.

Schmader, T. (2002). Sex identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38* (2), 194-201.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. J*ournal of Personality and Social Psychology, 85* (3), 440-452.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Seta, C. E., Seta, J. J., & Wang, M. A. (1988). The effects of evaluation on organizational processing. *Personality and Social Psychology Bulletin, 14* (3), 604-609.

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

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

Spielberger, C. D. (1985). Assessment of *state* and *trait anxiety*: Conceptual and methodological issues. *Southern Psychologist, 2* (4), 6-16.

Stangor, Carr, and Kiang, (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75* (5), 1191-1197.

Steele, C. M. (1999). Thin ice: Stereotype threat and black college students. *The Atlantic Monthly, 284* (2).

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual ability and performance. *American Psychologist, 52* (6), 613-629.

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

Steele, C. M., & Davies, P. G. (2003). Stereotype threat and employment testing: A commentary. *Human Performance, 16* (3), 311-326.

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* (6), 1213-1227.

Stricker, L. J. (1998). Inquiring about examinees' ethnicity and sex: Effects on AP Calculus AB Examination Performance. College Board Report No. 98-1

Stricker, L.J., & Bejar, I. I. (1999). Test difficulty and stereotype threat on the GRE general test. *GRE Research Report 96-06R*. Princeton: Educational Testing Service.

Stricker, L. J., & Ward, W. C. (1998). Inquiring about examinees' ethnicity and sex: Effects on computerized placement test performance. College Board Report No. 98-2

Tajfel, H., & Turner, J. C. (1985). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *The social psychology of intergroup relations*. Monterey, CA: Brooks/Cole. 33-47.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, *4* (4), 525 – 547.

Trafimow, D., Triandis, H. C., & Goto, Sharon C. (1991). Some tests of distinction between the private and the collective self. *Journal of Personality and Social Psychology*, *60* (5), 649-655.

Wah, D. M., & Robsinson, D. S. (1990). *Examinee score trends for the GRE general test: 1977-1978, 1982-1983, 1986-1987, and 1987-1988*. Princeton: Educational Testing Service.

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on sex differences in mathematical problem solving. *Sex Ro*les, *41* (3/4), 219-240.

Whaley, A. L. (1998). Issues of validity in empirical tests of stereotype threat theory. *American Psychologist*, *53*, 679-680.

Wheeler, S. C., Jarvis, W. B. G., & Petty, R. E. (2001). Think unto others: The self-destructive impact of negative racial stereotypes. *Journal of Experimental Social Psychology*, *37*, 173-180.

Yee, D. K., & Eccles (1988). Parent perceptions and attributions for children's math achievement, *Sex Roles*, *19*, 317-333.