EDUCATION & PSYCHOLOGY I+D+i
Research, Innovation and Solutions on-line®
Electronic Journal of Research in Educational Psychology
ejrep
Editorial EOS

Reducing the Impact of Stereotype Threat on Women's Math Performance: Are Two Strategies Better Than One?

Paul R. Jones

Pacific Institute for Research and Evaluation (PIRE), Calverton, MD

USA

Correspondence: Paul R. Jones. Pacific Institute for Research and Evaluation (PIRE), 11720 Beltsville Drive, Suite 900, Calverton, Maryland 20705-3102. USA. E-mail: jones@pire.org

Abstract

Introduction. Two studies examined whether stereotype threat impairs women's math performance and whether concurrent threat reduction strategies can be used to offset this effect.

Method. In Study 1, collegiate men and women (N = 100) watched a video purporting that males and females performed equally well (*gender-fair*) or males outperformed females (*gender differences*) on an imminent math test. In Study 2, (N = 44) women viewed the gender differences video, followed by misattribution (cue *present*, *absent*) and self-affirmation (*present*, *absent*) manipulations, before taking the aforesaid test.

Results. In the initial study, women underperformed men on the test after receiving the gender differences video, whereas no gender differences emerged in the gender-fair condition. In Study 2, affirming the self led to better performance than not doing so. Planned contrasts indicated, however, that only women receiving a misattribution cue and self-affirmation opportunity outperformed their counterparts not given these reduction strategies.

Discussion. These findings are discussed relative to Stereotype Threat Theory and educational implications are provided.

***Keywords*:** *stereotype threat*, *performance*, *self-affirmation*, *misattribution*

Received: 03/07/11 *Initial acceptance: 03/15/11* *Final acceptance: 09/07/11*

Reducción del impacto de la amenaza de estereotipo en el rendimiento de las mujeres en las matemáticas. ¿Valen más dos estrategias que una?

Resumen

Introducción. Se llevaron a cabo dos estudios para investigar si la amenaza de estereotipo influye en el rendimiento que las alumnas alcanzan en matemáticas, y si las estrategias simultáneas de reducción de amenaza pueden llegar a compensar ese efecto.

Método. En el Estudio 1, estudiantes universitarios de ambos sexos (N = 100) observaron un video que mostraba a hombres mujeres obteniendo iguales resultados (*igualdad de género)* o a los hombres superando a las mujeres (*diferencias de género*) en una prueba de matemáticas. En el Estudio 2, (N = 44) mujeres vieron el video de las diferencias de género, seguido por manipulaciones de falsa atribución (clave *presente*, *ausente*) y auto-afirmación (*presente, ausente*), antes de tomar la prueba.

Resultados En el Estudio 1 las mujeres, después de ver el vídeo de las diferencias de género, obtuvieron peores resultados en la prueba que los hombres; mientras que bajo las condiciones de igualdad de género no se observe ninguna diferencia de género. En el Estudio 2, la auto-afirmación incrementó el rendimineto en la prueba. Sin embargo, la realizacion de contrastes pre-establecidos indicaron que sólo las mujeres que recibieron manipulaciones de falsa atribución y oportunidades de auto-afirmación superaron a sus contrapartes que no recibieron estas estrategias de reducción.

Discusión. Estos resultados se discuten en relación a la teoría de la Amenaza de Estereotipo y se proporcionan las implicaciones educativas.

Palabras Clave: *amenaza de estereotipo, rendimiento, auto-afirmación, falsa atribución*

Recibido: 07/03/11 *Aceptación inicial: 15/03/11* *Aceptación final: 09/07/11*

"Had anyone told me 20 years ago that I would ever say that statistics are fun, I would have laughed out loud. I was sure that I was 'not good at math' and dealt with that self-perception by only taking whatever math was unavoidable, and holding my nose."

–Nancy Dess, Ph.D., Senior Scientist, APA Science Directorate (Dess, 2001)

Introduction

Nancy Dess's remarks about her math ability and reluctance to engage in mathematical pursuits convey the female math inferiority stereotype—that is, *women do not perform as well as men in math*. Studies have shown that the top male students outperform their female counterparts in the quantitative sciences and on many college and graduate entrance exams (Callahan, 1991; Stumpf & Stanley, 1998). For instance, the scores of women on the quantitative sections of the Graduate Record Examinations (GRE) lag behind those of men by 20 to 90 points (Educational Testing Service, 2002; Halpern, 1989). Although the predictive validity and utility of these exams have been debated (Callahan, 1991; Neisser, 1998; Sackett, Schmitt, Ellingson, & Kabin, 2001), there is little doubt that they play a pivotal role in access to educational resources and admission to selective institutions. However, what phenomena might account for these gender differences and how might such effects be minimized?

Stereotype Threat Theory and its Generalizability

Some researchers stress the importance of genetic factors in the underperformance of women in math (Benbow, 1988; Benbow & Minor, 1986; Benbow & Stanley, 1980). An alternative explanation, however, is offered by *Stereotype Threat Theory* (STT; Steele, 1992; 1997). According to STT, an individual may experience apprehension about the possibility of validating a stereotype that exists for his or her social group (Steele, 1997; Steele & Aronson, 1998). This situational predicament, known as *stereotype threat*, is conceptually defined as "... the discomfort targets feel when they are at risk of fulfilling a negative stereotype about their group; the apprehension that they could behave in such a way as to confirm the stereotype—in the eyes of others, in their own eyes, or both at the same time" (Aronson, Quinn, & Spencer, 1998, p. 86). Additionally, STT posits that this phenomenon will have a stronger impact on those who are most invested in a domain—that is, those with a high level of *domain identification*—and prolonged exposure to threat can lead to *disidentification*—the process whereby a threatened individual chronically disengages from tasks in this area (Steele, 1997).

In a set of seminal experiments, Steele and Aronson (1995) found that (after adjusting for prior performance) African-American undergraduates performed worse on a high-stakes exam than their Anglo-American counterparts when it was described as diagnostic of ability (high-threat). No racial differences surfaced, however, when the task was described as non-diagnostic (low-threat). Similar effects have been shown among other stigmatized groups (e.g., Latinos; Gonzales, Blanton, & Williams, 2002), in published and unpublished work, and across studies, manipulations, and tasks (for reviews, see Nguyen & Ryan, 2008; Wheeler & Petty, 2001). Importantly, this classic pattern of results has been replicated among women showing that, when compared to men, their math performance is hampered under threatening conditions (Ben-Zeev, Fein, & Inzlicht, 2005, Study 2; Brown & Josephs, 1999; Quinn & Spencer, 2001; Spencer, Steele, & Quinn, 1999).[1] Gender differences do not occur, however, in nonthreatening contexts.

Mediation and Reduction of Stereotype Threat

What factors might mediate or moderate threat effects? Wheeler and Petty's (2001) meta-analysis identified 18 potential mediators of stereotype threat effects on performance ranging from task confidence perceptions to distracting thoughts. Unfortunately, many of these mechanisms were not tested by formal mediational analysis. Even more recent meta-analytic reviews (Nguyen & Ryan, 2008) have not resolved this issue. In instances where the appropriate statistical rigor has been applied, results have often been mixed or null (Jones & Stangor, 2003; Smith, 2004).

State anxiety has received considerable attention as a potential mediator in the threat literature (Smith, 2004). Though early threat research uncovered modest direct support for this mechanism (e.g., Spencer et al., 1999; Study 3),[2] more recent evidence indicates that, among women, the activation of stereotype threat is associated with increased physiological arousal (Ben-Zeev et al., 2005; O'Brien & Crandall, 2003), anxiety (Spencer et al., 1999), and decreased cognitive capacity (Beilock, Rydell, & McConnell, 2009; Johns, Inzlicht, & Schmader, 2008) that, in turn, leads to academic underperformance (Rydell, McConnell, & Beilock, 2009). In sum, it appears the mediation of stereotype threat among women in the academic domain is multivariate (Mayer & Hanges, 2003; Smith, 2004; Steele & Aronson, 1995) and involves cognitive and physiological mechanisms.

Thus far, the literature has focused heavily on the factors necessary to induce threat, individual differences associated with this phenomenon, and the contexts in which threat effects occur. For instance, studies have tested how domain identification (Aronson et al., 1999; Stone, Lynch, Sjomeling, & Darley, 1999) and personal theories of intelligence (Aronson, Fried, & Good, 2002) moderate stereotype threat effects. Other research has examined the role of task characteristics (e.g., task difficulty; Spencer & Steele, 1992) in evoking threat. Disproportionately few studies, however, have investigated the conditions necessary to reduce (or remove) stereotype threat in testing contexts.

Misattribution and Self-Affirmation as Reduction Strategies

To date, a host of strategies have been offered as ways to reduce stereotype threat including rendering the belief incorrect or irrelevant (Spencer et al., 1999), redefining the situation as nonthreatening (Steele & Aronson, 1995) and teaching stigmatized individuals about this phenomenon (Johns, Schmader, & Martens, 2005). Two of the more popular reduction strategies involve misattribution (Ben-Zeev et al., 2005; Brown & Josephs, 1999; O'Brien & Crandall, 2003; Stone et al., 1999) and self-affirmation processes (Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009; Martens, Johns, Greenberg, & Schimel, 2006). [3]

For instance, Ben-Zeev et al. (Ben-Zeev et al., 2005; Study 2) assigned highly math-identified women to complete a difficult math exam in the presence of other women (*same-sex* condition) or men (*minority status* condition). Participants were placed in front of an ostensible "subliminal noise generator" and told the device emitted a tone that either was (*misattribution* condition) or was not (*control* condition) associated with physiological side-effects. As expected, women in the minority status condition performed worse than women in the same-sex condition when no misattribution cue was provided. However, there were no performance differences when the cue information was presented.

Prior research also shows that self-affirmation processes can reduce the impact of threat in academic contexts (Aronson & Damiani, 1997 as cited in Aronson, Cohen, & Nail, 1999; Cohen et al., 2009). For example, Martens et al. (2006) found that, although men were unaffected by the experimental manipulation, women did worse on an exam assumed to gauge math ability when compared to women under nondiagnostic conditions. Importantly, women

who were threatened, but were allowed to affirm the self, performed on par with women in the nondiagnostic condition.

When taken together, these studies highlight the utility of the aforesaid reduction strategies. Additionally, they underscore the role of cognitive and physiological mechanisms in stereotype threat effects.

Shortcomings of the Stereotype Threat Literature

Despite its counterintuitive appeal and interest from scholars, policy makers, and the media (Ad Council & Girl Scouts of the USA, 2004; Barnard, Burley, Olivarez, & Crooks, 2008, Chandler, 1999; Gladwell, 2008; McFarland, Lev-Arey, & Ziegert, 2003; Sackett, Hardison, & Cullen, 2004; Shih, Pittinsky, & Ambady, 1999; Stangor, 1999), several theoretical and empirical issues surrounding stereotype threat require clarification. First, the issue of which underlying mechanisms account for this phenomenon remains unclear (Steele & Aronson, 1995). The lack of formal meditational tests and methodological concerns associated with the measurement of multiple mediators in a given study (e.g., heightened demand characteristics) has undoubtedly contributed to this state of affairs (Stroessner & Good, 2010). Second, despite evidence indicating that misattribution and self-affirmation are viable moderators of stereotype threat effects, prior research has not always lent itself to straightforward accounts of the study results (Stone et al., 1999, Study 2). Third, the question, "Is there benefit to employing more than one threat-reduction strategy (concurrently) in a single study?" (i.e., misattribution and self-affirmation) has not, to my knowledge, been addressed in the published stereotype threat literature.

Goals of the Present Research

The studies reported herein were part of a larger research program designed to investigate the aforesaid shortcomings (Jones, 2005). Its goals were threefold. First, I intended to replicate the standard stereotype threat effect on women's math performance in Study 1 (Spencer et al., 1999). Second, I sought to understand the effects (if any) of combining two reduction strategies on the performance of stereotype threatened women in Study 2. Finally, I assessed whether selected cognitive (e.g., task confidence) and psychophysiological (e.g., anxiety) mediators underlie the relation between threat and performance in Studies 1 and 2.

To minimize article length without compromising substance, I only report on key findings as results on other variables are presented elsewhere (see Jones, 2005).

Study 1

In Study 1, women and men were asked to complete a math test after being told that the task had produced *gender differences* or was *gender-fair*. Participants completed measures of perceived stereotype threat before, during, and after the test. Consistent with STT, I hypothesized that women in the gender differences condition would perform worse on the task and experience greater levels of stereotype threat (e.g., increased anxiety) than men, whereas no between-groups differences in performance or perceived threat were expected in the gender-fair condition.

Method

Design and Participants

This quasi-experiment took the form of a 2 (gender: *male, female*) X 2 (instructional set: *gender differences*, *gender-fair*) factorial design with math performance as the primary dependent measure. Prior research shows that domain identification moderates stereotype threat effects (Aronson et al., 1999), therefore, all participants indicated how important math pursuits are to them using the 9-item math subscale of the *Domain Identification Measure* (DIM; Smith & White, 2001), their scores on the verbal and quantitative SAT (VSAT; QSAT), and their high school and college GPA (HS-GPA; C-GPA) on a survey administered weeks before the experiment. Items on the DIM were scored on a 5-point scale (1 = *strongly disagree* to 5= *strongly agree*; e.g., *Mathematics is one of my best subjects.*).

One hundred University of Maryland undergraduates participated in this experiment in exchange for course credit. The data from 7 participants (6 females and 1 male) were excluded because they failed the treatment check. This left 93 participants (63 females and 30 males) who were randomly assigned to one of two experimental groups. The mean math DIM score (α = .91) for the sample was 3.0, while the average VSAT, QSAT, C-GPA, and HS-GPA were 596, 618, 3.3, and 3.7, respectively. Given the controversy surrounding the use of QSAT scores as a covariate in ANCOVA designs (see Sackett et al., 2004; Steele & Aronson, 2004; Wicherts, 2004), I did not use this variable as a covariate in the analyses reported herein.

Procedure

Several weeks after the initial survey, participants reported to the lab where they were met by a male experimenter who told them that they would take part in a study designed to assess "the psychology of problem solving." [4] The experimenter explained that all participants would complete a set of timed tasks on a computer. Participants were urged to follow all directions and pay close attention to any video content because they would be asked to recall this information later in the experiment.

Each participant was led to a room equipped with a computer, speakers, and a folder labeled "Task 1." Participants were then seated at the computer, given a watch, pencil, and scrap paper, and told that the computer would tell them if they needed to refer to the folder. The experimenter asked if there were any questions, entered a randomly assigned code number to start the program, and exited the room.

Stereotype Threat Manipulation

After starting the program, participants were randomly assigned to one of two instructional sets embedded in a video. Similar to Spencer et al. (1999), participants in the gender differences (threat) condition watched a video depicting a European-American male, named "Patrick Smith," who was presumably a researcher from the Psychology Department. The character informed participants that he was studying why there are gender differences on standardized exams and trying to understand why males outperform females on the upcoming task. Participants in the gender-fair (no threat) condition viewed the same character, but were told that his research was a collaborative project with the Women's Studies Department and neighboring universities to develop a gender-fair test. He further stated that his findings had shown that males and females performed equally well on the upcoming task.

One of two charts was embedded in the video for 10 seconds to reinforce the gender performance information. The chart in the gender differences condition depicted the performance of men as higher than that of women, whereas the chart in gender-fair condition depicted the performance of men and women as equal. The video concluded by telling participants that they would be completing a math task.

Pretest Measures

After the video, participants were informed that they would answer several questions before the math task including the following:

Motivation and expectancy. Since stereotype threat is not assumed to lower task motivation among those high in domain identification, we asked participants to indicate whether they were motivated to do well on the math task and estimate how well they would perform on it. Both single-item measures were scored on a 7-point scale (1 = *not at all motivated; very poorly* to 7 = *extremely motivated; very well*).

Performance Measures

Math performance. Participants were then given 15 minutes to complete a 10-item math test. All questions were taken from a Graduate Record Examinations (GRE) study guide (Educational Testing Service, 1994) and geometry-based as prior studies show that these items often present the most difficulty for women because sound spatial skills are needed (Liben, 1978; Stangor & Sechrist, 1998). Only difficult items were selected (i.e., ≤ 50% of the testing population completed each question correctly) to remain consistent with earlier stereotype threat research (Spencer et al., 1992; Steele et al., 1995).

Reaction time and perceived task confidence (per item). After responding to each item, reaction time was recorded in milliseconds (ms) and participants provided task confidence ratings on a 9-point scale (1 = *not at all confident* to 9 = *extremely confident*).

Posttest Measures

State anxiety and demographic information. After the math task, participants completed the STAI (state version; Spielberger, 1972) and provided demographic information. The former 20-item inventory (α = .95) uses a 4-point scale (1 = *not at all* to 4 = *very much so*; e.g., *I feel anxious*.) with higher mean scores indicating greater momentary anxiety.

Treatment check. To verify the effectiveness of the threat manipulation, participants were asked to recall what they were told in the video regarding the math test (i.e., *non-gender biased*; *found gender differences*; *no such information was given to me*; *I do not remember*).

Results and Discussion

Treatment check. To test the experimental manipulation's effectiveness, participants were asked to recall how men and women had performed on the math task. Seven participants answered this item incorrectly—one in the *gender differences* condition and six in the *gender-fair* condition—and were excluded from the analysis. A chi-square (χ^2) test showed that the remaining participants were more likely to correctly recall that the test was gender neutral in the gender-fair condition (85%) and gender-biased in the gender differences condition (83%), when compared to participants that did not recall this information (15% and 17%, respectively) in the aforesaid conditions, $\chi^2(3, n = 92) = 77.78$, $p < .01$.

Math Performance

I analyzed the performance and mediator data using a series of 2-way ANOVAs (gender X instructional set). Although three participants—two males and one female—failed to complete the math test in the allotted time, these data were used in the analysis as their inclusion did not change magnitude or direction of the results. Given that 73% of the participants answered item 8 incorrectly—a higher percentage than any other item—it was excluded in the tabulation of performance scores. Table 1 shows that the results were in line with my hypotheses: A marginally significant gender main effect emerged on the number of items correct (across the remaining nine items, $M_{\%\ \text{incorrect}} = 52\%$), F (1, 89) = 3.48, $p = .065$, that was qualified by a significant gender X instructional set interaction, F (1, 89) = 5.62, $p = .02$. [5]

Table 1. ***Mean Number of Items Correct as a Function of Gender and Instructional Set***

Gender of the Participant	Instructional Set					
	Gender Differences			Gender-Fair		
	M	SD	N	M	SD	N
Female	3.93_b	2.09	33	4.47	2.37	30
Male	6.29_a	3.00	14	4.19_b	2.17	16

Note. Means not sharing a common subscript differ at $p < .05$. Higher means indicate better performance.

As predicted, women ($M = 3.93$, $SD = 2.09$) performed significantly worse than men ($M = 6.29$, $SD = 3.00$) in the gender differences condition, $t(45) = 3.10$, $p < .01$, $d = -0.95$; however, no differences emerged between women ($M = 4.47$, $SD = 2.37$) and men ($M = 4.19$,

SD = 2.17) in the gender-fair condition, $t(44) = 0.39$, $p = .70$. Men in the gender differences condition also performed significantly better than did men in the gender-fair condition, $t(28) = 2.22$, $p = .04$, $d = 0.84$. This finding is consistent with the *stereotype lift* effect whereby men perform significantly better under threat conditions when compared to non-threatened men. No significant differences emerged, however, between women in the gender differences and gender-fair conditions, $t(61) = 0.96$, $p = .34$.

Mean reaction time (per item). Scores on this measure were log(10) transformed to correct for skew and I deleted the data from three participants who exceeded the time limit. The analysis revealed only a significant main effect of gender on mean response time scores (per item), F (1, 86) = 4.33, $p = .04$. Specifically, women ($M = 4.43$, $SD = 0.24$) completed each item significantly faster (on average) than men ($M = 4.52$, $SD = 0.11$). No other significant effects were found (all p's > .38).

Potential Mediators

State anxiety, motivation, and expectancy. Separate analyses on these measures yielded only significant gender main effects. Women ($M = 4.92$, $SD = 1.05$) had significantly lower task expectations than men ($M = 5.90$, $SD = 0.89$; $F(1, 89) = 18.71$, $p < .01$) and significantly higher levels of anxiety than men ($M = 2.40$, $SD = 0.67$ vs. $M = 1.88$, $SD = 0.53$; $F(1, 89) = 13.80$, $p < .01$). No other significant effects emerged (all p's > .28). As expected, these data indicate that sheer motivation does not account for the effects described herein.

Task confidence (per item). Analyses on the average task confidence ratings (per item) revealed a significant gender main effect, $F(1, 89) = 8.84$, $p < .01$, that was subsumed by a significant gender X instructional set interaction, $F(1, 89) = 4.77$, $p = .05$ (Table 2).

Table 2. ***Mean Task Confidence as a Function of Gender and Instructional Set***

Gender of the Participant	Instructional Set					
	Gender Differences			Gender-Fair		
	M	SD	N	M	SD	N
Female	5.26_b	1.26	33	5.43	1.51	30
Male	6.80_a	0.95	14	5.67_b	1.44	16

Note. Means not sharing a common subscript differ at $p < .05$. Higher means indicate greater task confidence.

Simple-effects tests showed that women (M = 5.26, SD = 1.26) were significantly less confident in their answers than were men (M = 6.80, SD = 0.95) in the gender differences condition, $t(45) = 4.10$, $p < .01$. However, the confidence ratings of women (M = 5.43, SD = 1.51) and men (M = 5.67, SD = 1.44) did not differ in the gender-fair condition, $t(44) = 0.52$, $p = .60$. Though the confidence ratings of men in gender differences condition were significantly higher than those of men in the gender-fair condition, $t(28) = 2.22$, $p = .04$, no significant differences were found between the ratings of women in the gender differences and gender-fair conditions, $t(61) = 0.49$, $p = .63$. No other effects approached significance (all p's > .11).

Mediational Analysis

To test whether the relation between the gender X instructional set interaction and math performance was mediated by task confidence perceptions, I used the four-step approach outlined by Baron and Kenny (1986). All four steps were achieved (see Figure 1) as the interaction effect significantly predicted math performance and was correlated with participants' confidence scores. When confidence scores were held constant, their effect on math performance was significant, while the previously significant interaction effect on math performance was no longer statistically reliable, $\beta = 0.28$, $t(92) = 1.17$, $p = .25$. A Sobel test (Kenny, Kashy, & Bolger, 1998) revealed a significant difference in the direct path from the interaction effect to math performance after controlling for participants' confidence scores, $t(67) = 2.19$, $p = .03$. Thus, participants' confidence perceptions explained a significant proportion of the variance in the relationship between the interaction effect and math performance.

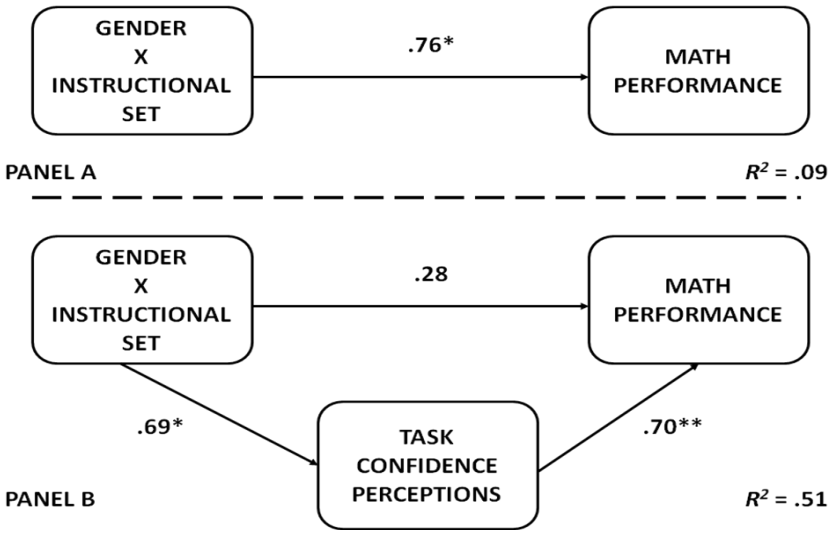

Figure 1. ***Mediation of the gender X instructional set interaction effect on task performance by task confidence perceptions.***

Note. The significant direct path from the interaction to task performance was reduced to nonsignificance when the effect of the mediator was statistically controlled. R^2 values reflect the proportion of variance explained by each model.
* = $p < .05$, ** = $p < .01$.

Further inspection of this model revealed that the confidence perceptions and performance of men were its driving forces. For instance, among men, I found a marginally significant instructional set--math performance relationship, $\beta = 0.34$, $t(29) = 1.92$, $p = .065$, and a significant correlation between instructional set and confidence perceptions, $\beta = 0.43$, $t(29) = 2.50$, $p = .02$. When confidence scores were held constant, their effect on math performance remained significant, $\beta = 0.81$, $t(29) = 6.36$, $p < .01$, while the marginally significant interaction effect on performance was no longer statistically reliable, $\beta = -0.002$, $t(29) = -0.16$, $p = .99$. There was no empirical support for confidence perceptions as a mediator of the threat--performance relationship among women.

Given that confidence perceptions were measured after each item, it is possible that the stereotype threat manipulation created gender differences in math performance, which in turn influenced participants' confidence scores. I examined this alternative with confidence scores as the dependent variable. Although unexpected, all four steps recommended by Baron et al. (1986) were met. In other words, the interaction effect significantly predicted confidence scores and the proposed mediator (i.e., math performance). When performance was held constant, these scores remained correlated with participants' confidence ratings, whereas the pre-

viously significant relation between the interaction effect and confidence scores was no longer reliable, $\beta = 0.18$, $t(92) = 0.75$, $p = .46$. Once again, a Sobel test revealed that, after controlling for math performance, there was a significant difference in the direct path from the interaction effect to confidence perceptions, $t(92) = 2.29$, $p = .02$. Therefore, this alternative model cannot be ruled out as a plausible account of these data.

When taken together, these findings replicate those of Spencer et al. (1999). There is an important distinction, however, between the present findings and the aforesaid work. The driving force behind my results was the performance of men, whereas the performance of women was responsible for the gender differences produced in Spencer et al.'s (1999) research.

Study 1 also illuminated the mechanisms that may mediate stereotype threat effects. For instance, Study 1 suggests that motivation, anxiety, and expectations did not underlie the effects found therein. Although task confidence perceptions mediated the relation between stereotype threat and performance, two findings make this straightforward account less clear. First, I found that a plausible alternative model fit these data equally well (i.e., performance as a mediator of the stereotype threat--task confidence relationship). Second, both models were heavily driven by the confidence perceptions and performance of men. Though this finding sheds light on the potential underlying mechanisms of *stereotype lift* and the role of cognitive factors in such processes, it provides little insight into what mediates stereotype threat among women. In short, this study leaves the question, "Which meditational model truly accounts for threat effects?" unresolved.

From the outset, I indicated that there is a burgeoning literature demonstrating basic stereotype threat effects (Nguyen & Ryan, 2008; Walton & Spencer, 2009), but comparatively less research addressing ways to reduce (or remove) this phenomenon in the situational context. Research addressing the efficacy (if any) of using multiple threat reduction strategies concurrently is less common. To explore these possibilities, I conducted a second study.

Study 2

In Study 1, taking an exam under threatening conditions impaired the math performance of women when compared to men. Could this effect be reduced (or removed) by self-affirmation, misattribution, or both processes? In addition, would combining these strategies

produce an added benefit to women's math performance or inhibit the potency of each mechanism? To examine these questions, I presented a sample of math-able women with the gender differences video described in Study 1. Additionally, participants were either given a misattribution cue, an opportunity to affirm the self, both of these strategies (misattribution + self-affirmation), or neither of these options (gender differences only) before completing a math test in Study 2. The presence of a dual strategies condition allowed me to test whether multiple threat reduction techniques produce effects that add to or inhibit the strength of each strategy. For instance, concurrent reduction strategies may safeguard women from threat in a way that produces a performance gain or have the unintended consequence of reducing the potency of each mechanism (e.g., by preventing participants from devoting sufficient attention to either strategy). [6]

In sum, I predicted that self-affirmation and misattribution processes would independently and concomitantly reduce the effects of stereotype threat on women's math performance. I expected performance differences to emerge between women given the gender differences instructions without a threat reduction strategy and participants in each of the remaining conditions, such that allowing women to self-affirm, misattribute arousal, or simultaneously engage in these strategies would lead to *better* performance than not doing so.

Method

Design and Participants

Study 2 took the form of a 2 (misattribution cue: *present, absent*) X 2 (self-affirmation opportunity: *present, absent*) between-participants factorial design with all recruits receiving the gender differences instructions and math test described in Study 1. I recruited 44 female undergraduates from the University of Maryland to participate in this experiment in exchange for course credit. The data from two participants (in the *misattribution cue absent*/*self-affirmation present* and *misattribution cue present*/*self-affirmation present* conditions) were excluded from the analysis due to response patterns consistent with not taking the math test seriously. This left 42 women who were randomly assigned to four experimental conditions. The mean VSAT, QSAT, C-GPA, HS-GPAs, and math DIM scores (α = .89) were 583, 607, 3.3, 3.7, and 3.2, respectively.

Procedure. The procedure and instructions for Study 2 were almost identical to those of Study 1, except for (1) how the experimental manipulations were introduced and (2) the manipulations checks used.

Malfunctioning Computer Cover Story

Upon arrival to the lab, participants were met by a male experimenter who provided the cover story described in Study 1. Each participant was led to a room containing two computers approximately 1 ft (0.3048 m) apart. For half of the participants (misattribution cue present condition) the right-most computer was rigged to appear as a malfunctioning machine (i.e., displaying a looped static image), while the left-most computer was in perfect working order with two envelopes labeled 'Task 1' and 'Task 2' placed next to it. Participants were seated at the left-most computer and told that the lab had been experiencing networking problems on some of its machines. They were further informed that a technician would be servicing the affected machines later in the day and, per his recommendation, everyone was instructed not to touch the malfunctioning computer.

Despite this shortcoming, participants were told that they were expected to complete all tasks on the functional computer; however, they should be aware that several participants indicated that the malfunctioning computer made them feel "anxious" while completing the tasks. The remaining half of the participants (misattribution cue absent condition) completed the experiment in the same context save the malfunctioning computer and anxiety side effect ploys. For these participants, the right-most computer remained but was turned off.

Self-Affirmation Manipulation

After the misattribution cue manipulation, the experimenter entered a random code number into the left-most computer, told participants to follow the on-screen directions, and exited the room. The computer then presented all participants with the gender differences (or threat inducing) video described in Study 1. They were further instructed to open the 'Task 1' folder and to complete its contents. Half of the participants (*self-affirmation present* condition) read the following excerpt and completed a free-format measure that asked them to write about their GPA:

"The average college student in America has a GPA of 3.0 on a 4.0 scale and a SAT score of 1,000. With that in mind, we'd like you to spend a few moments writing a brief paragraph about your GPA..." [7]

The remaining half of the participants (*self-affirmation absent* condition) opened a 'Task 1' envelope that directed them to open the 'Task 2' folder. For all participants, the latter envelope contained pretest items identical to those of Study 1. Each participant completed these items followed by the math test and posttest items described in the previous study.

Misattribution and self-affirmation treatment checks. Two items at the end of the experiment examined the effectiveness of the misattribution and self-affirmation manipulations. All participants were asked to recall what they were told about the feelings of others toward the malfunctioning computer (i.e., *It made them feel anxious*; *No such information was given to me*; *I do not remember*). Additionally, participants indicated the nature of the initial task (i.e., *I was asked to write about my GPA*; *I did not have to complete Task 1*; *I do not remember*). After completing these items, participants were probed for suspicion, debriefed, and dismissed.

Results and Discussion

Treatment checks. After removing one participant in the misattribution cue absent condition who failed the treatment check, participants' scores on the misattribution recall measure were subjected to a chi-square (χ^2) analysis. As expected, participants had little difficulty in remembering what they were told about the feelings of others toward the malfunctioning computer, $\chi^2(2, n = 42) = 27.90$, $p < .01$. Nearly all of the participants in the misattribution cue present and misattribution cue absent conditions (86% and 95%, respectively) correctly recalled this information. Similarly, all participants across self-affirmation conditions were able to correctly recall whether (or not) they were given a chance to affirm the self, $\chi^2(1, n = 42) = 42.00$, $p < .01$.

Math Performance

Ninety-eight percent (98%) of the participants were able to affirm the self based on their GPA, SAT score, or both indices. One participant, however, reported a GPA less than 3.5 and a SAT score below 1,000. Removal of this participant from the analysis did not change the direction of the results; therefore, this information was retained. Once again, I analyzed these data using a series of 2-way ANOVAs (misattribution cue X self-affirmation op-

portunity). Tests on participants' math scores and reaction times (per item) only revealed a significant self-affirmation opportunity main effect on the former measure, F (1, 38) = 4.29, p < .05. Namely, women in the self-affirmation present condition (M = 5.35, SD = 1.93) performed significantly better than did their self-affirmation absent condition counterparts (M = 4.14, SD = 1.86; d = 0.64). No other significant main effects or interactions emerged (all p's > .13). In keeping with my hypothesis, however, I conducted focused comparisons (t-tests) to test whether women in the misattribution cue absent/self-affirmation absent condition would perform worse than women in each of the remaining conditions. Indeed, these planned contrasts showed that women who had an opportunity to misattribute arousal and affirm the self (M = 5.90, SD = 2.13) performed significantly better than women not given these opportunities (M = 4.00, SD = 2.10), t(19) = 2.06, p < .03, d = 0.90, one-tailed. Comparisons of women in the misattribution cue absent/self-affirmation absent condition to those in the misattribution cue present/self-affirmation absent (M = 4.27, SD = 1.68) and misattribution cue absent/self-affirmation present conditions (M = 4.80, SD = 1.62), however, did not reach significance (all t's < 1.00). In addition, no significant main effects or interactions were detected among participants' response times (all p's > .13).

Potential Mediators

Motivation, expectancy, anxiety, and task confidence. Analyses on participants' motivation and expectancy scores did not yield any significant findings (all p's > .57). Once again, these data suggest that motivation did not mediate the effects described herein. Analyses on participants' STAI scores (α = .96) only revealed a marginally significant main effect of self-affirmation opportunity, F (1, 38) = 3.90, p = .056. Namely, women in the self-affirmation present condition (M = 2.32, SD = 0.73) reported significantly less anxiety than those in the self-affirmation absent condition (M = 2.71, SD = 0.55). Tests on participants' task confidence scores showed only a marginally significant misattribution cue X self-affirmation opportunity interaction, F (1, 38) = 2.94, p = .09. Simple effects tests revealed that women in the misattribution cue present/self-affirmation present condition (M = 6.32, SD = 1.18) reported significantly higher task confidence scores when compared to participants in misattribution cue present/self-affirmation absent condition (M = 5.02, SD = 1.49), t(19) = 2.20, p = .04. No appreciable differences emerged between participants in the remaining misattribution cue absent conditions irrespective of whether (or not) an opportunity to affirm the self was given (M = 5.36, SD = 0.94 vs. M = 5.58, SD = 1.85, respectively), t(19) = 0.34, p = .74.

In short, Study 2 generated several interesting results related to the performance of threatened women. The finding that reestablishing self-integrity before taking a difficult math test lead to better performance than not doing so, is consistent with prior research (Martens et al., 2006). Importantly, the results indicated that, while affirming the self was associated with improved performance, *only* women who concurrently affirmed the self and had a readily available misattribution cue experienced better performance when compared to threatened women devoid of either reduction strategy. The latter effect was neither additive, nor inhibitory based on the fact that women in the dual reduction strategies condition were the only group who experienced a performance gain relative to women merely exposed to threat.

The finding that no performance differences emerged as a function of the misattribution cue manipulation, however, was inconsistent with previous research (Ben-Zeev et al., 2005). Though it is possible that this manipulation lacked impact, (at least) two findings make this explanation less tenable. First, nearly all of the participants correctly completed the treatment check, suggesting it was not too subtle to produce differences. Second, it is unlikely that the inability of the misattribution cue manipulation to affect anxiety scores was due to these items being either too insensitive or delayed. In addition, the fact that the STAI has shown adequate reliability and generalizability in the past (Spielberger, Gorsuch, & Lushene, 1970; Spielberger, 1985) makes the latter account less plausible. More simplistic possibilities are that the STAI was tapping into a different construct in the current context or stereotype-threatened participants were unwilling to report on their phenomenological experiences (Bosson, Haymovitz, & Pinel, 2004). Unfortunately, the null results associated with the misattribution manipulation in Study 2 and inconclusive findings produced in other threat research using comparable frameworks (e.g., Stone et al., 1999) suggest that further replications with similar instruments and procedures are required before the full merits of this paradigm in reducing stereotype threat can be determined.

General Discussion

Collectively, the present research expands our understanding of stereotype threat in several important ways. For instance, prior studies have shown that activating negative group-based stereotypes can depress the math performance of women (Spencer et al., 1999). Study 1 replicated this effect by demonstrating that the performance of women, when compared to men, could be exacerbated by merely informing them that a math test had produced gender

differences. When the exam was described as having produced no gender differences, women and men performed equally well.

Additionally, reviews of the stereotype threat literature have often concluded that the mediation of the stereotype threat effects remains unclear (Jones & Stangor, 2003). Among women, relatively few studies (e.g., Spencer et al., 1999; Rydell et al., 2009) have produced models showing a significant link between threat, performance, and the mechanism(s) that may underlie this relation. Study 1 sought to clarify this mediational mire but only generated support for task confidence perceptions as a mechanism that may explain the *boost* (or stereotype lift) effect experienced by men under stereotype threat conditions (Walton & Cohen, 2003). Therefore, though it appears that task confidence perceptions play a role in stereotype lift processes among men—a finding supported by emerging research (e.g., Schmader, Forbes, Zhang, & Mendes, 2009)—the mediation of threat effects among women requires further exploration.

Also recall the research question guiding Study 2: "Is there benefit in using concurrent reduction strategies in a single study?" The simple answer is *yes*. Though misattribution processes failed to produce performance gains in isolation, self-affirmation was associated with a significant performance increase among the sample. However, it was *only* when women received the misattribution cue and an opportunity to affirm the self that they were buffered from the effects of stereotype threat and their performance improved relative to women in stereotype threat (or reduction strategy absent) condition.

Limitations

Despite its virtues, there are limitations to the present research. First, given that both studies were conducted in a lab, it is unclear whether these results would generalize to the real-world. Evidence from the field, however, suggests that they would (Ambady, Shih, Kim, & Pittinsky, 2001; Danaher & Crandall, 2008; Sternberg et al., 2002)—though a recent study has produced conflicting findings with respect to the applicability of threat reduction strategies to such contexts (Gillespie, Converse, & Kriska, 2010). Second, an alternative account of the meditational chain in Study 1 could not be ruled out—in other words, math performance acted as a mediator of the relationship between the stereotype threat X gender interaction and task confidence perceptions. I believe the failure to discount this explanation was rooted in the inherent tradeoff in threat research between reducing the likelihood of demand characte-

ristics and establishing a temporal order for potential mediators. Future threat researchers must give careful consideration to this issue in their methods and devise better ways of maximizing temporal clarity, while minimizing experimental demand.

Third, one can argue that without an effective control (or no-stereotype threat) condition, it is hard to determine whether threat was created in the experimental conditions of Study 2. If threat was not created, then it would be impossible to test whether there is efficacy in using two reduction strategies. Based on theoretical and empirical grounds, however, I am inclined to believe that threat was invoked (and held constant) across the conditions in this study. In terms of theory, several studies indicate that this situational predicament is the default state of affairs for stigmatized individuals in a testing context (Brown & Day, 2006; Inzlicht & Kang, 2010). Thus, creating stereotype threat in lab or field contexts should be easy to accomplish. As Steele and Davies (2003) argue, its "finding a way to eliminate stereotype threat in a testing situation [that] is the difficult thing to do..." (p. 315; brackets mine). From an empirical standpoint, examination of the means for women in the high stereotype threat conditions of Studies 1 (gender differences condition; M = 3.93, SD = 2.09) and 2 (gender differences + misattribution absent/self-affirmation absent condition; M = 4.00, SD = 2.10) suggests that my manipulations had a similar impact on participants across studies. By and large, these arguments and data support the notion that stereotype threat was created across conditions in Study 2. Evidence showing that the combination of two reduction strategies produced a significant performance increase relative to the threat induction only condition indicates that this approach is a viable way to offset this phenomenon. The lack of a true control condition does, however, limit my ability to say *how much* of a performance gain occurred.

Future Directions and Educational Implications

Although the present studies have expanded our understanding of stereotype threat, further research is warranted. For instance, research is needed to clarify the impact of this phenomenon in real-world contexts. Though the majority of published research has examined threat effects in laboratory contexts, Steele and Davies (2003) estimated that lab studies may actually *underestimate* the impact of threat when compared to what is experienced in applied settings. In addition, empirical tests of STT's disidentification hypothesis are lacking.

These findings also have implications for those in education and public policy, and can aid in setting research priorities and selecting interventions most likely protect women from stereotype threat. For example, results from Study 2 showed that using concurrent reduction strategies was a legitimate way to reduce gender-based performance gaps—particularly if offered *before* a math test. One must, however, consider the portability of these strategies to the field. Indeed, nuanced self-affirmation interventions are easy to implement, more realistic, and offer real-world applicability, whereas misattribution-based interventions may be less applicable to such contexts.

Finally, securing funds for higher education is becoming increasingly more difficult. Given that enrichment programs, college entry, and scholarships are inextricably linked to standardized test scores, underperformance in math can have a profound impact on the developmental and career trajectories of women. In addition to anecdotal evidence showing that women are making strides toward closing the gender gap in mathematics (e.g., winning prominent scientific competitions; Cogito.org, 2010), the present findings offer practical guidance in understanding how stereotypes affect women's performance and the ways in which these beliefs can be overcome. By adopting curricula and policies that address stereotype threat, scientists, educators, and policymakers can better understand, predict, and control this phenomenon in a cost-efficient and yet dramatic way.

Acknowledgments

Funding: The preparation of this manuscript was supported in part by NIH grant number 1R21CA154258-01 and the Office of Behavioral and Social Sciences Research (OBSSR). The content described herein is solely the responsibility of the author and does not necessarily represent the official views of the NCI, OBSSR, or NIH.

Additional Contributions: This article was based, in part, on my doctoral dissertation conducted under the direction of Charles Stangor at the University of Maryland. I am indebted to Hal Sigall, Judson Mills, Paul Hanges, and Seppo Iso-Ahola for their committee guidance. I owe a debt of gratitude to Ted Miller, Dexter Taylor, and Cecelia Snowden for comments on an earlier draft. Special mention goes out to Anthony Ramirez and Eduardo Romano for their help in Spanish translation and Alma Lopez for her assistance in proofreading.

Footnotes

[1] Whereas Ben-Zeev et al. (2005, Study 2) controlled for prior SAT performance in their analysis, the remaining studies cited did not.

[2] Spencer et al. (1999; Study 3) could not rule out anxiety (controlling for evaluation apprehension) as a potential mediator of stereotype threat effects in their analysis. However, evidence for anxiety as a univariate mediator of the threat--performance relationship was weak.

[3] This listing of reduction strategies is not exhaustive.

[4] Although it is possible that a male experimenter could make female participants more nervous, we followed the lead of other threat research in doing so (Oswald & Harvey, 2000-2001) and reasoned that whatever this impact is, it should be consistent across conditions given our experimental design and controlled for.

[5] Therefore, reaction times and task confidence ratings in Study 1, and the performance, reaction time, and confidence scores in Study 2 were also based on these nine items.

[6] The former reasoning is in line with Cohen et al. (2009) who found that self-affirmation processes can have a lasting, positive impact on the performance of stigmatized individuals.

[7] GPA and SAT values were based on data from introductory psychology students at the University of Maryland.

References

Ad Council & Girl Scouts of the USA. (2004). Girls go tech: It's her future. Do the math. Retrieved March 4, 2004, from http://www.girlscouts.org/for_adults/leader_magazine/2003_spring/its_her_future.asp

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*, 385-390.

Aronson, J., Cohen, G. L., & Nail, P. R. (1999). Self-affirmation theory: An update and appraisal. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance theory: Progress on a pivotal theory in social psychology*. Washington, DC: American Psychological Association Books.

Aronson, J., & Damiani, M. (1997). *Stereotype threat, attributional ambiguity, and fragile self-competence*. Unpublished manuscript, Austin, TX.

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

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

Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). Stereotype Threat and the Academic Underperformance of Minorities and Women. In J. K. S. C. Stangor (Ed.), *Prejudice: The Target's Perspective* (pp. 83-103). San Francisco, CA: Academic Press.

Barnard, L., Burley, H., Olivarez, A., & Crooks, S. (2008). Measuring vulnerability to stereotype threat. *Electronic Journal of Research in Educational Psychology, 6,* 51-64.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2009). Stereotype threat and working memory: Mechanisms, alleviation, and spill-over. *Journal of Experimental Psychology: General, 136,* 256-276.

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

Benbow, C. P. (1988). Sex differences in mathematical ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences, 11,* 169-232.

Benbow, C. P., & Minor, L. L. (1986). Mathematically talented males and females and achievement in high school sciences. *American Educational Research Journal, 23,* 425-436.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210,* 1262-1264.

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

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

Brown, R. P., & Day, E. A. (2006). The difference isn't Black and White: Stereotype threat

and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology, 91*, 979-985.

Callahan, C. (1991). An update on gifted females. *Journal for the Education of the Gifted, 14*, 284-311.

Chandler, M. (Writer) (1999). Secrets of the SAT [Televison series episode]. In M. Chandler (Producer), *Frontline*. Virginia: Public Broadcasting Service (PBS).

Cogito.org. (2010, March). *2010 Intel Science Talent Search winners*. Retrieved June 28, 2011, from http://cogito.cty.jhu.edu/17913/2010-intel-science-talent-search-winners/

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science, 324*, 400-403.

Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*, 1639-1655.

Dess, N. K. (2001, Spring). Statistics: Grin Big and Bear it! *The APAGS Newsletter, 13*, 20.

Educational Testing Service (1994). *GRE: Practicing to take the General Test* (9th ed.). Princeton, NJ: Educational Testing Service.

Educational Testing Service (2002). *Sex, Race, Ethnicity, and performance on the GRE General Test*. Princeton, NJ: Educational Testing Services.

Gillespie, J. Z., Converse, P. D., & Kriska, S. D. (2010). Applying recommendations from the literature on stereotype threat: Two field studies. *Journal of Business and Psychology, 25*, 493-504.

Gladwell, M. (2008). *Outliers: The story of success*. New York, NY: Little, Brown and Company.

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

Halpern, D. F. (1989). The disappearance of cognitive gender differences: What you see depends on where you look. *American Psychologist, 44*, 1156-1158.

Inzlicht, M., & Kang, S. K. (2010). Stereotype threat spillover: How coping with threats to social identity affects aggression, eating, decision making, and attention. *Journal of Personality and Social Psychology, 99*, 467-481.

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: The influence of emotion regulation. *Journal of Experimental Psychology:*

General, 137, 691-705.

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

Jones, P. R. (2005). The effects of stereotype threat on the test performance and task choices of women. *Dissertation Abstracts International, 66*, 3465.

Jones, P. R., & Stangor, C. (2003). *The moderators and mediators of stereotype threat: A meta-analysis*.Unpublished manuscript, College Park, MD.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 233-265). Boston, MA: McGraw-Hill.

Liben, L. S. (1978). Performance on Piagetian spatial tasks as a function of sex, field dependence, and training. *Merrill Palmer Quarterly, 24*, 97-110.

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*, 236-243.

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

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

Neisser, U. (1998). *The Rising Curve: Long-term gains in IQ and related measures*. Washington, DC: American Psychological Association.

Nguyen, H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*, 1314-1334.

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

Oswald, D. L., & Harvey, R. D. (2000-2001). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology: Developmental, Learning, Personality, Social, 19*, 338-356.

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. Special Issue: Stigma: An insider's perspective. *Journal of Social Issues, 57*, 55-71.

Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology, 96,* 949-966.

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist, 59,* 7-13.

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

Schmader, T., Forbes, C. E., Zhang, S., & Mendes, W. B. (2009). A meta-cognitive perspective on cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin, 35*, 584–596.

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

Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review, 16,* 177-206.

Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 6*, 1040-1057.

Spencer, S. J., & Steele, C. M. (1992). The effect of stereotype vulnerability on women's math performance. In J. Jonides (Chair), *Psychological mediators of academic achie vement and intervention programs*. Symposium conducted at the Annual Convention of the American Psychological Association, Washington, DC.

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

Spielberger, C. D. (1972). *Anxiety: Current trends in theory and research* (Vol. 1). New York: Academic Press.

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

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *State-Trait Anxiety Inventory manual*. Palo Alto, CA: Consulting Psychologists Press.

Stangor, C. (1999, November 4-6). *Consequences of stereotypes on children.* Paper presented at the Multidisciplinary Research Conference on Stereotypes, Prejudice, Tolerance,

and Rights: Educating children for living in diverse cultures, Bethesda, MD.

Stangor, C., & Sechrist, G. B. (1998). Conceptualizing the determinants of academic choice and task performance across social groups. In J. K. S. C. Stangor (Ed.), *Prejudice: The target's perspective* (pp. 105-124). San Diego: Academic Press.

Steele, C. M. (1992). Race and the schooling of Black Americans. *The Atlantic Monthly*.

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

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

Steele, C. M., & Aronson, J. (1998). Stereotype Threat and the Test Performance of Academically Successful African Americans. In C. J. M. Phillips (Ed.), *The Black-White Test Score Gap* (pp. 401-427). Washington, DC: Brookings Institution Press.

Steele, C. M., & Aronson, J. (2004). Stereotype threat does not live by Steele and Aronson (1995) alone. *American Psychologist, 59*, 47-48.

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

Sternberg, R. J., Jarvin, L., Leighton, J., Newman, T., Moon, T., Callahan, C., et al. (2002). *Girl's can't do math? The disidentification effect and gifted high school students' math performance*.Unpublished manuscript, New Haven, CT: Yale University.

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

Stroessner, S., & Good, C. (2010). What are unresolved issues about stereotype threat? Retrieved June 4, 2010, from http://reducingstereotypethreat.org/unresolved.html

Stumpf, H., & Stanley, J. C. (1998). Stability and change in gender-related differences on the College Board Advanced Placement and Achievement Tests. *Current Directions in Psychological Science, 7*, 192-196.

Walton, G., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456-467.

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20*, 1132-1139.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*, 797-826.

Wicherts, J. M. (2004). Stereotype threat research and the assumptions underlying analysis of covariance. *American Psychologist, 59*, 7-13.