

## ABSTRACT

Title of Document: THE EFFECTS OF STEREOTYPE THREAT ON THE  
TEST PERFORMANCE AND TASK CHOICES OF  
WOMEN

Paul R. Jones, Ph.D., 2005

Directed by: Professor Charles G. Stangor, Department of Psychology

Can the activation of a prevalent stereotype alleging female math inferiority influence the math performance and task choice behavior of women? If so, what mediates each of these effects? In addition, what strategies can be used to reduce the impact of this stereotype on the performance of women? Three studies examined these questions by using techniques derived from stereotype threat (Steele, 1992), self-affirmation (Steele & Liu, 1983), misattribution (Schachter, 1964) and stigma-threat (Blascovich et al., 2001a) research. In Studies 1 and 2, collegiate women and men were (or were not) presented with a gender differences (or no gender differences) instructional set either prior to completing a math test or prior to selecting an upcoming task, respectively. Study 1 demonstrated that women performed more poorly on a math test after receiving the gender differences instructional set when compared to their male counterparts. However, no gender differences emerged when women and men received a gender fair instructional set. In addition, Study 1 revealed that the gender X instructional set interaction effect on performance was mediated by task confidence perceptions—although the confidence perceptions of men heavily influenced this effect. Study 2 found a trend that suggests that

the instructional set manipulation may also have implications for participants' choice behavior. Whereas women appeared to be more likely to choose a math task over a proofreading task, when presented with a gender differences instructional set, women displayed the opposite choice pattern after receiving a gender fair instructional set. The trend amongst men suggested that they were more likely to choose a math task over a proofreading task irrespective of instructional set. Study 3 examined whether the performance deficits experienced by women could be reduced by employing either self-affirmation or misattribution processes. The results demonstrated that these deficits were alleviated when women were allowed to affirm the self prior to completing a math task. These findings are discussed in relation to stereotype threat theory and to potential educational interventions. Future directions for stereotype threat research are also discussed.

THE EFFECTS OF STEREOTYPE THREAT ON THE TEST PERFORMANCE  
AND TASK CHOICES OF WOMEN

by

Paul R. Jones

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2005

Advisory Committee:

Professor Charles Stangor, Chair  
Professor Harold Sigall  
Professor Judson Mills  
Professor Paul Hanges  
Professor Seppo Iso-Ahola

© Copyright by

Paul R. Jones

2005

## **DEDICATION**

The present dissertation is dedicated to my grandparents, Frank and Virginia Jones and Pedro and Iva Torres. They have often served as a source of inspiration and continue to be an integral part of my life. Without their unwavering support, the present dissertation would not have been possible. In addition, I would also like to recognize my late aunts Mabel Armstrong and Adeline Jones. Their kind hearts, warm spirits, and commitment to family will be sorely missed.

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Creator. Without His wisdom and guidance this document would not have been possible. I would also like to extend my sincere appreciation to Dr. Charles Stangor for his vision and encouragement throughout the dissertation process. His guidance and tireless efforts on the present manuscript have truly made it a rewarding experience and I'm blessed to have had him as an advisor. Dr. Stangor has been a magnificent role model for me and I would like to express my sincere gratitude to him for his support and guidance over the years.

In addition, I would like to acknowledge the members of my dissertation advisory committee, Drs. Charles Stangor (Chair), Harold Sigall, Judson Mills, Paul Hanges, and Seppo Iso-Ahola. Their insights on the manuscript were extremely helpful and have undoubtedly contributed to the quality of the final document.

To my parents, Paul and Maria Jones, my brothers, Justin and Brian Jones, my nephew Devin Jones, my uncle, Roosevelt Brown III, my cousin, Frank Jones III, my close friends, Russell Sisco, Jr. and Dwayne Cooper, my 'extended' parents, Russell and Viola Sisco, and my God-daughter, Atiana Sisco, there are no words that can describe the sincere love and appreciation that I have for you all. Your love and support has been instrumental in both my personal and professional development and I would not be where I am today without you.

I would also like to recognize my mentors, Drs. James M. Jones, Terell Lasane, Clayton Stansbury, Carrol Perrino, Robert Smith, Henrietta Hestick, Earl Walker, and

Kim Nickerson. You've all been extremely important in my educational journey and I can not begin to thank you enough for all that you've done for me. Your insights, patience, and understanding have truly made an indelible mark on my life.

I would like to further recognize two special people in my life, Toni Betton and Anna Couch, whom I consider lifetime friends. I'd like to sincerely thank you both, from the deepest recesses of my heart, for every moment that we have spent together. You are truly masters in the art of listening and it was through your kind words, inspiration, and compassion, that I have learned the true meaning of friendship, patience, and forgiveness.

I would like to acknowledge my Pacific Institute for Research and Evaluation family—Monique Sheppard, Cecelia Snowden, Charles Hayes, Ted Miller, Dexter Taylor, Bruce Lawrence, Amy Owens, Doug Hill, Mark Johnson, Latifa Boyce, Stephanie Stevens, and Debra Furr-Holden, my research assistants—Margarita Chukhina and Megan Wareham, as well as, Fiona Ewan, the Wahabs, Fritzi and Aaron Hart, Charlene Weaver, and Leah Strong. Your wisdom, advice, patience, and encouragement have been pivotal in my personal and professional development.

And finally, I would like to acknowledge Cynthia, Marina, and Raymond Mendoza, John, Claudia, and Miguel Battle, Carolina, José, Cynthia, Mark Anthony, Jalen, and Anna Flores, LaJuan Watson, Eric Betters, and Roxanna and Giovanni Contreras. My experience with all of you has been a rewarding one and I'm truly blessed to have made your acquaintance. I will always have fond memories of our time together.

## TABLE OF CONTENTS

DEDICATION .....	ii
ACKNOWLEDGEMENTS .....	iii
TABLE OF CONTENTS .....	v
LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
INTRODUCTION.....	1
Gender Differences in Standardized Testing .....	3
Theoretical Explanations for Gender Differences in Performance .....	4
Biological Explanations of Performance Differences .....	5
Social/Psycho-social Explanations of Performance Differences .....	7
Stereotype Threat Theory.....	9
Disidentification Hypothesis.....	11
Vanguard Hypothesis .....	11
Generalization of Stereotype Threat .....	13
Stereotype Threat and Non-performance Based Outcomes .....	14
Stereotype Lift.....	16
Mediation of Stereotype Threat .....	17
Alleviating Stereotype Threat .....	20
Additional Directions for the Alleviation of Stereotype Threat.....	22
Self-Affirmation Theory .....	22
Generalization of Self-affirmation Theory.....	24
Self-affirmations and Stereotype Threat .....	26
Misattribution of Arousal.....	27
Misattribution and Threat .....	29
Perceptions of “Threat” vs. “Challenge” .....	33
Psychophysical Measurement of Threat and Challenge Motivational States .....	34
Empirical Support for “Threat” vs. “Challenge” Perceptions and Their Impact on Task Performance .....	35
Integrating “Threat” vs. “Challenge” Perceptions into the Stereotype Threat Paradigm .....	37
Goals and Research Hypotheses .....	40
STUDY 1 .....	45
Method .....	45
Design and Participants.....	45
Exclusion of QSAT Score as a Covariate .....	46
Procedure .....	46
Stereotype Threat Manipulation.....	48
Gender and Self-doubt Activation .....	50
Post-test Measures .....	54
Results.....	55



Manipulation Check.....	55
Task Performance .....	56
Potential Mediators: Implicit Measures .....	58
Potential Mediators: Self-Report Measures .....	59
Mediation Analysis .....	62
Potential Moderators.....	67
Discussion.....	70
STUDY 2 .....	72
Method .....	72
Design and Participants.....	72
Procedure .....	73
Task Choice and Strength of Choice.....	74
Results.....	75
Manipulation Check.....	75
Task Choice .....	76
Strength of Choice .....	78
Reaction Time for the Chosen Alternative .....	78
Other Dependent Measures .....	79
Exploratory Measures .....	83
Discussion.....	84
STUDY 3 .....	87
Method .....	87
Design and Participants.....	87
Procedure .....	88
Malfunctioning Computer Cover Story and Misattribution Manipulation .....	88
Self-Affirmation Manipulation .....	90
Results.....	92
Manipulation Checks .....	92
Task Performance .....	93
Potential Mediators: Implicit Measures .....	94
Potential Mediators: Self-Report Measures .....	96
Mediation Analysis .....	98
General Discussion .....	103
Implications for Stereotype Threat Theory .....	106
Implications for Educational Environments .....	107
Future Directions for Stereotype Threat Research.....	108
FOOTNOTES .....	111
APPENDICES .....	115
Appendix A - Consent Form.....	115
Appendix B - List of Critical Items Used on the Gender and Self-Doubt Activation Measures .....	116
Appendix C - Items Used on the Performance Task in Studies 1 and 3 .....	117
Appendix D - Achievement Motivation Measure.....	122

REFERENCES .....	vii
	123

**LIST OF TABLES**

Table 1.	Mean Number of Items Correct as a Function of Gender and Instructional Set.....	57
Table 2.	Mean Task Confidence as a Function of Gender and Instructional Set.....	61
Table 3.	Mean Number of Gender Word Completions as a Function of Gender and Instructional Set.....	80
Table 4.	Mean Number of Gender Word Completions as a Function of Self-affirmation and Misattribution Opportunity.....	95

## LIST OF FIGURES

Figure 1.	Charts depicted in the video to reinforce the performance differences of men and women on the upcoming task in the gender fair and gender differences conditions.....	49
Figure 2.	Mediation of the gender X stereotype threat interaction effect on task performance via task confidence perceptions.....	64
Figure 3a.	Number of participants choosing a given type of task as a function of gender in the gender fair conditions.....	77
Figure 3b.	Number of participants choosing a given type of task as a function of gender in the gender differences conditions.....	77
Figure 4.	Visual depiction of each computer's orientation.....	89
Figure 5.	Mediation of the self-affirmation effect on task performance via anxiety.....	100

## INTRODUCTION

*“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, PhD, Senior Scientist, APA Science Directorate*

Nancy Dess’s (2001) remarks about her math ability and her reluctance to engage in mathematical tasks convey a commonly held female math inferiority stereotype. Although this topic has been well documented and has spawned considerable research interest (Herrnstein & Murray, 1994; Jencks & Phillips, 1998; Jensen, 1969)—with potential explanations for this phenomenon stemming from those rooted in biology to those rooted in sociology—the present thesis focuses on the social psychological perspective in this debate.

I believe that gender-based stereotypes play an integral role in the experiences of women with profound implications for their performance and task choices in the academic domain. I further reason that by examining the impact of these stereotypes on both the performance and the academic choices of women, I can identify the conditions under which women are more (or less) likely to succumb to these beliefs and buffer these individuals from their deleterious effects.

The primary goals of this thesis were fourfold. First, I wanted to understand how the explicit activation of a stereotype alleging female math inferiority would impact the math performance and situational perceptions of women. Second, I wanted to examine

what affective, cognitive, or motivational processes may underlie these effects. Third, I wanted to understand whether the opportunity to affirm the self or to misattribute arousal would buffer women from the deleterious effects of this stereotype. Finally, I sought to examine whether activating the female math inferiority stereotype would have any implications for the types of tasks that women would be interested in engaging in.

The sections of this manuscript that follow present the relevant background, theory, and methodology that accompany three experiments designed to address the primary goals stated above. In the introduction, I describe the relevant literature on gender differences in standardized testing. I then outline several theoretical perspectives on this topic and present stereotype threat theory (Steele, 1992) as an alternative approach to understanding this phenomenon. The theoretical parameters of stereotype threat theory are presented along with empirical research supporting this model. I then introduce self-affirmation theory (Steele & Liu, 1983) and Schachter's (1964) two-factor theory of emotion and discuss their theoretical parameters and empirical research supporting these models. I further discuss how these theories are relevant to stereotype threat theory particularly in terms of how both self-affirmations and misattribution processes can serve as potential reduction strategies for stereotype threat outcomes. The potential parallels between stereotype threat theory and threat vs. challenge perceptions—as presented in a bio-psychosocial model of stigma threat (Blascovich, Mendes, Hunter, Lickel, Kowai-Bell, 2001)—are also discussed. I conclude the introduction with a general overview of the proposed experiments and a presentation of the research hypotheses.

In the final section of the manuscript, I present three experiments and detail the

relevant findings generated in these studies. I then provide a discussion of the results followed by my suggestions for future research.

## **Gender Differences in Standardized Testing**

In the area of standardized testing, performance differences between men and women are readily apparent. The performance levels of the top male students are superior to that of their female counterparts on many of the major college entrance exams (i.e., Preliminary Scholastic Aptitude Test [PSAT], Scholastic Aptitude Test [SAT], American College Teaching Program [ACT]) (Callahan, 1991). For instance, on the math sub-section of the SAT, women lagged some 40 points below the scores posted by men (Halpern, 1989).

Recent Graduate Record Examination (GRE) scores convey a similar story (Educational Testing Service [ETS], 2002) with women continuing to lag behind men in the area of mathematics. In 2000-01, women scored well below the performance level of men on the quantitative portion of the GRE (GRE-M) ( $M_{\text{women}} = 545$ ,  $SD = 144$ ;  $M_{\text{men}} = 641$ ,  $SD = 139$ ). Even more perplexing is that whereas women have been demonstrated to score lower, on average, than men on standardized exams—even when they have received comparable preparation—when grades are used as a measure of performance, women score consistently higher than men even when both groups are equally prepared (Callahan, 1991; Wainer & Steinberg, 1992). This paradox is one of the major reasons why researchers continue to search for explanations of this phenomenon. And although there continues to be a debate regarding the predictive validity and utility of standardized

examinations in the educational and psychological literatures<sup>1</sup> (Callahan, 1991; Sackett, Schmitt, Ellingson, & Kabin, 2001), there is little doubt that these tests play a pivotal role in access to educational resources (e.g., enrichment programs, scholarships, fellowships), admission to selective institutions, and in the career paths chosen by women.

### **Understanding the Career Path Implications of Gender Performance Differences**

Further evidence regarding the long-term implications of the gender performance gap is evident in the literature devoted to the representation of women in science and engineering (for a review, see National Science Foundation [NSF], 1996). Women are sorely underrepresented in these fields composing less than a quarter (22%) of the nation's scientists and engineers (NSF, 1996). This finding is even more distressing considering that women compose more than half of the U. S. population (51%) and nearly one-half of its workforce (46%) (NSF, 1996). Even when women do choose career paths in these fields, they tend to be concentrated in specific sub-specialty areas (e.g., psychology) as opposed to a more broad representation in these disciplines. These findings suggest that women's performance, particularly in the quantitative areas, may be involved in their career decision-making process. If this is the case, then it becomes important to further uncover the determinants of such outcomes and to understand how women use their performance in these areas to inform their decisions about what career path they will pursue (for a review, see Stangor & Sechrist, 1998).

### **Theoretical Explanations for Gender Differences in Performance**

Several theoretical perspectives have been advanced to account for gender



differences in performance. These theoretical frameworks can be further classified into one of two general camps that stress either (1) biological or (2) social/psycho-social mechanisms as the key to understanding both gender and racial divides in academic performance.

### ***Biological Explanations of Performance Differences***

Theorists in the *biological mechanisms* camp (Benbow & Stanley, 1980; Herrnstein & Murray, 1994; Jensen, 1969) stress the importance of genetic factors as primary contributors to the underperformance of women and minorities (for a discussion, see Jenks & Phillips, 1998; Kolata, 1980). Although this perspective has been met with sharp criticism (for a review, see Graves, Jr., 2002), several studies have generated empirical support for this approach (Benbow, 1988; Benbow & Minor, 1986; Benbow & Stanley, 1980; Benbow & Stanley, 1984). In one provocative study, Benbow and Stanley (1980) administered both a math and verbal portion of the SAT to precocious male and female junior high school students. As predicted, male students outperformed female students on the mathematics portion of the examination, whereas both groups performed comparably on the verbal portion. These researchers posited that since these students had not been taught the basic principles and methodologies necessary to solve such problems (given their status as junior high school students), any performance differences that emerged were at least due in part to genetics.

Despite these findings, more recent accounts (Jenks & Phillips, 1998; Neisser, 1998) continue to de-emphasize the impact of genetic factors in gender and racial

performance differences. Proponents of this perspective cite several lines of empirical research that contradict the biological approach and view these data as evidence that a purely biological model can not fully account for such performance differences. These findings include the following:

(1) When minorities or mixed-race children are reared by European American families, they experience a boost in their standardized examination scores.<sup>2</sup>

(2) Examination scores are subject to fundamental socio-environmental changes (e.g., school desegregation). For instance, a phenomenon referred to as the *Flynn effect* has been documented across several administrations of IQ exams. More specifically, this phenomenon is characterized by the robust finding that over time, massive gains in IQ scores will invariably emerge (Flynn, 1998; Neisser, 1998). Based on the data from 20 countries, Flynn (1998) has estimated that on IQ tests of *fluid* intelligence—that is, “...the minds ability to solve problems at the moment...” (p. 26), a consistent 3-point gain per decade increase has been observed on these exams since their inception.

(3) There has been a gradual narrowing of prominent performance gaps over the past several decades (e.g., a reduction in the African American/European American test score gap). As evidenced by scores on the National Assessment of Educational Progress Exam (NAEP)—a nationally administered exam believed to be more representative than the other standardized exams—the performance experienced by African Americans and Hispanics (when compared to European Americans) has been substantially reduced over the last 30 years (Neisser, 1998).

These findings cast doubt upon the plausibility of a purely biological model as a

viable explanatory framework of accounting for gender and racial performance differences. In addition, these findings cast doubt upon a major tenet of the biological approach—the notion that intelligence is a *fixed* phenomenon (Neisser, 1998).<sup>3</sup> Given the tendency for biological effects to occur more gradually, the relatively short time frame necessary to produce a narrowing of gender and racial gaps suggests that a biological approach is not likely to be solely responsible for these effects. Moreover, the relative paucity of direct evidence for the biological perspective has added to the large degree of skepticism associated with this approach (Neisser, 1998).

### ***Social/Psycho-social Explanations of Performance Differences***

Accounts from the social/psycho-social mechanisms camp have offered both social and societal factors as the root of both gender and minority performance differences. Some theorists from this perspective posit that these performance differences are heavily rooted in societal disadvantages (Steele, 1992). Such disadvantages may be manifest in both socialization (Boykin & Toms, 1985; Kolata, 1980; Steinberg, Dornbusch, & Brown, 1992) and socioeconomic differences (for a review, see White, 1982).

Anecdotal evidence has often supported the notion that socioeconomic differences directly influence academic performance (White, 1982). However, in a meta-analysis of over 100 empirical studies examining the relationship between socioeconomic status (SES) and academic achievement, White (1982) discovered that the correlation between these variables was relatively weak—especially when SES and academic achievement were computed using individual data (as opposed to aggregate data) as the unit of

analysis. Such findings cast doubt upon the contention that SES differences can explain the performance deficits experienced by women and minorities.

Other theorists (Steele, 1998) have suggested that societal disadvantage, socioeconomic hardship, and poor academic achievement are rooted in defective value systems. Steele argues that core principles such as personal responsibility and a devotion to excellence in all pursuits (that is not tied to the dominant culture in the form of affirmative action, welfare, etc.) must be internalized, communicated, and understood (p. 108). It is assumed that through the internalization of these core values, progressive social reform can be achieved in both the economic and academic domains. Although intuitively appealing, such internalization processes have rarely been examined empirically.

Other researchers (Eccles, 1987; Jussim & Fleming, 1996) have suggested that psycho-social mechanisms, such as expectancies, are directly involved in the performance decrements experienced by women and minorities. Indeed there is a burgeoning literature devoted to the examination of how task relevant expectancies influence task performance (Bandura, 1977; Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Dolly, Bell, Reynolds, & Saunders, 1979; d'Ydewalle, Swerts, & De Corte, 1983; Shih, Pittinsky, & Ambady, 1999; Stangor & Carr, 1997; Stangor, Carr, & Kiang, 1998; Stangor & Sechrist, 1998). For instance, Jussim and Fleming (1996) provide a narrative review of research devoted to self-fulfilling prophecies and their impact on task performance. These researchers note that, "...a self-fulfilling prophecy occurs when an erroneous social belief leads to its own fulfillment" (p. 161) and point out that such effects have been

demonstrated empirically in many areas including education, commercial banking, and job interviewing (Jussim, 1989; Jussim, 1991).

In a classic study by Rosenthal and Jacobsen (1968), teachers who were led to believe that some of their students were “late bloomers” actually influenced these students to make substantial performance gains over their non-labeled classmates. Similarly, Dickstein and Kephart (1972) found that female participants performed better on an intelligence test when an experimenter explicitly provided a high expectancy for these students. These studies further underscore the power that task relevant expectations maintain over task performance.

### **Stereotype Threat Theory**

Claude Steele and his colleagues at Stanford University have proposed stereotype threat theory (Steele, 1992; 1997) as a promising theoretical approach in explaining the performance deficits experienced by women and minorities. Stereotype threat theory maintains that an individual may experience apprehension about the possibility of validating a (negative) stereotype that exists for their respective group in a given domain (Steele, 1997; Steele & Aronson, 1998). This situational predicament has been referred to as *stereotype threat* or *stereotype vulnerability* and has been 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).

According to this approach, relevant group-based stereotypes are purported to account for the performance decrements experienced by women and stigmatized group members as opposed to a genuine lack of ability. Moreover, stereotype threat theory posits that the minimal conditions necessary to invoke stereotype threat are threefold: one must (1) “simply have the test recognizable as a test...” and it must be both (2) difficult and (3) diagnostic of ability (Steele & Davies, 2003, p. 10).

In a set of seminal studies, Steele and Aronson (1995; Studies 1 & 2) examined the performance of African American and European American college students under conditions designed to either evoke or nullify—that is, by rendering the stereotype irrelevant to the current situation via leading participants to perceive the task as a problem solving instrument—stereotype threat. Participants were randomly assigned to complete a difficult verbal task described as either *diagnostic of ability*—hence, making the stereotype of African American underachievement relevant to the situation—or as *non-diagnostic of ability*. As predicted, African Americans performed worse than European Americans on the task when it was described as diagnostic of ability, whereas no racial differences emerged when the task was described as non-diagnostic. The remaining studies in this set (Studies 3, 4, & 5) produced similar results using a more subtle manipulation of stereotype threat (a race prime) and this pattern of results has been replicated by many studies within the stereotype threat literature (Gonzales, Blanton, & Williams, 2002; McKay, Doverspike, Bowen-Hilton, & Martin, 2002).

### ***Disidentification Hypothesis***

A relatively unexplored tenet of stereotype threat theory addresses the behavior of individuals who experience this situational predicament over time. Steele (1992; 1997) posits that prolonged exposure to stereotype threat can lead to a process referred to as *disidentification*—a more chronic type of domain avoidance. In an effort to buffer their self-esteem, the threatened individual may elect to disengage from tasks in a stigmatized domain and may no longer view their performances in that area as a vital part of their self-concept (Osborne, 1995; Spencer et al., 1999). Although both the inhibited task performance and disidentification processes are integral components of stereotype threat theory, the later tenet has been virtually ignored within this literature.

### ***Vanguard Hypothesis***

One interesting moderator of stereotype threat effects is the extent to which the individual views a given domain as an important part of the self-concept. It is presumed that stereotype threat will have a stronger impact upon those who are the most invested in a respective domain. Given that those who are highly identified with a domain are likely to be confident in their abilities within that area, the threat of confirming a negative group-based stereotype should be particularly salient to these individuals—especially when completing a difficult and diagnostic test of ability. In contrast, individuals who maintain lower levels of domain identification are presumed to be less invested in their performance within that area. Thus, the prospect of confirming a negative group-based stereotype is assumed to be less salient to these individuals and, in turn, substantially

reduces the likelihood that stereotype threat processes will impair their performance.

Empirical support for this contention has been demonstrated in several studies (e.g., Aronson et al., 1999; Leyens, Désert, Croizet, & Darcis, 2000). For instance, Aronson et al. (Study 2) presented both highly and moderately domain-identified European American male calculus students with an Asian math superiority stereotype prior to completing a difficult math test. As predicted, high math identification participants confronted with the stereotype performed more poorly on the task when compared to their non-stereotype activated counterparts. However, for those moderately identified with math, activating the Asian math superiority stereotype actually led to superior performance when compared to their counterparts for whom the stereotype remained non-activated.

The benefits of these studies were twofold. First, these studies underscore the importance of individual differences in domain identification as an important moderator of stereotype threat. Second, these studies demonstrated the generality of stereotype threat to members of non-stigmatized groups (e.g., European American males) which further supports the notion that the presence of a chronic stigma is not required to experience threat outcomes. Indeed, stereotype threat can influence anyone given that there is a negative stereotype associated with their social group in a given domain and that these individuals place importance on the respective domain. And since domain identification is an important variable in stereotype threat research, I assessed the extent to which participants were identified with mathematics in the studies reported in this thesis.



### ***Generalization of Stereotype Threat***

Steele (1997) maintains that stereotype threat is a general phenomenon that can be described as “...a situational threat—a threat in the air—that, in general form, can affect the members of any group about whom a negative stereotype exists (e.g., skateboarders, older adults, White men, gang members).” (p. 614). This contention is further illuminated when he notes, “...everybody experiences stereotype threat because we’re all members of one group or another that is negatively stereotyped in society” (as cited in Chandler, 1999). Therefore, stereotype threat may be experienced by anyone, providing that the individual is highly identified with a domain and ascribes subjective value or importance to it (Aronson et al., 1999; Aronson, Steele, Salinas, & Lustina, 1998b).

To date, stereotype threat has been examined in almost 100 empirical studies and this effect has been found in both published and unpublished work, dissertations, and theses (Jones & Stangor, 2003). Steele and Davies (2003) note that, “...the effect (of stereotype threat) has now been demonstrated in different groups, on different tests and behaviors, under different conditions, in several countries, and by many different investigators.” (p.10; brackets mine). A recent meta-analytic review by Jones and Stangor (2003) has established a medium effect size ( $d = .40$ ) for the impact stereotype threat on the task performance of stigmatized individuals with this effect being demonstrated across studies, manipulations (e.g., race primes, minority status), tasks (e.g., political knowledge, Math, sports, memory) and participants (e.g., women, children, the elderly, low SES, Blacks, and Whites). Therefore, there is little doubt as to whether stereotype threat is a robust phenomenon. This fact has undoubtedly contributed to the substantial

amount of interest in this phenomenon from psychologists, educators, policy makers, and the media at large (Ad Council & Girl Scouts of the USA, 2004; Chandler, 1999; Sackett, Hardison, & Cullen, 2004; Mayer & Hanges, 2003; McFarland, Lev-Arey, & Ziegert, 2003).

### ***Stereotype Threat and Non-performance Based Outcomes***

Given the mounting empirical evidence supporting the existence of stereotype threat in women and stigmatized group members on performance-based tasks, there is little doubt that stereotype threat exists. However, relatively little empirical attention has been devoted to the impact of this phenomenon on non-performance based outcomes within this literature. Although embedded within the deep structure of stereotype threat theory's disidentification hypothesis, important dependent variables such as task choice have been rarely examined empirically. Clearly, the types of tasks that stereotyped individuals choose to engage in have clear relevance to their future academic opportunities (e.g., college entry, scholarships/fellowships). And it can be argued that stereotype threat processes are not only expected to influence the performance of threatened individuals, but they are also expected to bear heavily on the decisions that stigmatized individuals make regarding their academic future (e.g., whether to take a challenging math course).

Steele's (1992; 97) domain identification hypothesis posits that prolonged exposure to stereotype threat can lead vulnerable individuals to disassociate from a given domain by making performance in that area no longer relevant to their self-concept. If

this is the case, then stereotype threat effects should have clear implications for the types of tasks that these individuals choose to engage in and their preference for these tasks. For instance, the disidentification hypothesis would predict that women and stigmatized individuals would be less likely to engage in tasks in a stereotyped domain and that they would also exhibit a lower level of preference for such tasks. It would follow that, based upon prior negative experiences in a stereotyped domain, these individuals would become reluctant to approach these tasks and would gradually fail to see their performance in this domain as a relevant part of their self-concept.

However, an alternative hypothesis can be gleaned from literature devoted to women's performance in the area of mathematics (e.g., Callahan, 1991). Findings in this literature demonstrate that when high school grades are used as a measure of performance, women score consistently higher than men irrespective of preparation. These data suggest that women may not necessarily avoid tasks in a stigmatized domain—as the disidentification hypothesis would predict—but that they may actually approach such tasks. In addition, given that college bound women are likely to have obtained some degree of success at math pursuits in the past (e.g., favorable course grades), it is assumed that they would have had to score well on a performance-based measure (e.g., a non-standardized test) at some point. If this is the case, then a *women's achievement hypothesis* might predict that women may actually be more likely to actively engage in tasks within a stereotyped domain as opposed to avoiding them.

Although both of these hypotheses are intriguing, the threat literature has yet to delve into whether threat effects would generalize to non-performance based outcomes

such as task choice. Therefore, the present research explored these possibilities within an experimental context.

## **Stereotype Lift**

Whereas a considerable amount of research attention has been devoted to examining the effects of stereotype threat on women and stigmatized group members, more recent research (Walton & Cohen, 2003) has focused on the impact of activating negative group-based stereotypes on the performance of non-stigmatized group members (e.g., Whites, High SES). Although women and stigmatized group members typically display a marked decrease in their task performance after activating a negative stereotype in a valued domain, non-stigmatized group members display a boost in performance that has been commonly referred to as *stereotype lift*. More specifically, stereotype lift is defined as a "...performance boost caused by the awareness that an out-group is negatively stereotyped" (Walton & Cohen, 2003; p. 456). Non-stigmatized individuals may benefit from this effect irrespective of whether the ability of stigmatized out-group members is made salient given that people tend to link negative out-group stereotypes at the pre-conscious level (Walton & Cohen, 2003).

A meta-analytic review conducted by Walton and Cohen (2003) has found compelling evidence for stereotype lift. These researchers examined over 40 relevant studies, and found a robust ( $d = .24$ ) stereotype lift effect for non-stereotyped group members when a negative group-based stereotype about an out-group was linked to performance. Although rarely discussed in the threat literature, the effect of stereotype

lift is evident in most threat studies. Thus, the potential for stereotype lift was examined within the present research.

### **Mediation of Stereotype Threat**

Wheeler and Petty (2001) have identified 18 potential mediators of the effect of stereotype threat on task performance including: (1) the presence of distracting thoughts, (2) perceptions of test bias, (3) thoughts concerning academic performance, (4) self-worth, (5) state anxiety, (6) frustration, (7) persistence, (8) guessing, (9) time allocation, (10) self-handicapping, (11) effort, (12) perceived difficulty, (13) perceived pressure, (14) evaluation apprehension, (15) confidence, (16) self efficacy, (17) performance expectancies, or (18) self-perceptions of skill (p. 12). A more parsimonious framework offered by Jones and Stangor (2003) organizes these mediators into either *threat-related*, *cognitive*, or *strategy* classifications. According to these researchers, strategy mediators are presumed to elicit behavioral changes such as alterations in test strategy or variations in perceived test difficulty. In contrast, cognitive mediators are only assumed to assess stereotype activation. Stereotype activation and performance expectations would fall under this rubric. Threat-related mediators are only presumed to impact individuals under stereotype threat. Motivational factors (e.g., self-reported motivation) and phenomenological experiences (e.g., anxiety, physiological arousal) are included in this category.

One debilitating phenomenological experience examined by Blascovich and colleagues (2001b) was the adverse hemodynamic effects (e.g., elevated arterial blood

pressure) of stereotype threat on African Americans. Such circulatory elevations have often been associated with maladaptive personality types such as John Henryism—a behavioral predisposition to cope with social and economic stressors through high-effort output—and more chronic and severe cardiac conditions including cardiovascular disease and hypertension (James, Hartnett, & Kalsbeek, 1983; James, Strogatz, Wing, & Ramsey, 1987). As Blascovich et al. discovered, the effects of stereotype threat are not only limited to performance outcomes, but also have profound impact upon one’s psyche and overall well-being.

Although the potential underlying mechanisms of stereotype threat have often been identified and classified, many of these proposed mediators have not been systematically tested—that is, via formal mediational tests. And in instances where such statistical rigor has been applied, the results have either been mixed or null (Jones & Stangor, 2003; Smith, 2004). For instance, state anxiety is often presumed to underlie stereotype threat effects and this variable has been examined more than any other single mediator within this literature—16 studies have tested this mechanism using seven different types of measures (Smith, 2004; Jones & Stangor, 2003). Despite its broad appeal, relatively little empirical support has been generated for this mechanism. Of the 16 studies that conducted formal mediational tests, only three studies were able to uncover empirical support for any of the potential mechanisms that were tested. Two of these studies found support for anxiety as a potential mediator, whereas a single study found support for stereotype activation (Jones & Stangor, 2003). Thirteen of the remaining studies failed to produce empirical support for any of the remaining

mechanisms.

It is important to note that, although mixed or null results have emerged on many of the proposed mediators, these results should be interpreted with some degree caution for several reasons. First, although it is possible that none of the proposed mediators actually underlie stereotype threat in isolation (a possibility advanced by Hanges & Mayer, 2003; Smith, 2004), given the small number of studies that have actually conducted formal mediational analyses, such a determination may be premature. Second, several studies in this literature have failed to examine the proposed mediators after manipulating stereotype threat but prior to measuring task performance. Although demand characteristics are always a concern, using an experimental paradigm that measured a proposed mediator prior to task performance would seem important in establishing a definitive causal chain for the mechanisms presumed to underlie threat effects. Third, one plausible reason that attempts to uncover potential mediators have been unsuccessful is that these variables may have subsided by the time they were assessed—particularly if measured after task performance. It is possible that measuring phenomenological experiences such as anxiety may be diluted if an individual is given a self-report measure after a prolonged period of time. Therefore, to avoid these potential methodological shortcomings and to allow for formal mediational tests to be conducted, the present research measured several presumed mediators—implicitly and explicitly—both prior to and after task performance.

## Alleviating Stereotype Threat

Thus far, the stereotype threat literature has focused heavily on the factors necessary to induce threat, the dispositional characteristics associated with threat, and the contexts in which threat effects occur. For instance, several studies have examined how dispositions such as domain identification (Aronson et al, 1999; Stone, Sjomeling, Lynch, & Darley, 1999), cultural/racial identification (Marks, 2000; Smith, 2002), and personal theories of intelligence (Aronson, Fried, & Good, 2002) moderate the effects of stereotype threat. Other studies have examined the role of task characteristics (e.g., task difficulty; Spencer & Steele, 1992) in inducing stereotype threat. However, relatively few research efforts have been conducted with the explicit goal of examining the conditions necessary to effectively alleviate stereotype threat from the testing context—*either before or after threat is initiated*.

Several strategies have been offered as potential ways to reduce the effects of stereotype threat including (1) rendering the stereotype incorrect, irrelevant, or non-applicable to the current situation (e.g., Spencer, Steele, & Quinn, 1999), (2) Re-defining the situation as non-threatening (e.g., Steele & Aronson, 1995; Blascovich et al., 2001b), (3) through diffusion of responsibility (McIntyre, Paulson, & Lord, 2003), or (4) via misattribution processes (Brown & Josephs, 1999; Stone et al., 1999). For instance, Spencer et al. effectively removed the effects of stereotype threat from women by rendering the female math inferiority stereotype irrelevant to the experimental context. As expected, and across two studies, women presented with a quantitative exam described as having produced *no gender differences* in the past performed equally as well



as their male counterparts. However, when the same test was described as having produced *gender differences* women performed worse than participants in all of the remaining experimental conditions.

Similarly, Blascovich et al. removed the effects of stereotype threat in African Americans by redefining the situation as less threatening. These researchers found that African Americans confronted with an experimenter of the same ethnicity, who described the upcoming task as culturally unbiased, performed on par with European American participants across conditions. Indeed, only African Americans confronted with a European American experimenter describing the task as a genuine test of intelligence performed significantly worse than all other groups combined on the task.<sup>4</sup> Thus, by making it unlikely that an individual will be judged according to a negative group-based stereotype the effects of stereotype threat were once again alleviated.

McIntyre and colleagues (2003) have had similar success in demonstrating that the effects of threat can be greatly reduced by providing stigmatized individuals with information regarding the accomplishments of other in-group members. According to their account, "...one might reassure people that their group could take care of itself regardless of their own performance, thus diffusing responsibility" (p. 8). Using this logic, McIntyre et al. provided half of the participants with information regarding female achievement (e.g., accomplishments of women in medicine) prior to completing a math task. The remaining half of the participants were not given this information. As expected, no gender differences emerged when the participants had an opportunity to read information regarding the achievements of women. However, women performed worse

than their male counterparts when no gender-based achievement information was provided.

And finally, Brown and Josephs (1999) have reduced the impact of threat via misattribution processes. These researchers provided half of the women and men in their experiment with an external handicap—to account for a potential poor performance—prior to completing a math task. The remaining half of the participants were not provided with this handicap under the assumption that by presenting women with a ready made excuse for a potential failure, the burden of confirming the female math inferiority stereotype could be removed. This is precisely what occurred as women performed equally as well as men when given an external handicap. However, women performed worse than their male counterparts when an external handicap was not provided.

### **Additional Directions for the Alleviation of Stereotype Threat**

Although the strategies described above have been shown to reduce the effects stereotype threat, the present thesis offers an additional strategy that may be effective in alleviating threat effects. More specifically, I posit that the impact of stereotype threat can be effectively reduced via self-affirmation (Liu & Steele, 1986; Steele, 1988; Steele & Liu, 1983) as well as via misattribution (Schachter & Singer, 1962) processes.

### **Self-Affirmation Theory**

Steele and colleagues (1983; 1986; 1988) have proposed a theory of self-affirmation which posits that there is a self-regulatory system for perceptions of self-

integrity—that is, perceptions of moral and ethical consistency, self-esteem, self identity, and/or perception of self-control (Steele, Spencer, & Lynch, 1993). According to these theorists, individuals are motivated to maintain a positive view of their self-integrity, which is composed of the self perceptions that one is, “...competent, good, coherent, unitary, stable, capable of free choice, capable of controlling important outcomes,...[etc.]” (p.262; brackets mine). More specifically, these researchers postulate that when an individual’s self image is threatened, he or she will actively engage in a process of rationalization and self-justification (through the constant re-interpretation of experiences) in an effort to restore balance to their positive sense of self-integrity. Such threats can take multiple forms ranging from negative evaluations made by others to behaviors that contradict one’s moral or ethical standards.

It is assumed that factors such as domain relevance and favorability of the self-concept (i.e., individual differences in resilience to self-image threats) moderate the impact of self-image threats on global perceptions of integrity. Thus, when an individual is confronted with a threat to their self-integrity, he or she will be less (or more) likely (or unlikely) to respond to the threat based on the extent to which they maintain a highly favorable and stable self-concept—that is, the extent to which one maintains a high “...global self-evaluation as determined by the balance of positive-to-negative self evaluation[,]. . .the balance of positive-to-negative self knowledge (in important domains of life), the nature of one’s attachments, the beliefs that one holds....(e.g., that all people are created equal), and so on.” (Steele et al., 1993; p. 886; brackets mine). When one’s self-evaluations of integrity are predominantly negative, he or she will be more likely to

respond to evaluative threats by engaging in self-affirmation processes to restore balance to the self-image. Individuals with more favorable self concepts are assumed to be more resilient to such self-image distress and are less likely to engage in rationalizations after image threat.

There is an additional tenet that is equally central to self-affirmation theory. More specifically, when an individual engages in the reaffirmation process, the domain of the affirmation need not be related directly to the source of the threat. For instance, a reaffirmation of the global sense of self-integrity is sufficient to reduce the impact of a specific stressor (Steele, 1988). Therefore, the use of self-affirmations as an adaptive coping strategy should be successful irrespective of whether the affirmation affirms the domain of the self-image threat or one's global self-integrity. However, it is assumed that self-affirmations will be most effective when they are relevant to the domain of the self-image threat.

### ***Generalization of Self-affirmation Theory***

Self-affirmation theory has been applied to many phenomena with consistent results, including as a potential buffer for the self-esteem of abused women (Lynch & Graham-Bermann, 2000) and for layoff survivors (Wiesenfeld, Brockner, Petzall, Wolf, & Bailey, 2001). For instance, Lynch and Graham-Bermann (2000) found that self-affirmations (and psychological maltreatment) were predictive of self-esteem, but only amongst women who suffered from physical abuse and not their non-abused counterparts. Similarly, Weinsenfeld and colleagues (2001) demonstrated that, amongst full-time

employees, job security perceptions were negatively associated with positive affect. However, when these participants were given an opportunity to re-affirm the self, this once statistically reliable relationship was eliminated. In both studies, these outcomes were taken as evidence that self-affirmations may serve as a buffer to the impact of traumatic stressors on one's general sense of self-worth.

In the dissonance literature, Steele and colleagues have also examined the utility of the self-affirmation framework as an alternative explanation to cognitive consistency theories. For instance, Steele and Liu (1983) examined the extent to which participants would engage in self-rationalizations after freely choosing to engage in writing a counter-attitudinal essay (in favor of a tuition increase) in a free choice paradigm. One-half of the participants were given an opportunity to affirm the self via a value affirmation (e.g., economic—political orientation), whereas the remaining participants were given no such opportunity prior to measuring their attitudes. The relevance of the self-affirmation domain to the specific threat was also varied. As expected, the results revealed that irrespective of the domain of the self-affirmation (whether relevant [e.g., value oriented] or irrelevant [i.e., non-value oriented] to the threat), participants were less likely to rationalize their behavior—and hence, change their attitude in accordance with the essay—after receiving an opportunity to affirm the self.

According to Steele and colleagues, since the ego is given priority in self-affirmation theory (as opposed to cognitive consistency), the reason participants in the aforesaid experiment experienced dissonance is because they perceived a discrepancy between their behavior (i.e., writing in favor of a tuition increase) and their self-concept

(i.e., being a moral, honest, and competent individual). This discrepancy was posited to serve as the impetus for their aversive feelings, based upon a self-esteem maintenance motive, which sharply contrasts cognitive dissonance theory and its prediction that there is a motive for having one's thoughts, feelings, and actions in consonance (Festinger, 1957). Subsequent studies by Steele and colleagues (1986; 1988; 1993) have demonstrated that self-affirmations play a pivotal role in the dissonance reduction process. However, a more detailed coverage of this theoretical debate is beyond the scope of the present thesis.

### ***Self-affirmations and Stereotype Threat***

It is likely that stereotype threat represents a threat to the overall sense of self-integrity maintained by stigmatized individuals in an important and relevant performance domain (e.g., academic performance). Presumably, the threat of being evaluated through the lens of a negative group-based stereotype—such as being perceived as inferior at math—would seem extremely inconsistent with the self-image of being a “competent student”. Thus, the opportunity to re-affirm the self may serve as an alternative means for stigmatized individuals to cope with a potential threat to their self-image in the form of stereotype threat. To my knowledge, no published empirical study incorporating self-affirmation theory with stereotype threat theory exists. Moreover, to tie together two literatures, usually examined in isolation, would seem to be a solid contribution to both theoretical perspectives and would reduce the possibility of the duplication of efforts. Thus, a major goal of this thesis is to explore this possibility by examining the extent to

which self-affirmations can alleviate the effects of stereotype threat.

### **Misattribution of Arousal**

Originally formulated by Schachter (1964), the psychology of emotion represented a broad research base that not only spawned interest in how emotions are formed and experienced, but also addressed the malleability of emotions as phenomenological experiences. According to Schachter's two-factor theory of emotion, there were two central aspects to this phenomenological experience, (1) an initial physiological arousal and a (2) subsequent cognitive labeling of the arousal. The first phase was posited to be experienced by everyone in a similar fashion, whereas the second phase was highly dependent on the situational context. Thus, the experience of emotion was a both malleable and context dependent. Moreover, since the labeling phase of emotion was cognitive in nature, it allowed for the possibility that emotion could be misattributed to other factors in the social milieu. And one could manipulate the arousal level, the cognitive label, or both (Cotton, 1981; p. 367).

In a classic study, Schachter and Singer (1962) injected participants with either a shot of epinephrine or a saline solution and subsequently provided a cognitive label for the injection. The participants were either provided with correct information regarding the injection, incorrect information about the injection, or no information at all. The situational context was varied by the introduction of a confederate who was instructed to either act euphoric (e.g., throwing paper in a trash can simulating basketball shooting) or disgruntled (i.e., after being expected to fill out a long survey with revealing and insulting

items, the confederate throws the paper into the ground walks off). Subsequent self-reports and behavioral measures of anger and happiness were recorded. As predicted, participants in the euphoric condition indicated a higher degree of happiness on both types of measures when provided with either the misinformation label or no label at all. Similarly, in the anger condition, participants given epinephrine with no cognitive label were found to exhibit more behavioral and self-reported anger, than correctly informed participants, although differences on the latter measure were not statistically reliable. When taken together, these researchers reasoned that when arousal is salient, without a pertinent label, the individual will (mis)attribute the arousal to an emotion (Cotton, 1981).

Such effects have been replicated with success across researchers, research paradigms (e.g., excitation transfer paradigm; Zillmann 1971; 1972), and domains including altruistic behavior (Harris & Hwang, 1973), the perception of humor (Schachter & Wheeler, 1962), interpersonal attraction (Dutton & Aron, 1974), sexual arousal (Cantor, Bryant, & Zillman, 1974) and most notably in dissonance research (Zanna & Cooper, 1974) to resolve theoretical disputes between the competing perspectives of cognitive dissonance (Festinger, 1957) and self-perception theory (Bem, 1972) (for a review, see Cotton, 1981). In a widely cited study, Zanna and Cooper (1974) utilized a misattribution of arousal paradigm to demonstrate that cognitive dissonance theory provided a more plausible theoretical account of attitude change effects, than self-perception theory. These researchers varied the nature of an external stimulus (e.g., a placebo) within a free-choice dissonance paradigm, and examined its impact on



subsequent attitude change. More specifically, the misattribution cue—a benign placebo—was given a side effect label of either being known to cause anxiety, relaxation, or no label was given (control). Participants were then either given a choice or assigned to write a counter-attitudinal essay using a dual experiments ploy. Counter to self-perception theory, Zanna et al. found that in the anxious label condition, participants in both the high and low choice conditions failed to exhibit attitude change, whereas in the control condition attitude change was moderated by choice consistent with cognitive dissonance theory. Moreover, alternative dissonance paradigms (e.g., hypocrisy paradigm; Fried & Aronson, 1995) have employed misattribution of arousal models which further underscore the utility of this framework.

### ***Misattribution and Threat***

One under-examined reduction strategy within the threat literature is offered by the misattribution of arousal paradigm. Of the 69 articles, dissertations, theses, and unpublished manuscripts uncovered by Jones and Stangor (2003), relatively few studies have attempted to examine the utility of misattribution processes in moderating threat outcomes. Three studies are of particular importance with regards to examining the potential impact of misattribution processes on stereotype threat which include O'Brien and Crandall (2003), Brown and Josephs (1999), and Stone and colleagues (1999). Each study has examined aspects of this process either directly or indirectly yielding relatively mixed results.

For instance, O'Brien and Crandall (2003) examined the influence of test

characterization and task difficulty on the performance of men and women. In their experiment, participants completed both an easy and a difficult math task after being informed that an upcoming task was either sensitive or insensitive to gender differences. As predicted, threat lowered performance of both groups on the difficult task and improved their performance on the easy task when compared to the control conditions. Similarly, women in the gender differences condition experienced performance deficits on the difficult tasks and performance boosts on the easy task. The performance of men was not influenced by this manipulation. These findings were taken as evidence that arousal was a potential mediator of threat effects.

Although provocative, this study still leaves open two important questions regarding the utility of a misattribution paradigm within a stereotype threat context. First, the absence of an arousal measure and the failure to conduct a formal mediational test of this potential mechanism makes it difficult to establish a definitive link between threat, arousal, and performance. Second, since a formal manipulation of misattribution was not present, it remains unclear whether such processes could provide any benefit to stigmatized group members under threatening conditions.

Research by Brown and Josephs (1999; Study 2) utilized misattribution processes by providing women and men with an external handicap—prior to completing a math task—that they could presumably misattribute a potential failure to. According to their logic, the burden of confirming the stereotype that “women don’t do well at math” would be greatly reduced for female participants since they would now be able to clearly attribute a poor performance to the external handicap. As expected, women given this

excuse performed equally as well as their male counterparts. However, amongst those not presented with an external handicap, the typical gender difference pattern emerged.

Although this study was taken as evidence that misattribution processes can moderate the effects of stereotype threat, it did not involve the misattribution of arousal which is characteristic of classic misattribution paradigms.

Stone and colleagues (1999; Study 2) opted to manipulate arousal in the classic sense by examining the impact of both framing techniques and misattribution processes on sports performance. In their study, high and low athletically engaged European Americans completed a golf task after being presented with a frame describing the upcoming task as either examining “natural athletic ability” or “the psychological factors involved in general sports performance”. The authors reasoned that misattribution processes would moderate the effect of threat, so they varied whether participants were (or were not) presented with a plausible external attribution for any anxiety. Using a clever manipulation, half of the participants were informed—in a letter ostensibly written by the psychology department—that recent building renovations had led some participants to feel “tense and uneasy” (p. 1220). Participants were further told that they would be asked to report on their lab experience at the conclusion of the experiment. As predicted, engaged participants given the natural ability frame performed worse when not buffered by the misattribution cue, when compared to engaged participants in all other conditions. However, no performance differences emerged for the disengaged participants.

It should be noted that there are at least three potential problems with the

misattribution manipulation in this experiment. First, checks of the misattribution manipulation gleaned from a combined measure of the perceived impacts of the lights, temperature, and noise in the experimental context revealed that high misattribution participants reported that the environmental factors had significantly less impact on their performance than did low misattribution participants. The authors reasoned that participants informed of the misattribution cue must have monitored the effect of the room on their performance and surmised that the renovations were not particularly problematic when compared to their low misattribution counterparts. This rationale suggests that the cue served as a cognitive disruption rather than operating via reducing anxiety as would be predicted by classic misattribution of arousal paradigms.

Second, and perhaps even more problematic for a misattribution of arousal interpretation, the experimental manipulations failed to produce any significant effects on the anxiety measures assessed in the experiment—Spielberger state anxiety inventory and the competitive state anxiety inventory; both measures generally demonstrate adequate levels of reliability—aside to significant time effect. Thus, neither threat, nor the misattribution cue, was linked to anxiety suggesting that a misattribution of arousal interpretation seemed less tenable for these findings.

Third, although the misattribution cue produced differences on the manipulation check (albeit in the opposite direction of what would be predicted by a misattribution of arousal interpretation), it is quite possible that this manipulation was not particularly strong or that these measures were insensitive. Perhaps, the combination of a more powerful manipulation and more sensitive measures would have been successful in both

producing and detecting the predicted fluctuations in arousal.

In sum, it remains unclear whether misattribution of arousal processes can be used as a viable means of removing stereotype threat based upon the mixed results presented above. And with the paucity of research devoted to examining this strategy, it appears that there are more questions than answers with regards to the usefulness of this approach. Therefore, I will further explore the potential utility of this reduction strategy for women in the present research.

### **Perceptions of “Threat” vs. “Challenge”**

Blascovich et al. (2001b) have linked stereotype threat to increased physiological activity in African Americans—manifest in heightened arterial blood pressure. Such heightened hemodynamic activity has been linked to maladaptive behavioral dispositions (e.g., John Henryism—defined as adapting to stressful situations via exerting increased effort even in the face of insurmountable obstacles and prevalent amongst low SES African Americans; James, 1994; James, Hartnett, & Kalsbeek, 1983), strong activation of the sympathetic nervous system (i.e., chronic high blood pressure and increased heart rate) and poor health outcomes such as hypertension and cardiovascular disease (Dressler, Bindon, & Neggers, 1998).

One interesting parallel to this line of research within the threat literature derives from the work of Blascovich and his colleagues (2001a; 2002) on their bio-psychosocial model of threat. More specifically, this research is devoted to the reactions of both the stigmatized and non-stigmatized to various motivated performance contexts (e.g.,

standardized testing, competitive tasks, and negotiations). The authors purport that the perceptions of stigmatized individuals in certain contexts (e.g., standardized exams) are associated with specific motivational states. Furthermore, they posit that individuals under such situational demands may evaluate a given task as either threatening or challenging.

Threat evaluations "...are characterized by the perception that the situational demands are 'outweighing' one's personal resources," whereas challenge evaluations are characterized by perceptions of one's personal resources approaching or exceeding the task demands (p. 254). In the former motivational state, individuals are presumed to make situational evaluations and maintain perceptions of their current context as consistent with "...danger, uncertainty, and required effort..." (p. 254). These evaluations appear to be rather consistent with the perceptions of stereotyped individuals under stereotype threat. In the latter motivational state, individuals may view current and future tasks through the lens of an opportunity to show off one's task relevant knowledge and abilities. Although both motivational states are associated with performance outcomes, threat evaluations are believed to reduce task performance, whereas challenge evaluations are believed to foster more positive performances. Therefore, both threat and challenge evaluations not only have implications for task perceptions, but they also have a profound impact on task relevant outcomes.

### ***Psychophysical Measurement of Threat and Challenge Motivational States***

Blascovich and his colleagues (2001a) have linked precise psychophysiological

reactivity to the exhibition of threat and challenge motivational states. For instance, they have linked challenge states to specific adrenal activation—that is, the “sympathetic-adrenal-medullary axis”—which enhances cardiac functioning and decreases resistance in the vascular system. Challenge motivational states are also associated with specific cardiovascular responses, such as significant increases in cardiac output (CO—the amount of blood being pumped by the heart in liters) and left-ventricular contractility (VC—indexed by a decrease in pre-ejection period), and accompanying decreases in total peripheral resistance (TPR—overall vasoconstriction occurring in the periphery).

Conversely, threat states have been linked to further adrenal activity (i.e., of the pituitary-adrenal-cortical axis) that is known to prevent decreases in resistance to the vascular system. In addition, threat motivational states are associated with cardiovascular reactivity manifest in the stabilization of CO and TPR, accompanied by increases in VC (Mendes et al., 2002). Such psychophysiological measurements are usually gauged by continuous blood pressure readings employed during an interaction, resting, or performance period.

### ***Empirical Support for “Threat” vs. “Challenge” Perceptions and Their Impact on Task Performance***

In a series of studies examining the stigma-threat hypothesis, Blascovich and colleagues (2001a) have demonstrated that when perceivers are confronted and paired with a confederate that bears a physical stigma (e.g., facial birth mark, race), this interaction has a profound impact upon their perceptions, psychophysiological functioning, and performance—on both cooperative and interdependent tasks. For

instance, Blascovich et al. varied the presence of a physical stigma (a facial birthmark) on a confederate and examined its impact on both verbal delivery and task performance (a cooperative word-finding task). Psychophysiological measures of VC, CO, and TPR were recorded as well as self-reported threat vs. challenge perceptions. As expected, on both tasks participants confronted with a stigmatized confederate exhibited reactivity consistent with threat—that is, significant increases in VC and TPR accompanied by slight, but non-significant, fluctuations in CO. Participants confronted with a non-stigmatized confederate exhibited challenge reactivity—that is, significant increases in VC and CO, with comparative decreases in TPR. Self-report measures yielded similar outcomes, with participants reporting having exerted significantly more effort and having perceived the task as more competitive when confronted with a stigmatized partner as opposed to being paired with a non-stigmatized partner. In addition, those paired with a stigmatized confederate performed significantly more poorly on the performance task. Thus, not only was threat reactivity (as evidenced by both physiological reactivity and self-reports) a function of the stigma associated with a potential partner, but it had a profound (negative) impact upon subsequent performance when paired with a stigmatized confederate.

Subsequent studies by Mendes and colleagues (2002) have replicated these effects while varying the nature of the stigma (e.g., Race, SES) associated with a potential partner. The general finding regarding the stigma-threat hypothesis is that when paired with a stigmatized partner, threat reactivity is triggered and reduced task performance occurs. However, when paired with a confederate free of physical, social, or socio-



economic stigma, challenge reactivity is triggered and performance is enhanced.

Blascovich, Mendes, Hunter, & Salomon (1999) have also examined the exhibition of threat vs. challenge reactivity on performance in the area of social facilitation. These researchers argue that conditions present in a social facilitation context may not only evoke evaluations of threat vs. challenge, but may also have a profound impact on task performance. For instance, Blascovich et al. hypothesized that performing in front of others would increase the likelihood that an individual would perceive his or her performance as goal relevant and, in turn, increase the likelihood that perceptions of threat vs. challenge would be evoked. More specifically, these researchers posited that performance on well learned tasks—believed to enhance performance when in the presence of others according to social facilitation theory (for a review, see Zajonc, 1965)—would lead to the exhibition of “challenge” evaluations and the typical psychophysiological patterns associated with this motivational state. In contrast, when performing unlearned tasks—assumed to inhibit performance according to social facilitation theory—these researchers maintained that threat evaluations would be exhibited. This is precisely what they found on both the performance and psychophysiological measures of threat vs. challenge.

### ***Integrating “Threat” vs. “Challenge” Perceptions into the Stereotype Threat Paradigm***

It seems apparent that the motivational states of threat and challenge in Blascovich et al.’s bio-psychosocial model do bear some resemblance to the phenomenon of stereotype threat in several ways. First, although the research regarding the mediation

of stereotype threat effects remains unclear (Jones & Stangor, 2003; Wheeler & Petty, 2001; Smith, 2004), there is some evidence that threat is linked to anxiety (Osborne, 2001; Walters, Shepperd, & Brown, 2003; Spencer et al., 1999, Study 3) and increases in arterial blood pressure (Blascovich, Spencer, Quinn, & Steele, 2001b).<sup>5</sup> And although anxiety has been measured via self-report in most threat studies, the assessment of anxiety via physiological arousal measures (e.g., palmar sweating, increased heart rate) is not uncommon (Smith, 2004). Moreover, the physiological reactivity that is considered to be characteristic of threat responses (e.g., the prevention of decreased resistance in the vascular system) appears to bear some resemblance to those present under stereotype threat (e.g., increased state anxiety—often operationalized in the psychological literature as increased arterial blood pressure) (Smith, 2004).

Second, the perceptions of individuals in threat motivational states (i.e., perceptions of danger, uncertainty, required effort, situational demands that are in excess of personal resources), seem to parallel those presumed to operate in individuals experiencing stereotype threat (e.g., lowered expectations, withdrawal of effort, cognitive interference, anxiety) quite well. Similarly, challenge perceptions (i.e., personal resources exceeding task demands) appear to be consistent with the evaluations of those for whom the impact of threat has been removed. Such compatibility would indicate that perceptions of threat and challenge would dovetail quite well with the current formulations of stereotype threat theory.

Third, it seems quite appropriate to consider that the experiences related to interactions with stigmatized others might bear some relationship to the perceptions that

threatened individuals maintain toward those they perceive as maintaining a negative group-based stereotype that they are at risk of confirming. For instance, perceptions consistent with challenge evaluations may be maintained by those for whom stereotype threat has been removed, whereas those under threat may maintain perceptions consistent with threat evaluations. Similarly, stigmatized individuals under (or not under) threat should experience physiological reactivity consistent with their situational evaluations. From this perspective, just as the prospect of being confronted with a stigmatized individual in a potential interaction can evoke perceptions of threat, I posit that the prospect of confirming a negative group-based stereotype “...in their own eyes, the eyes of others, or both at the same time” (Aronson et al., 1998a, p. 86) can also prime such perceptions and cardiovascular reactivity.

And finally, if perceptions of “threat” vs. “challenge” do map onto the phenomenological and physiological responses of participants both under stereotype threat and removal conditions, then it follows that it would be advantageous to use removal strategies that incorporate features of a classic misattribution paradigm—that is, leading participants to (mis)attribute any arousal away from a performance task and toward a salient external stimulus. If successful, the incorporation of misattribution strategies and threat vs. challenge motivational states into the current formulation of stereotype threat theory would help to tie together three literature bases normally examined in isolation. Such a strategy can reduce the duplication of efforts and help us further understand the underlying mechanisms of these phenomena more clearly.

## Goals and Research Hypotheses

The present research had five overarching goals that produced five corresponding research hypotheses. These goals and hypotheses are presented below:

Goal 1: In Study 1, I attempted to replicate the standard stereotype threat effect on performance in a sample of collegiate women and men. Consistent with stereotype threat theory, I predicted that women would perform more poorly on a math task than men when presented with a gender differences instructional set, whereas there would be no gender differences when participants were provided with a gender fair instructional set. More specifically I predicted that:

*Hypothesis 1.* A significant 2 (gender: male, female) X 2 (instructional set: *gender differences* instruction; *gender fair* instruction) interaction would emerge such that women would perform worse than men after receiving a *gender differences* instructional set. However, no gender differences were predicted for women and men who received the *gender fair* instruction.

Goal 2: Given the assumption that stereotype threat would differentially impact women and men on performance based vs. non-performance based tasks, Study 2 examined the effects of stereotype threat on two novel non-performance based dependent measures—task choice and strength of choice. And given that prior research has rarely examined the impact of threat on non-performance based measures, Study 2 was an exploratory effort with the goal of further understanding the generalization of stereotype threat to the types of choices that stigmatized individuals make and the strength associated with these choices.

Once again, I utilized a sample collegiate women and men and recorded both their task choice and strength of preference for these choices after manipulating the

instructional set. I then examined the applicability of two competing hypotheses to these data—one which was rooted in the disidentification tenet of stereotype threat theory and a second that was rooted in research devoted to the achievements of college bound women. The former hypothesis assumed that women would be more likely disassociate from tasks in a stereotyped domain, such as math, by actively avoiding them. This avoidance behavior was presumed to be the resultant of consistent poor performance in the stereotyped domain that was a function of stereotype threat. Moreover, threatened women were presumed to maintain a low level of preference for such tasks when given the opportunity to partake in them. However, the task choices and task preferences of men were presumed to be uninfluenced by this situational pressure, given that these individuals were not at risk of confirming a negative group-based stereotype.

A competing alternative hypothesis predicted that whereas stereotype threat was hypothesized to lower the task performance of women, this phenomenon was not expected to impact their task choices in a similar manner. Contrary to the disidentification hypothesis, the *women's achievement hypothesis* predicted that women would actually be likely to select a task within a stereotyped domain when they were threatened (as opposed to when they were not threatened). Given that college bound women are likely to have experienced prior success in a stigmatized domain, such as math, they may be likely to approach such tasks when confronted with a negative group based stereotype regarding the performance of women. These success experiences may also lead these individuals to maintain a higher degree of preference for these tasks when under threatening conditions. However, similar to the disidentification hypothesis, the

*women's achievement hypothesis* assumes that the selection process and preferences of men would not be influenced by the presence (or absence) of threatening conditions given (1) their potential success in the respective domain and (2) given the absence of any negative group-based stereotypes regarding men. This rationale led me to generate the following series of competing hypotheses:

Hypothesis 2a: If the disidentification hypothesis was more applicable to these data, then in Study 2 a significant 2 (gender) X 2 (instructional set) interaction would emerge on both the task choice and strength of choice measures—only amongst those who chose the math alternative on the latter measure—such that women in the *gender differences* condition would be *less* likely to select a math task over a proofreading task, and would display a weaker preference for the selected alternative when compared to women in the *gender fair* condition. However, the choice behavior and task preference of men was expected to remain consistent across instructional sets.

Hypothesis 2b: If the women's achievement hypothesis was more applicable to these data, then in Study 2 a significant 2 (gender) X 2 (instructional set) interaction would emerge on both the task choice and strength of choice measures—only for participants who chose the math alternative on the latter measure—such that women in the *gender differences* condition would be *more* likely to select a math task over a proofreading task, and would prefer the selected task more, when compared to women in the *gender fair* condition. However, the choice behavior of men was expected to remain consistent across instructional sets.

Goal 3: In Study 3, I examined whether the effect of stereotype threat could be alleviated via either misattribution or self-affirmation processes in a sample of stereotype threatened collegiate women. I predicted that both self-affirmations and misattribution processes could be used to reduce the effects of stereotype threat on the task performance of women. I further posited that a combination of multiple removal strategies might produce an additive effect that would further insulate these individuals from the impact of threat. Therefore, I hypothesized that by allowing women to self-affirm, to misattribute arousal, or to engage in both of these removal strategies prior to engaging in a math task,

the impact of stereotype threat on task performance would be reduced. These predictions led to the following hypothesis:

Hypothesis 3: A significant 2 (misattribution opportunity: present, absent) X 2 (self-affirmation opportunity: present, absent) interaction would emerge for women who received the *gender differences* instruction such that when given an opportunity to either affirm the self or to misattribute arousal, task performance for these participants should markedly increase when compared to participants who did not receive either removal strategy. An additive effect was predicted for women that engaged in both removal activities when compared to their counterparts who were not given a removal opportunity.

Goal 4: In all three studies I had the goal of assessing a variety of mediators that might play an integral role in the threat-performance relationship including motivation, expectancies, threat vs. challenge perceptions, gender stereotype and self-doubt activation, self-esteem, task confidence, and state anxiety. Several of these variables were measured implicitly via a word fragment completion task (e.g., gender and self-doubt activation), whereas several additional variables were measured explicitly via self-report (e.g., motivation, expectancies, threat vs. challenge perceptions) prior to completing the dependent measure of interest. I employed this strategy to maximize my potential to detect the mechanisms that may underlie stereotype threat, while circumventing some of the methodological shortcomings encountered in other studies. The remaining items were measured explicitly (e.g., anxiety, self-esteem) after completing the dependent measure. Task confidence was only measured in Studies 1 and 3 during the performance task—after each successive item. Given that both empirical and anecdotal evidence exists regarding the influence of stereotypes on women's math performance (e.g., Spencer et al., 1999), I expected women to experience more stereotype threat than men in contexts where they recognized that their math ability may be assessed.

In contexts where women's math ability was less likely to be evaluated, I expected such concerns to be reduced. Since no negative stereotype exists regarding the math performance of men, these individuals should not perceive either context as threatening. Therefore, similar to Brown and Josephs (1999), Steele and Aronson (1995), McKay and colleagues (2002), and Blascovich et al. (2001a), I posited that women would be more likely to experience threat in math-related contexts than men. More specifically, I posited:

Hypothesis 4: In Studies 1-3, women given a *gender differences* instructional set would experience greater levels of stereotype threat than their male counterparts, whereas no gender differences would emerge amongst those given a *gender fair* instructional set. The predicted gender differences were expected to occur on measures of gender and self-doubt activation, state anxiety, and threat vs. challenge perceptions. Although assessed, the predicted gender differences were not anticipated to emerge on measures of motivation, expectancies, self-esteem, reaction time, and task confidence given the findings produced in prior research with respect to women (Jones & Stangor, 2003).

Goal 5: Finally, I examined whether individual differences in performance motivation would moderate the effects of threat on both performance and task choice by including an exploratory achievement motivation measure (Midgley et al., 1998) in Studies 1 and 2. Although this possibility has been recently proposed (Smith, 2004), the manner in which such motivations would impact stereotype threat outcomes remains unclear. Therefore, I did not make any explicit predictions regarding how individual differences in this motive would impact the performance and task choices of the participants in the studies reported herein.



## STUDY 1

### Method

#### Design and Participants

Study 1 used a 2 (gender: male, female) X 2 (instructional set: *gender differences*; *gender fair*) between participants factorial design with math performance serving as the primary dependent measure. I recruited 101 University of Maryland students to participate in this experiment in exchange for course credit during the spring semester of 2003. All participants indicated their level of identification with the domain of mathematics based upon their scores on a domain identification measure (Smith & White, 2001) which was administered in a mass testing session prior to the experiment.<sup>6</sup> The data from 8 participants were excluded from the analysis because they failed the manipulation check (as described below). This left 93 participants (63 female and 30 male) who were randomly assigned to one of two experimental conditions. The ethnic background of the sample included 50 European Americans, 20 people who self-identified as 'Other', 13 African Americans, and 10 Asian Americans.<sup>7</sup> The four most prevalent majors were Psychology (32%), Undecided (12%), Education (6%) and Letters & Sciences (5%). The mean verbal and quantitative SAT score (VSAT; QSAT), College GPA, and High School GPA were 596, 618, 3.3, and 3.7, respectively. The mean math domain identification score for the sample was 3.0.

## **Exclusion of QSAT Score as a Covariate**

Given the controversy within the threat literature surrounding the use of QSAT scores as a covariate in ANCOVA designs, I decided not to use this variable as a covariate in any analyses reported in this manuscript. The rationale behind this decision was twofold. First, given the potential for one's QSAT scores to be influenced by stereotype threat, there is a possibility that scores on this measure and threat manipulations are not mutually exclusive—a fundamental assumption of ANCOVA designs (for a review, see Wicherts, 2004). Second, using QSAT scores as covariate could produce spurious results in quasi-experimental stereotype threat studies and limit one's potential to attribute gender and racial differences on standardized tests to stereotype threat (Sackett, Hardison, & Cullen, 2004; Wicherts, 2004).

## **Procedure**

Several weeks prior to participating in this experiment, participants completed the domain identification measure and provided their QSAT and VSAT scores as part of a mass testing session. These indices were imbedded within a battery of measures, which were completed and simultaneously collected. By collecting this information prior to the experiment, I was able to measure the participants' level of identification with and ability within mathematics without heightening their sensitivity to the true nature of Study 1.

After the initial mass testing phase, participants were recruited to participate in the study (for a copy of the consent form, Please see Appendix A). All participants reported to the lab individually, where they were met by a male experimenter who provided them

with the cover story of the experiment. Participants were informed that they would be taking part in an experiment designed to assess the “the psychology of problem solving” and that would entail completing a set of tasks from “several general content areas.” The experimenter further noted that these tasks would be chosen at random and that he was unaware of which content areas the items would be sampled from.

Participants were also informed that the entire experiment would be conducted on computer and that they were expected to pay close attention to all on-screen directions since the experimenter would not be able to assist them while they completed the upcoming tasks. They were further informed to pay close attention to any video content since they would be asked to recall this information at a later point in the experiment. Participants were then told that they would be timed while completing these tasks and that if they did not have a watch, the experimenter would provide them with one. In addition, all participants were provided with a writing utensil and scrap paper.

At this point, all participants were led to a small laboratory room equipped with a computer, a set of speakers, and a folder labeled ‘Task 1’ placed to their right. Participants were then seated at the computer and it was reiterated that they should pay close attention to any video content presented on-screen. The experimenter repeated this instruction to maximize the likelihood that participants would retain the performance difference information embedded within the video content.

Participants were further told that the computer would inform them if they would need to refer to the Task 1 folder at any point in the experiment. They were instructed to

only refer to the folder if they were given explicit instructions to do so. The experimenter then asked if there were any further questions, initiated the program, and then exited the room.

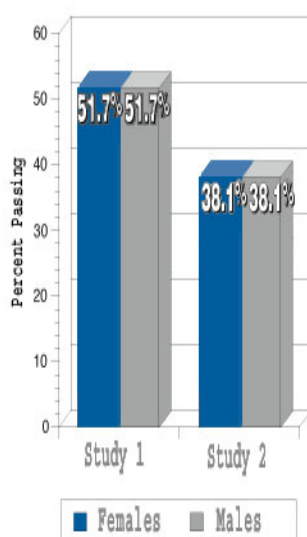
### ***Stereotype Threat Manipulation***

After the program was initiated, the computer randomly assigned participants to one of two instructional sets embedded within a video. Similar to a manipulation used in prior studies (Blascovich et al., 2001b; Spencer et al., 1999), participants in the *gender differences* condition were presented with a video depicting a male, named ‘Patrick Smith’, who was presumably a researcher from the Psychology Department. The male character informed participants that his research program was investigating why there are gender differences on standardized exams and that it was his goal to try to further understand why males tend to outperform females on the upcoming problem solving task. Participants in the *gender fair* condition were introduced to the same male character. However, these participants were told that he was conducting a collaborative research effort with the Women’s Studies Department and several neighboring universities in an effort to develop a gender fair test. The male character then mentioned that his research has demonstrated that males and females perform equally well on the upcoming task.

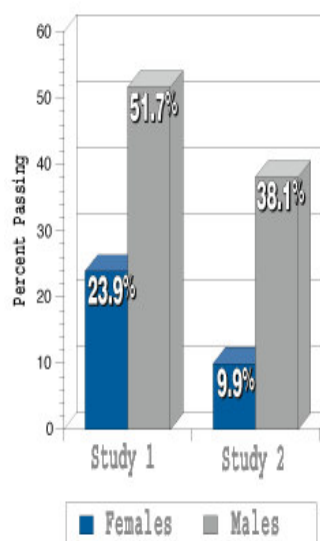
In an effort to strengthen this manipulation, one of two charts was embedded within the video content (see Figure 1). Each chart graphically reinforced the gender differences information described by the male character and remained on the screen for 10 seconds in each condition. Both charts were entitled, “Performance of Collegiate

Males and Females Across Two Preliminary Studies” and the University’s insignia was imprinted on these graphics to maximize their perceived authenticity. I reasoned that by informing participants in the *gender differences* condition about the nature of these differences both orally and graphically, the possibility of evoking threat in this condition would be heightened, whereas it would be minimized in the *gender fair* condition.

Performance of Collegiate Males and Females  
Across Two Preliminary Studies



Performance of Collegiate Males and Females  
Across Two Preliminary Studies



*Figure 1.* Charts depicted in the video to reinforce the performance differences of men and women on the upcoming task in the *gender fair* and *gender differences* conditions, respectively.

The male character then reappeared in the video content and informed all participants that they would be completing a math task. He further noted that the task was composed of 10 items and that they would have a 15-minute time limit. The communicator concluded by reminding participants to pay attention to all directions.

### ***Gender and Self-doubt Activation***

After viewing the video, the computer informed participants that they would be completing an initial problem solving task prior to completing the main task. All participants were then instructed to open the folder labeled Task 1 and to complete its contents.

Task 1 was composed of a modified 54-item word fragment completion measure that has been utilized in prior research (Brown & Josephs, 1999, Study 1; Steele & Aronson, 1995, Study 3) to detect the implicit activation of stereotype threat and self-doubt (Please see Appendix B for a listing of the critical items).<sup>8</sup> A total of 16 items were used to assess gender and self-doubt activation—which are described in more detail in the sections that follow—whereas the remaining 38 items served as fillers. For both measures, I assumed that the critical word fragments would be completed in a manner consistent with both gender and self-doubt laden associations by participants experiencing stereotype threat. In addition, activation was deliberately assessed after the experimental manipulation, but prior to measuring performance, to allow for formal mediational tests to be conducted.

*Gender activation.* Nine of the 16 critical word fragments items on the word-fragment completion measure were designed to detect the subtle activation of gender-related constructs or images associated with women. The gender activation measure was a modified version of Steele and Aronson's (1995; Study 3) stereotype threat activation measure. Their measure contained the following word fragments and target words: \_ \_

CE (Race); LA \_\_ (Lazy); \_\_ ACK (Black); \_\_ OR (Poor); CL\_S\_ (Class); BR \_ \_ \_ \_  
 \_ (Brother); \_ \_ \_ TE (White); MI \_ \_ \_ \_ \_ (Minority); WEL \_ \_ \_ \_ (Welfare); CO \_ \_  
 \_ (Color); and TO \_ \_ \_ (Token). The fragment “\_\_ \_\_ CE”, for instance, could  
 be completed in several ways including “race”, “mice”, “rice”, or “vice”. For each word  
 fragment, these authors allowed at least 2 letter spaces to be vacant. They reasoned that  
 this strategy would increase the number of possible ways that each fragment could be  
 completed, while reducing the possibility that ceiling effects would emerge.

To assess gender activation, a modified version of Steele & Aronson’s (1995)  
 measure was used to detect the activation of gender-related constructs. Several of the  
 target words offered by Steele et al. were substituted to create the following completions:  
 \_ \_ \_ \_ER (Gender); MA\_\_ (Math); \_ \_ \_ AN (Woman); \_ E \_ \_ LE (Female); SI \_ \_ \_ \_  
 (Sister); \_ \_ LE (Male); TO \_ \_ \_ (Token); \_ \_ \_ MAL (Normal); \_ \_ \_ \_AGE (Average).  
 Similarly, the fragment “\_ \_ \_ \_ER” could be completed in several ways including “gender”,  
 “tender”, “fender”, or “Denver”. In addition, at least 2 letter spaces were left vacant for  
 each completion to increase the number of potential completions for each alternative.  
 Scores on this measure were assessed by assigning a point for each word fragment that  
 was completed in a manner consistent with the target word. Consistent completions were  
 then summed to create an overall gender activation score with higher scores indicating  
 increased activation.

*Self-doubt activation.* A 7-item measure that was identical to the one employed  
 by Brown and Josephs (1999, Study 1) was used to assess self-doubt activation. These

items were presumed to tap into the implicit activation of self-doubt and included the following completions: LO \_ \_ \_ (Loser); DU \_ \_ (Dumb); SHA \_ \_ (Shame); \_ \_ \_ ERIOR (Inferior); FL \_ \_ \_ (Flunk); \_ AR \_ (Hard); W \_ \_ K (Weak). Scores on this measure were tabulated in a manner consistent with the gender activation measure described above. After completing the word fragments embedded in Task 1, participants then placed the measure back into the folder and indicated that they had completed the task via computer. The remainder of the experiment was completed on the computer.

*Motivation.* Participants' task motivation was then assessed by a single item that asked them to indicate the extent to which they were motivated to do well on the main task. This measure was assessed on a 7-point scale ranging from 1 (not at all motivated) to 7 (extremely motivated).

*Expectancies (overall).* Participants were then given a single item which asked them to indicate how well they believed they would do on the performance task overall. This measure was rated on a 7-point scale ranging from 1 (very poorly) to 7 (very well).

*Threat vs. Challenge Perceptions.* Participants were then presented with 3-items devoted to assessing the extent to which the performance task was perceived as a potentially threatening (or less challenging) situation. Participants were asked to indicate their agreement or disagreement with the following statements: "I believe that the upcoming task will be stressful," "I plan to exert maximum effort on the upcoming task," and "I believe I will do well on the upcoming task,"—all rated on a 9-point scale ranging from 1 (strongly disagree) to 9 (strongly agree) with the first item reverse coded. These



items were scored and combined to create a measure of perceived challenge. I predicted that whereas women would respond more negatively to these items when under threat (threat pattern), their responses would be more positive when not under threat (challenge pattern).<sup>9</sup> I further posited that men's threat vs. challenge perceptions would be uninfluenced by the instructional set manipulation.

*Math Performance.* After completing the threat vs. challenge measure, participants were given 15 minutes to complete a 10-item math task (see Appendix C). All items were obtained from a test bank of prior GRE exams (Educational Testing Service, 1994) and only items that were difficult, but not impossible, for students to complete—e.g., only items that 50% or less of the testing population completed correctly—were used. All of these items contained either four or five response options and were geometry-based. Geometry items were chosen because prior research has shown that items in this area typically present the most difficulty for women—particularly because superior spatial skills are required (Liben, 1978; Stangor & Sechrist, 1998).

*Perceived task confidence (per item).* After responding to each item on the math task, participants provided a corresponding task confidence rating on a 9-point scale ranging from 1 (not at all confident) to 9 (extremely confident).

*Reaction time.* Reaction time measures were also recorded in milliseconds (ms) after the completion of each item on the math task.

### ***Post-test Measures***

*State anxiety.* After completing the math task, participants completed the state portion of the Spielberger state anxiety inventory (Spielberger, 1972). This 20-item measure—anchored on a 4-point scale ranging from 1 (not at all) to 4 (very much so)—was designed to identify the extent to which participants exhibited anxiety when completing the math task. A typical item from this measure was, “I feel anxious” and this inventory has been shown to be both reliable and generalizable across a multitude of contexts (Spielberger & Diaz-Guerrero, 1976).

*Self-esteem.* Participants then completed the 10-item Rosenberg Self-Esteem Scale (RSES) (Rosenberg, 1965; 1989) which examined the extent to which these individuals maintained a positive overall self-view. This measure was rated on a 4-point scale ranging from 1 (strongly disagree) to 4 (strongly agree) and has been demonstrated to exhibit both internal consistency (e.g.,  $\alpha = .88$ ) and stability over time (e.g., test-retest reliability = .82) (Fleming & Courtney, 1984). A typical item on this measure was “I am satisfied with myself.”

*Demographic information.* After the completing the RSES, participants were asked to provide demographic information (e.g., ethnicity) in a battery of items.

*Manipulation check.* Participants then completed a single item designed to examine the effectiveness of the threat manipulation. All participants were asked to recall what they were told by the person in the video stimulus regarding the nature of the problem solving tasks that they would be completing. This item had four response

alternatives including (1) non-gender biased, (2) had found gender differences; (3) no such information was given to me, or (4) I do not remember.

*Achievement motivation.* Since goal orientations have been linked to anxiety and performance outcomes in the achievement motivation literature (Dweck, 1986), and since reviews of the threat literature have suggested that such motives may help to illuminate the mechanisms that may underlie this phenomenon (Smith, 2004), an exploratory achievement motivation measure (Midgley et al., 1998) was given to participants. This measure consisted of 12 items (see Appendix D), six devoted to measuring performance-approach motivation (e.g., “An important reason why I do my school work is because I want to get better at it”) and the remaining six items measuring performance-avoidance motivation (e.g., “I want to get out of having to do school work”). Each item was rated on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). After completing this measure, participants were probed for suspicion, debriefed, thanked, and dismissed.

## **Results**

### **Manipulation Check**

To verify the effectiveness of the instructional set manipulation, participants were asked to recall the nature of the performance differences that had been described to them by the communicator regarding an upcoming task. After removing eight participants that failed the manipulation check—seven in the *gender differences* condition and one in the

*gender fair* condition—I found that participants had relatively little difficulty recalling this information across conditions,  $\chi^2(3, n = 92) = 77.78, p < .01$ . In the *gender fair* condition, all but a seven participants (85%) were able to correctly recall that the main task had been found to be gender neutral in the past. Of the remaining participants in this condition, five (11%) did not recall the nature of the gender relevant information, whereas only two (4%) did not recall ever being given this information.

In the *gender differences* condition, nearly all of the participants (83%) were able to correctly identify the gender differences information. Only 8 participants (17%) in this condition failed to correctly identify the nature of the gender differences—four did not recall the nature of the gender differences, whereas the remaining 4 failed to recall being given such information. These data suggest that the instructional set manipulation was successful in allowing participants to correctly decipher the video feedback in a manner consistent with the gender differences information they received.

### **Task Performance**

Due to a disproportionate number of participants (73%) answering item 8 incorrectly in the present study, this item was not included in the tabulation of performance scores. The remaining nine items appeared to be sufficiently challenging, but not impossible, with the percentage of participants in Study 1 getting a particular item correct ranging from 38% to 65%. I then subjected this 9-item measure to a 2 (gender) X 2 (instructional set) ANOVA.<sup>10</sup>

Although the main effect of instructional set was not statistically reliable,  $F(1, 89)$

= 2.01,  $p = .16$ , a marginally significant effect of gender on performance did emerge,  $F(1, 89) = 3.48, p = .065$ . Overall, women ( $M = 4.19, SD = 2.23$ ) performed worse than men ( $M = 5.17, SD = 3.12$ ) on the math task. As predicted, this main effect was qualified by a significant gender X instructional set interaction,  $F(1, 89) = 5.62, p = .02$ , that was consistent with Hypothesis 1. Means, standard deviations, and sample sizes per cell for this interaction are presented in Table 1.

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		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
Female	3.93 <sub>b</sub>	2.09	33	4.47	2.37	30
Male	6.29 <sub>a</sub>	3.00	14	4.19 <sub>b</sub>	2.17	16

*Note.* Means sharing different subscripts are significantly different at the .05 alpha level. Higher means indicate better performance.

I conducted t-tests to examine the simple effects of this interaction. Consistent with Hypothesis 1, in the *gender differences* condition, women ( $M = 3.93, SD = 2.09$ ) performed significantly more poorly on the task than men ( $M = 6.29, SD = 3.00$ ),  $t(45) = 3.10, p < .01$  (Cohen's  $D = -0.95$ ). However, in the *gender fair* condition, women ( $M = 4.47, SD = 2.37$ ) actually performed slightly better than their male counterparts ( $M = 4.19, SD = 2.17$ ) although this difference was not statistically reliable,  $t(44) = 0.39, p = .70$  (Cohen's  $D = 0.11$ ). Although women in the *gender differences* condition performed

worse than women in the *gender fair* condition, this difference was not statistically significant,  $t(61) = 0.96, p = .34$  (Cohen's  $D = -0.22$ ). In contrast, men in the *gender differences* condition performed significantly better than men in the *gender fair* condition,  $t(28) = 2.22, p = .04$  (Cohen's  $D = 0.84$ ).

*Reaction time (per item)*. I then tested whether the manipulation had an impact on the reaction times of participants. I applied a  $\log(10)$  transformation to normalize these data after deleting three participants from this analysis due to their failure to complete the performance measure within the allotted time frame. Response times per item were then averaged and subjected to a 2 X 2 ANOVA. This analysis only revealed a significant main effect of gender. On average, women ( $M = 4.43, SD = 0.24$ ) were significantly faster than men ( $M = 4.52, SD = 0.11$ ),  $F(1, 86) = 4.33, p = .04$ . No other significant main effects or interactions reached significance, all  $p$ 's  $> .38$ . This finding supported Hypothesis 4.

### **Potential Mediators: Implicit Measures**

*Gender word fragment completions*. A slight positive skew ( $skewness = 0.80$ ) appeared on this measure with the majority of the values clustered near its lower limit. Given the large proportion (55%) of participants who failed to complete any of the word completions in a gender-related manner, I applied a logarithmic transformation to these data in an effort to try to help normalize this distribution. However, after applying a linear transformation to these data, I did not find that this distribution was markedly different in shape than the previous one. Therefore, I subjected non-transformed scores

on this measure to a 2 X 2 ANOVA. A significant main effect of instructional set did emerge,  $F(1, 89) = 4.51, p = .04$ , indicating that across gender, those in the *gender fair* condition ( $M = 0.85, SD = 0.89$ ) completed significantly more gender-related word completions than those in the *gender differences* condition ( $M = 0.49, SD = 0.75$ ). This effect was unexpected and inconsistent with Hypothesis 4. No other significant effects or interactions were found, all  $p$ 's  $> .50$ .

*Self-doubt activation.* Similar to the gender activation measure, a positive skew ( $skewness = 1.37$ ) was evident in these data with 61% of the scores on this measure clustered at its lower limit. Once again, I attempted to normalize these data, but the resulting distribution was similar to that of the non-transformed data. Therefore, I subjected these non-transformed data to a 2 X 2 ANOVA. However, no significant main effects or interactions emerged, all  $p$ 's  $> .35$ . Once again, these results did not support Hypothesis 4.

### **Potential Mediators: Self-Report Measures**

*Task motivation.* Participants' responses on the task motivation measure were also subjected to a 2 X 2 ANOVA. As expected, no significant main effects or interactions were detected, all  $p$ 's  $> .36$ . These data suggest that sheer motivation does not explain the performance differences experienced by women and men in the *gender differences* condition. This finding was consistent with Hypothesis 4.

*Expectancies (overall).* Participants' overall performance expectations were subjected to a 2 X 2 ANOVA which revealed only a significant gender main effect,  $F(1,$

89) = 18.71,  $p < .01$ . Overall, women ( $M = 4.92$ ,  $SD = 1.05$ ) had significantly lower task expectations than men ( $M = 5.90$ ,  $SD = 0.89$ ). No other significant main effects or interactions emerged, all  $p$ 's  $> .58$ . This finding supported Hypothesis 4.

*Threat vs. challenge perceptions.* I examined participants' responses to the threat vs. challenge measure and subjected these data to a 2 X 2 ANOVA. Although, the measure was found to exhibit only a modest level of internal consistency,  $\alpha(93) = .52$ , a significant gender main effect did emerge,  $F(1, 89) = 6.05$ ,  $p = .02$ . I found that, overall, women ( $M = 6.51$ ,  $SD = 1.16$ ) viewed the math task as less of a challenge than men ( $M = 7.16$ ,  $SD = 1.11$ ). No other significant main effects or interactions emerged, all  $p$ 's  $> .43$ . This finding was inconsistent with Hypothesis 4.

*State Anxiety.* I subjected participants' scores on the Spielberger state anxiety inventory—which was found to be a reliable index,  $\alpha(93) = .95$ —to a 2 X 2 ANOVA. Only a significant main effect of gender emerged,  $F(1, 89) = 13.80$ ,  $p < .01$ . Women ( $M = 2.40$ ,  $SD = 0.67$ ) reported experiencing greater levels of anxiety than men ( $M = 1.88$ ,  $SD = 0.53$ ). No other significant main effects or interactions emerged, all  $p$ 's  $> .28$ . This finding did not support Hypothesis 4.

*Self-Esteem.* Participants' scores on the RSES—which was found to be internally consistent,  $\alpha(92) = .87$ —were also subjected to a 2 X 2 ANOVA. As predicted in Hypothesis 4, no significant main effects or interactions were detected, all  $p$ 's  $> .38$ .

*Task Confidence (per item).* I then averaged participants' task confidence ratings (per item) and subjected them to a 2 X 2 ANOVA—excluding ratings for item 8. Similar



to the pattern of results obtained on the performance measure, the main effect of instructional set was not statistically reliable,  $F(1, 89) = 2.63, p = .11$ . However, a significant gender main effect did emerge,  $F(1, 89) = 8.84, p < .01$ , which revealed that women ( $M = 5.34, SD = 1.37$ ) expressed significantly lower levels of task confidence than their male counterparts ( $M = 6.19, SD = 1.34$ ). This main effect was qualified by a significant gender X instructional set interaction,  $F(1, 89) = 4.77, p = .05$ . Table 2 presents the means, standard deviations, and cell sizes per condition on the task confidence measure as a function of gender and instructional set.

Table 2

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

Gender of the Participant	Instructional Set					
	<u>Gender Differences</u>			<u>Gender Fair</u>		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
Female	5.26 <sub>b</sub>	1.26	33	5.43	1.50	30
Male	6.80 <sub>a</sub>	0.95	14	5.67 <sub>b</sub>	1.44	16

*Note.* Means sharing different subscripts are significantly different at the .05 alpha level. Higher means indicate greater task confidence.

Once again, I conducted t-tests to examine the simple effects of this interaction. In the *gender differences* condition, women ( $M = 5.26, SD = 1.26$ ) expressed significantly lower levels of task confidence when compared to their male counterparts ( $M = 6.80, SD$

= 0.95),  $t(45) = 4.10, p < .01$ . However, in the *gender fair* condition, the task confidence levels expressed by women and men did not differ ( $M = 5.43, SD = 1.50; M = 5.67, SD = 1.44$ , respectively),  $t(44) = 0.52, p = .60$ . The task confidence scores posted by women in the *gender differences* and *gender fair* conditions also did not differ,  $t(61) = 0.49, p = .63$ , whereas the confidence ratings of men in *gender differences* condition were significantly greater than those of men in the *gender fair* condition,  $t(28) = 2.22, p = .04$ . This finding was both unexpected and inconsistent with hypothesis 4.

### **Mediational Analysis**

I used the Baron and Kenny (1986) approach to determine whether the gender X instructional set interaction effect on task performance was partially or fully mediated by task confidence—since this was the only potential mediating variable that produced a significant interaction effect.

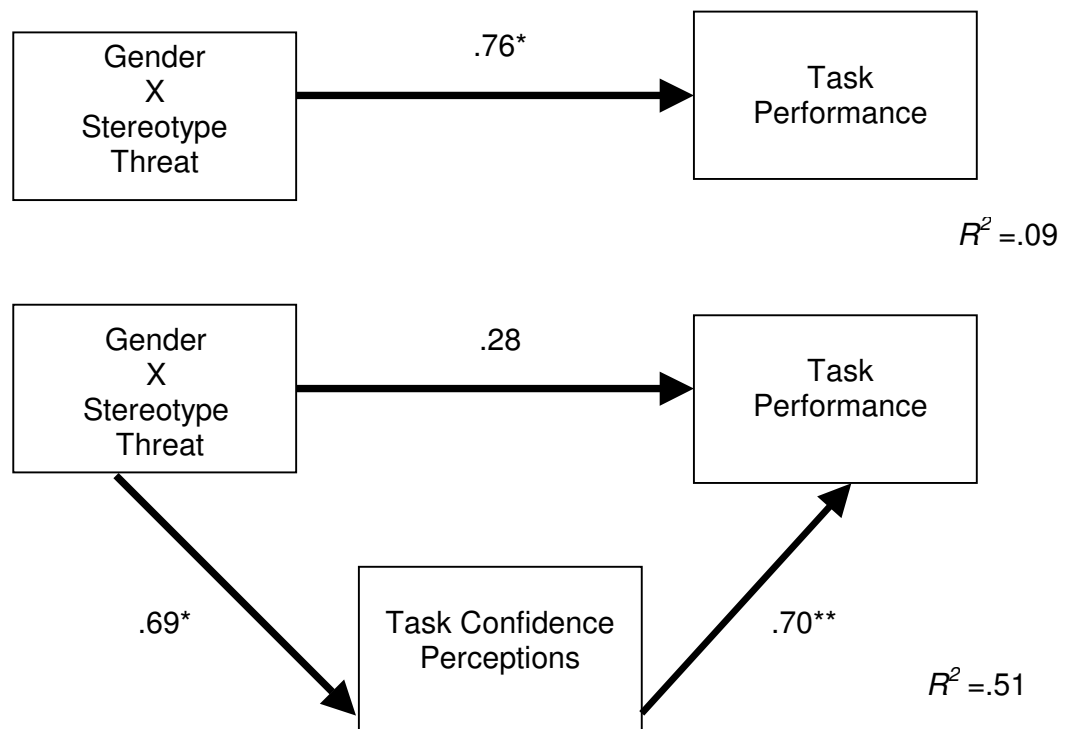
According to this statistical approach, four steps are necessary to demonstrate that the gender X instructional set interaction effect on task performance was mediated by task confidence perceptions. First, a significant relationship must be found between the gender X instructional set interaction and task performance. This stipulation was met—I found that there was a significant relationship between these variables. Not only was this stipulation met by both the overall model—including gender, instructional set, and the gender X instructional set interaction simultaneously in the model— $R = .31, F(3, 89) = 3.04, p = .03$ , but it was also met by the independent gender X instructional set interaction effect on task performance,  $\beta = 0.76, t(92) = 2.37, p = .02$ . Indeed, both approaches were

successful in predicting performance.

The second step in the mediational process consisted of demonstrating that the gender X instructional set interaction was correlated with task confidence. This step was achieved when I found that both the overall model (as described above) significantly predicted task confidence ratings,  $R = .37$ ,  $F(3, 89) = 4.60$ ,  $p = .01$ , as did the independent effect of the gender X instructional set interaction,  $\beta = 0.69$ ,  $t(92) = 2.18$ ,  $p = .03$ .

The third and fourth steps in the mediational process consisted of showing that task confidence predicted task performance and that after statistically controlling for this mediator, the significant gender X instructional set interaction effect on task performance was either reduced in magnitude (partial mediation) or removed (full mediation). These stipulations were tested by including task confidence, gender, instructional set, and the gender X instructional set interaction, respectively, in a regression equation predicting task performance. Although the overall model including task confidence scores did a better job of predicting task performance than the original model,  $R = .72$ ,  $F(4, 88) = 23.26$ ,  $p < .01$ —as evidenced by a significant change in the proportion of variance explained by the model,  $R^2_{\text{change}} = .42$ ,  $F_{\text{change}}(1, 88) = 76.22$ ,  $p < .01$ —closer inspection of the independent effects of these variables on task performance demonstrated that task confidence perceptions mediated the gender X instructional set-performance relationship. When task confidence scores were entered into the regression equation (first), the independent effect of this variable remained significant,  $\beta = 0.70$ ,  $t(92) = 8.73$ ,  $p < .01$ , whereas the previously significant gender X instructional set interaction on task

performance,  $\beta = 0.76$ ,  $t(92) = 2.37$ ,  $p = .02$ , was no longer statistically reliable,  $\beta = 0.28$ ,  $t(92) = 1.17$ ,  $p = .25$ . An additional Sobel test analysis, as suggested by Kenny, Kashy, and Bolger (1998), revealed a significant difference in the direct path from the gender X instructional set interaction to task performance after statistically controlling for task confidence perceptions,  $t(67) = 2.19$ ,  $p = .03$ . Therefore, the third and fourth tenets of the mediational process were met and the impact of the gender X instructional set interaction on task performance was fully mediated task by confidence perceptions. Figure 2 presents the mediation of the gender X stereotype threat interaction effect on task performance by task confidence perceptions.



*Figure 2.* Mediation of the gender X stereotype threat interaction effect on task performance via task confidence perceptions. Note that the significant direct path from the interaction to task performance was reduced to non-significance when the effect of the mediator was statistically controlled.  $R^2$  values reflect the proportion of variance explained by each model, respectively. \* =  $p < .05$ , \*\* =  $p < .01$ .

It should be noted that the driving force behind this mediational model appeared to be task confidence perceptions and performance of male participants. When I examined the role of these perceptions in the stereotype threat-performance relationship amongst men, I found a marginally significant relationship between the instructional set manipulation and performance (step 1),  $\beta = 0.34$ ,  $t(29) = 1.92$ ,  $p = .065$ . In addition, I found that the instructional set manipulation was associated with task confidence perceptions (step 2) for these participants,  $\beta = 0.43$ ,  $t(29) = 2.50$ ,  $p = .02$ . When task confidence scores were entered into the regression equation (first), the independent effect of this variable remained significant,  $\beta = 0.81$ ,  $t(29) = 6.36$ ,  $p < .01$ , whereas the previously marginally significant gender X instructional set interaction on task performance,  $\beta = 0.34$ ,  $t(29) = 1.92$ ,  $p = .065$ , was no longer statistically reliable even at marginal levels,  $\beta = -0.002$ ,  $t(29) = -0.16$ ,  $p = .99$ . Thus, there was at least some evidence that the task confidence perceptions of men played a role in driving this mediational model.

When I examined the role of these task confidence perceptions in the stereotype threat-performance relationship amongst women, I failed find any empirical support for this mediator. I did not find a significant relationship between the instructional set manipulation and performance (step 1),  $\beta = 0.12$ ,  $t(63) = 0.94$ ,  $p = .35$ , nor did I find any evidence that the instructional set was associated with task confidence perceptions (step 2) for these participants,  $\beta = 0.06$ ,  $t(63) = 0.48$ ,  $p = .63$ . Therefore, task confidence did

not mediate the stereotype threat-performance relationship for women.

Although the present model demonstrated that task confidence perceptions fully mediated the relationship between gender X instructional set interaction and performance, there is an alternative model that may also explain these data. Given that task confidence perceptions were measured just after each item on the performance task, it is possible that the instructional set manipulation led to gender differences in task performance, which in turn influenced the task confidence perceptions of these participants. To rule out this alternative account, I conducted a mediational analysis (as described above) with a model that simultaneously included gender, instructional set, and the gender X instructional set interaction as predictor variables, task performance as the potential mediator, and task confidence perceptions as the outcome variable. I found that not only was the overall model—including all of the predictor variables—significant in predicting task confidence perceptions,  $R = .36$ ,  $F(3, 89) = 4.60$ ,  $p < .01$ , but the independent gender X instructional set interaction effect also predicted task performance (step 1),  $\beta = 0.69$ ,  $t(92) = 2.18$ ,  $p = .03$ . In addition, I found that both the overall model (as described above),  $R = .31$ ,  $F(3, 89) = 3.04$ ,  $p = .03$ , and the independent effect of the gender X instructional set interaction,  $\beta = 0.76$ ,  $t(92) = 2.37$ ,  $p = .02$ , significantly predicted task performance (step 2).

When task performance scores were entered into the regression equation (first), in addition to the predictors described above, I found that although the overall model including task performance scores did a better job of predicting task performance than the

original model,  $R = .73$ ,  $F(4, 88) = 25.42$ ,  $p < .01$ , closer inspection of the independent effects of these variables on task confidence perceptions revealed that task performance mediated the gender X instructional set-task confidence perceptions relationship. When task performance scores were entered into this regression equation (step 3), the independent effect of this variable on task confidence perceptions was significant,  $\beta = 0.66$ ,  $t(92) = 8.73$ ,  $p < .01$ , whereas the previously significant gender X instructional set interaction on task confidence,  $\beta = 0.69$ ,  $t(92) = 2.18$ ,  $p = .03$ , was no longer statistically reliable (step 4),  $\beta = 0.18$ ,  $t(92) = 0.75$ ,  $p = .46$ . An additional Sobel test analysis also revealed a significant difference in the direct path from the gender X instructional set interaction to task confidence perceptions after statistically controlling for task performance,  $t(92) = 2.29$ ,  $p = .02$ . These findings demonstrate that task performance fully mediated the effect of the gender X instructional set interaction on task confidence perceptions and that this alternative model can not be ruled out as a plausible account of these data.

### **Potential Moderators**

After discovering that the original achievement motivation measure failed yield performance motivation subscales that displayed acceptable levels of internal consistency— $\alpha_{performance\ approach} (91) = .61$  and  $\alpha_{performance\ avoidance} (91) = .69$ , respectively—I ran an exploratory factor analysis on the achievement motivation measure—using a varimax rotation—with following factor selection criterion: 1) only factors with eigen values above one were considered and (2) only factors that contained at least three items

with factor loadings above .6 were retained. After rotation, a four factor solution emerged. Since the fourth factor had only a single item that adhered to second selection criterion, it was not retained. The remaining three factors were interpreted as measures of *General Work Avoidance Motivation (GWA-M)*—composed of items 4, 8, and 12—*General Motivation to Performance Better Than Peers (GPP-M)*—composed of items 1, 5, and 9—and *General Motivation to Avoid of Ineptitude Perceptions (GAIP-M)*—consisting of items 2, 6, and 10. I then subjected these data to reliability and moderator analyses.

*General Work Avoidance Motivation (GWA-M)*. I conducted analyses on these data to examine whether GWA-M moderated the gender X instructional set interaction effect on performance. Participants scores on the GWA-M—which proved to be internally consistent,  $\alpha(91) = .80$ —were then dichotomized, using a median split procedure, and subjected to a 2 (gender) X 2 (instructional set) X 2 (GWA-M: High, Low) ANOVA. Only one main effect was significant—the main effect of general work avoidance,  $F(1, 83) = 7.92, p < .01$ . High GWA-M participants ( $M = 5.02, SD = 2.72$ ) performed significantly better on the task than Low GWA-M participants ( $M = 3.849, SD = 2.12$ ) on the math task. There was also a significant gender X GWA-M interaction,  $F(1, 83) = 8.57, p < .01$ . This interaction revealed that there women who were high in GWA-M ( $M = 4.21, SD = 2.29$ ) performed more poorly on the task when compared to their high GWA-M male counterparts ( $M = 6.47, SD = 2.87$ ). However, Women ( $M = 4.32, SD = 2.07$ ) who were low in GWA-M performed better than their low GWA-M male counterparts ( $M = 2.50, SD = 1.72$ ). However, this interaction was not qualified by



a significant three-way interaction amongst these factors,  $F(1, 83) = 0.19, p = .67$ , which suggests that general work avoidance motivation did not moderate the instructional set X gender interaction effect on performance.

*General Motivation to Perform Better Than Peers (GPP-M).* I conducted similar analyses on these data to examine whether GPP-M moderated the gender X instructional set interaction effect on performance using the statistical procedures described above. Although this measure proved to be reliable,  $\alpha(47) = .82$ , the gender X instructional set X GPP-M interaction failed to reach statistical significance,  $F(1, 83) = 0.23, p = .64$ . Given that the significant gender X instructional set interaction effect on performance remained intact,  $F(1, 83) = 7.02, p = .01$ , there was no evidence that GPP-M moderated this effect. No other significant main effects or interactions emerged, all  $p$ 's > .18.

*General motivation to avoid ineptitude perceptions (GAIP-M).* Individual differences in GAIP-M,  $\alpha(47) = .79$ , were examined as a potential moderator of the gender X instructional set interaction effect on performance using the same analyses described above. The analysis revealed a marginally significant gender main effects,  $F(1, 83) = 3.61, p = .061$  as well as a statistically reliable GAIP-M effect,  $F(1, 83) = 7.97, p < .01$ . In addition, there was a significant gender X GAIP-M interaction,  $F(1, 83) = 6.82, p = .01$ , which revealed that women who scored high on the GAIP-M ( $M = 4.30, SD = 2.22$ ) experienced performance deficits when compared to their male counterparts ( $M = 6.85, SD = 2.88$ ). Amongst participants who were low in this motive, women ( $M = 4.19, SD = 2.16$ ) performed better than their male counterparts ( $M = 3.69, SD = 2.68$ ).

However, given that the significant gender X instructional set interaction effect on performance remained intact,  $F(1, 83) = 5.89, p = .02$ , and given that the gender X instructional set X GAIP-M interaction failed to reach statistical significance,  $F(1, 83) = 0.00, p = .98$ , there was no evidence that GPP-M moderated this effect. No other significant main effects or interactions approached significance, all  $p$ 's  $> .23$ .

## Discussion

The primary purpose of Study 1 was to replicate previous findings by establishing a link between stereotype threat and task performance in women. As predicted, when female participants were presented with a math test that was described as having produced *gender differences*, these participants performed significantly more poorly than their male counterparts. In contrast, when the test was described as *gender fair*, women performed equally as well as men. A similar pattern of results occurred for women and men on the task confidence measure. These findings dovetail nicely with the results of Spencer et al. (1999; Study 2) who found that by informing women that there were gender differences on an upcoming math task, women performed more poorly than men. However, no gender differences emerged when the task was purported to be a gender fair test.

It should be noted that, although the pattern of results in Study 1 are consistent with those found by Spencer et al., there is an important distinction between the findings in both studies. The driving force behind the results on the performance measure in Study 1 was the performance of men. However, this was not the case in the Spencer et al.

study, which found that the performance of women were responsible for differences produced on the performance measure.

Evidence for stereotype lift was also discovered in Study 1. Indeed I found that men who were presented with a gender differences instructional set performed significantly better than their male counterparts in the gender fair condition. This finding was consistent with the extant literature on this phenomenon and underscored the notion that the activation of negative out-group can have implications for non-stigmatized individuals (Walton & Cohen, 2003).

A secondary goal of Study 1 was to examine the mechanisms that may underlie stereotype threat effects. Although I discovered that the impact of stereotype threat on performance was mediated by task confidence perceptions, there were two additional findings that make this rather straightforward account less clear. First, I found that a plausible alternative mediational model—with performance mediating the relationship between stereotype threat and task confidence—fit these data equally well. And given that the latter model could not be ruled out due to the temporal sequence in which the performance and task confidence measures were administered, it remains unclear as to which model truly accounts for these data. Second, it should be noted that both of these mediational models were heavily driven by the task confidence perceptions and performance of men. Whereas this finding may shed light on the potential underlying mechanisms of stereotype lift, it provides relatively little insight into what mediates the stereotype threat experience of women. And consistent with many reviews of this literature (e.g., Jones & Stangor, 2003), the mediational picture as it relates to women

remains illusive.

A final goal of Study 1 was to examine whether individual differences in achievement motivation moderated the effects of stereotype threat on task performance. I failed to find any evidence that performance motives, such as the motivations to avoid work and to avoid perceptions of ineptitude, moderated the impact of threat on the performance of women and men. These findings suggest that individual differences in performance do not play an integral role in moderating stereotype threat effects.

## **STUDY 2**

### **Method**

#### **Design and Participants**

This experiment took the form of a 2 (gender) X 2 (instructional set) between participants factorial design that was almost identical to Study 1. Task choice and task preference were the primary dependent measures. I recruited 71 male and female University of Maryland undergraduates to participate in this experiment in exchange for course credit during the fall semester of 2004. The data from 3 participants were removed from the analysis due to a computer error and a fourth participant was excluded due to having failed the manipulation check (as described below). This left a total of 67 participants (37 male and 30 female) randomly assigned to two experimental conditions. The ethnic breakdown of the sample included 37 European Americans, 11 African Americans, 10 Asian Americans and 9 participants self-identified as 'Other'. The most

prevalent majors were Psychology (34%), Biology (7%), Computer Science (7%) and Undecided (7%)—the latter three disciplines were tied for the second ranking. The mean VSAT, QSAT, College and High School GPAs, and math domain identification scores were 621, 622, 3.3, 3.6, and 3.4, respectively.

## **Procedure**

The procedure for Study 2 was almost identical to that of Study 1. The lone exceptions were that (1) a task choice measure served as the primary dependent measure and (2) that task preference was assessed after participants made this selection. Participants were given the same cover story and instructional set manipulation that was described in Study 1. After being randomly assigned to an instructional set condition, participants were informed via computer that they would be completing two initial tasks prior to completing a math task. They were further informed that they would be given an opportunity to choose the nature of the second task from amongst a set of randomly generated alternatives. Once again, the first task was designed to implicitly detect gender and anxiety activation, whereas the second task ostensibly gave participants an opportunity to choose the nature of this task—this selection served as the task choice measure—and asked them to indicate how much they preferred the chosen alternative over the non-chosen alternative—responses to this item served as the strength of choice measure. Both measures are described in more detail in the section that follows.

### ***Task Choice and Strength of Choice***

After completing the initial task, participants were given an opportunity to choose the nature of a second task on the computer. Participants were first shown a screen which read, “We would like you to choose the nature of the upcoming task from two randomly generated alternatives.” The computer then displayed a screen that stated, “Now generating alternatives” followed by the word “Working.” After a brief pause, the computer presented two alternatives that were presumably generated at random. The two alternatives were “1 = Math” and “2 = Proofreading”. Both the math and proofreading alternatives were selected based upon prior research which revealed that these two types of tasks were congruent in terms of their favorability ratings (Jones, 2003). These tasks were demonstrated to be uncorrelated in terms of their perceived favorability and were both preferred and chosen equally by a sample of collegiate men and women.

The computer then asked participants to “Please select the type of task you would like to complete on the upcoming task by pressing the number ‘1’ for a math task or the number ‘2’ for a proofreading task. The order in which these tasks were presented was counterbalanced. In addition, the reaction time associated with this choice was recorded.

After participants selected a task, they were asked to indicate how strongly they preferred the chosen alternative over the non-chosen alternative on a 7-point scale ranging from 1 (“I do not at all prefer the chosen alternative over the non-chosen alternative”) to 7 (“I strongly prefer the chosen alternative over the non-chosen alternative). Once participants’ scores on this measure were recorded, they completed a battery of post-

choice measures prior to completing what was presumed to be the second task. These items were almost identical to the pre-test and post-test measures that were almost identical to those administered in Study 1.<sup>11</sup> After completing these items, the computer program stopped and participants were probed for suspicion, debriefed, thanked, and dismissed.

## Results

### Manipulation Check

To verify the effectiveness of the instructional set manipulation, participants were asked to recall the nature of the performance differences that had been described to them by the communicator. As expected, after removing one participant in the *gender fair* condition that failed the manipulation check, I found that participants across conditions had little difficulty in recalling this information,  $\chi^2(3, n = 67) = 67.00, p < .01$ . In the *gender fair* condition, all but a single participant (97%) were able to correctly recall that the upcoming task had been known to produce no gender-differences in the past. Similarly, in the *gender differences* condition, nearly all of the participants (97%) were able to correctly identify that prior research had demonstrated gender differences on the main task. These data suggest that the threat manipulation was successful in allowing participants to interpret the video feedback in a manner consistent with the experimental manipulation.

## Task Choice

I submitted these data to a logistic regression analysis to examine whether there were any significant main effects or interactions on the task choice measure. Although the overall model—including gender, instructional set, and the gender X instructional set interaction—was successful in predicting task choice,  $\chi^2(3, n = 67) = 7.87, p = .05$ , Cox & Snell  $R^2 = .11$ , Nagelkerke  $R^2 = .15$ , no single predictor was statistically reliable in predicting task choice, all  $p$ 's  $> .20$ . This finding was contrary to Hypotheses 2a and 2b.

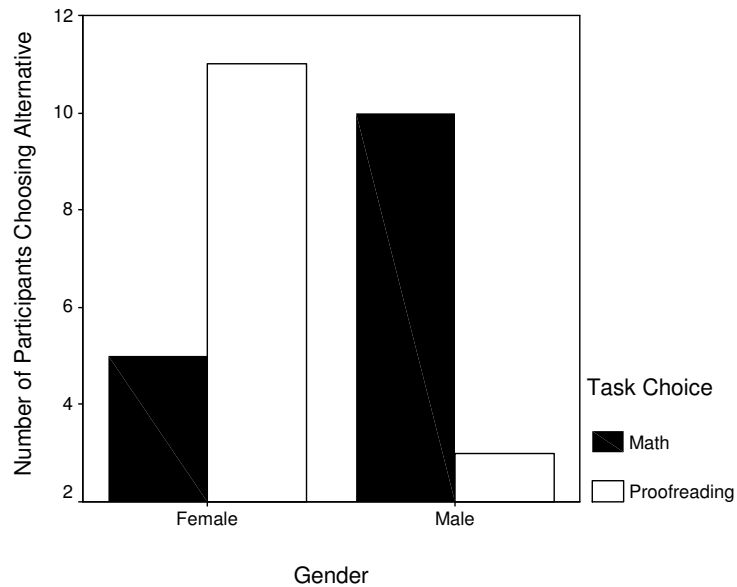
Although no significant main effects or interactions were detected on this measure, I examined whether any trends existed in these data by inspecting the simple effect of gender on task choice. In the *gender fair* condition, I found a significant association between these variables,  $\chi^2(1, n = 29) = 6.00, p = .01$ . As evidenced in Figure 3a, women were more likely to select a proofreading task (69%) over a math task (31%), whereas the opposite tendency in task selection emerged for men (23% vs. 77%, respectively).

Whereas the pattern of results for women in the *gender differences* condition was markedly different from those who received the *gender fair* instruction, the trend for men remained the same. As evidenced in Figure 3b, women were more likely to select a math task (57%) over a proofreading task (43%) in the *gender differences* condition. This trend was consistent with Hypothesis 2b. Men maintained their tendency to select a math task over a proofreading task (71% vs. 29%, respectively). Although these trends are intriguing with respect to the disidentification hypothesis, they should be interpreted with

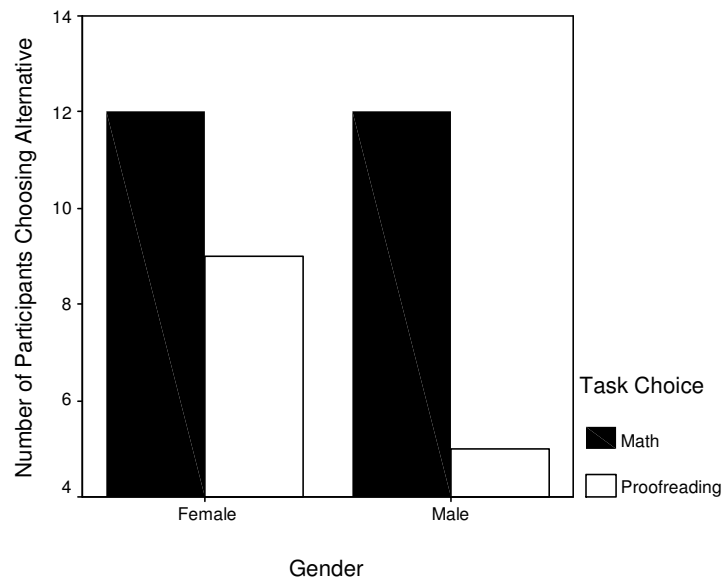


caution given that the relationship between gender and task choice for these participants

was non-significant,  $\chi^2(1, n = 38) = 0.73, p = .39$ .



*Figure 3a.* Number of participants choosing a given type of task as a function of gender in the *gender fair* conditions.



*Figure 3b.* Number of participants choosing a given type of task as a function of gender in the *gender differences* conditions.

## Strength of Choice

Participants' scores on the task preference measure were submitted to a 2 (gender) X 2 (instructional set) ANOVA. For those participants who selected the math task, only a marginally significant main effect for gender emerged,  $F(1, 35) = 3.71, p = .062$ , indicating that women ( $M = 5.47, SD = 1.87$ ) preferred the math task less than men ( $M = 6.55, SD = 1.01$ ) did. This finding is only partially consistent with Hypothesis 2a. No other significant main effects or interactions were detected, all  $p$ 's  $> .50$ .

Amongst participants who chose the proofreading task, only a significant condition main effect was detected,  $F(1, 24) = 5.90, p = .02$ , indicating that participants in the gender differences condition ( $M = 6.21, SD = 1.05$ ) preferred the proofreading task significantly more than their counterparts who received the gender fair instruction ( $M = 4.86, SD = 1.61$ ). No other significant main effects or interactions emerged, all  $p$ 's  $> .60$ .

## Reaction Time for the Chosen Alternative

I conducted an analysis to determine if the manipulation or participant gender had an impact on the reaction time measure. After applying a  $\log(10)$  transformation to these data, I subjected them to a 2 X 2 ANOVA. For participants who chose math task, only a marginal effect of gender emerged,  $F(1, 35) = 3.79, p = .06$ , revealing that amongst those who chose the math task, women ( $M = 4.00, SD = 0.19$ ) took longer to make their selection than their male ( $M = 3.88, SD = 0.24$ ) counterparts. This result was consistent with Hypothesis 4. No other main effects approached significance, all  $p$ 's  $> .40$ .

For participants who chose the proofreading task, no significant main effects or interactions emerged, all  $p$ 's > .60.

### Other Dependent Measures

*Gender activation.* A floor effect emerged on this measure with the majority of the values clustered near its lower limit—contributing to the positive skew in the distribution of these data (*skewness* = 1.61). Considering that more than one-half of the participants (54%) failed to complete any of the word completions in a manner consistent with gender-relevant constructs, I applied a logarithmic(10) transformation to normalize these data. However, after applying the transformation an even stronger positive skew emerged (*skewness* = 2.92). Thus, I conducted a 2 X 2 ANOVA on the non-transformed variable.

Although no significant main effects were detected on this measure,  $p$ 's > .50, there was a marginally significant gender X instructional set interaction,  $F(1, 63) = 3.09$ ,  $p = .084$ . As presented in Table 3, the simple effect of gender on stereotype activation revealed that, in the *gender differences* condition, women ( $M = 0.38$ ,  $SD = 0.59$ ) generated fewer gender word completions than their male counterparts ( $M = 0.76$ ,  $SD = 0.66$ )—this effect was marginally significant,  $t(36) = 1.87$ ,  $p = .07$ . In the *gender fair* condition there were no statistically reliable gender differences,  $t(27) = 0.85$ ,  $p > .40$ . Women and men did not differ in the number of gender word completions that they generated across instructional sets,  $t(35) = 1.69$ ,  $p = .10$ ,  $t(28) = 0.78$ ,  $p = .44$ , respectively.

Table 3

*Mean Number of Gender Word Completions as a Function of Gender and Instructional Set*

Gender of the Participant	Instructional Set					
	<u>Gender Differences</u>			<u>Gender Fair</u>		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
Female	0.38 <sub>a</sub>	0.59	21	0.87	1.15	16
Male	0.76 <sub>b</sub>	0.66	17	0.54	0.88	13

*Note.* Means sharing a different subscript are significantly different at marginal levels,  $p = .07$ . Higher means indicate a greater number of gender completions.

Even when these data were analyzed only for participants who chose the math alternative, an almost identical pattern of results emerged—no significant main effects, all  $p$ 's > .20, and a significant gender X instructional set interaction effect,  $F(1, 35) = 5.03$ ,  $p = .03$ . The relevant means and standard deviations were as follows: women,  $M_{gender\ differences} = 0.50$ ,  $SD = 0.67$ ;  $M_{gender\ fair} = 1.20$ ,  $SD = 0.84$ ; men,  $M_{gender\ differences} = 0.75$ ,  $SD = 0.75$ ;  $M_{gender\ fair} = 0.40$ ,  $SD = 0.52$ . For both analyses, the pattern of results obtained for women and men in the *gender differences* condition were unexpected and inconsistent with Hypothesis 4.

Amongst participants who chose the proofreading alternative, no significant main effects or interactions emerged, all  $p$ 's > .33.

*Self-doubt activation.* I also discovered that a positive skew (*skewness* = 1.43) was evident in participants' self-doubt activation scores. Considering the number of participants who failed to complete any of the word completions in a manner consistent with anxiety activation (66%), I applied a logarithmic(10) transformation in an attempt to normalize these data. However, after applying the transformation an even more profound positive skew emerged (*skewness* = 2.61). Therefore, I analyzed scores on the non-transformed measure with a 2 X 2 ANOVA. No significant main effects or interactions emerged irrespective of whether these data were analyzed overall or for participants who selected either the math or proofreading alternatives, respectively—all  $p$ 's > .10. Once again, this result was contrary to Hypothesis 4.

*Threat vs. Challenge perceptions.* Although this measure was not found reliable by conventional standards,  $\alpha(67) = .56$ , I subjected it to a 2 X 2 ANOVA. No significant main effects or interactions emerged, all  $p$ 's > .10. When analyzed only for participants who selected the math alternative, a significant effect of gender was discovered,  $F(1, 35) = 4.11$ ,  $p = .05$ . More specifically, women perceived the math task as significantly less of a challenge ( $M = 6.59$ ,  $SD = 1.64$ )—and presumably as more of a threat—than men ( $M = 7.67$ ,  $SD = 0.95$ ). This finding was only partially consistent with Hypothesis 4. No other significant main effects or interactions emerged for these participants, all  $p$ 's > .16. Similarly, no significant main effects or interactions were found when these data were analyzed for participants who selected the proofreading alternative, all  $p$ 's > .19.

*State Anxiety.* A 2 X 2 ANOVA was run on the Spielberger state anxiety inventory which was found to be internally consistent,  $\alpha(67) = .94$ . However, no significant main effects or interactions were detected, all  $p$ 's  $> .30$ . When analyzed for participants who selected the math alternative, only a marginal gender main effect emerged,  $F(1, 35) = 3.75, p = .06$ . This effect demonstrated that women reported higher levels of state anxiety ( $M = 1.76, SD = 0.63$ ) than men ( $M = 1.43, SD = 0.31$ ). This outcome was in partial support of Hypothesis 4. No other significant main effects or interactions were found, all  $p$ 's  $> .80$ . Similarly, no significant main effects or interactions emerged when analyzed for participants who chose the proofreading task, all  $p$ 's  $> .21$ .

*Task motivation.* I also conducted a 2 X 2 ANOVA on the task motivation measure. As expected no significant main effects or interactions were detected, all  $p$ 's  $> .20$ . However, when these data were analyzed only for participants that selected the math alternative, a marginally significant condition effect emerged,  $F(1, 35) = 3.19, p = .083$ . More specifically, I discovered that participants in the *gender fair* condition ( $M = 5.40, SD = 1.06$ ) reported being significantly more motivated than participants in the *gender differences* condition ( $M = 4.58, SD = 1.44$ ). This finding was only partially consistent with Hypothesis 4. No other significant main effects or interactions were found for these participants, all  $p$ 's  $> .70$ . Similarly, when analyzed only for participants who chose the proofreading alternative, no significant main effects or interactions emerged, all  $p$ 's  $> .34$ .

*Expectancies (overall).* I conducted a 2 X 2 ANOVA on the task expectancy

measure, which revealed only a significant gender main effect,  $F(1, 63) = 6.41, p = .01$ . More specifically, women ( $M = 5.14, SD = 1.06$ ) reported significantly lower task expectations than men ( $M = 5.80, SD = 0.93$ ). No other significant main effects or interactions emerged, all  $p$ 's  $> .20$ . Similarly, when I analyzed these data only for participants that selected the math alternative, a significant gender effect was detected,  $F(1, 35) = 5.39, p = .03$ . Once again, women ( $M = 4.88, SD = 1.22$ ) self-reported lower task expectations than men ( $M = 5.95, SD = 1.00$ ). However, no other main effects or interactions were statistically reliable, all  $p$ 's  $> .14$ . This finding was only partially consistent with Hypothesis 4. When analyzed only for participants who selected the proofreading alternative, once again, no significant main effects or interactions were found, all  $p$ 's  $> .89$ .

### **Exploratory Measures**

I examined whether individual differences in *General Work Avoidance Motivation (GWA-M)*,  $\alpha(67) = .79$ , *General Motivation to Performance Better Than Peers (GPP-M)*,  $\alpha(67) = .63$ , and *General Motivation to Avoid of Ineptitude Perceptions (GAIP-M)*,  $\alpha(67) = .63$ , played a role in predicting participants' task choice behavior. Therefore, I subjected each of performance motivation to a separate logistic regression equation prediction task choice. Each of these models also included gender, instructional set, and the gender X instructional set interaction with these variables to serve as predictors of choice.

Although all of the overall models were able to predict choice—albeit at marginal

levels— $\chi^2_{\text{GWA-M}}(4, n = 67) = 7.90, p = .09$ ,  $\chi^2_{\text{GPP-M}}(4, n = 67) = 7.92, p = .095$ , and  $\chi^2_{\text{GAIP-M}}(4, n = 67) = 8.66, p = .07$ , respectively, no single predictor in any of these separate models was statistically reliable in predicting task choice, all  $p$ 's  $> .17$ .

## Discussion

Study 2 generated several interesting trends with respect to the stereotype threat phenomenon. First, despite the presence of a non-significant gender X instructional set interaction, an interesting trend emerged which indicated that women in the *gender differences* condition were more likely to select a math task over a proofreading task. However, women in the *gender fair* condition were more likely to select a proofreading task over a math alternative. Men did not display any systematic differences in their selection of a math task over a proof-reading task irrespective of which instructional set that they were confronted with. Although suggestive, these trends support that notion that not only is task performance impacted by the effects of stereotype threat, but it is quite possible that the decision-making process of women and stigmatized group members may also be influenced by this phenomenon.

According to stereotype threat theory's disidentification hypothesis it is posited that, over time, women and stigmatized group members may disassociate from a domain and no longer view performance in that area as a vital part of the self-concept—despite having initially maintained a strong identification with the respective domain. Thus, one might predict that collegiate women—who are presumed to have had been threatened by the prospect of confirming a negative stereotype about their group at some point in their



academic history—would be *more* likely to avoid math tasks when under stereotype threat. In contrast, non-threatened women may be expected to approach such tasks. The women's achievement hypothesis predicts that women may actually approach such tasks initially and only disidentify from a given domain after experiencing consistent failure in that area. The findings in Study 2 did not support either of these hypotheses directly. Clearly more research is needed to further explore these hypotheses before any definitive conclusions can be reached.

The findings in Study 2 also suggest that strength of choice is a variable that deserves future consideration in threat research. Although paradoxical in nature, it was found that amongst women who did choose to engage in a math task, irrespective of whether under the vice of stereotype threat or not, they actually preferred this alternative less when compared to their male counterparts. The fact that women actively choose to engage in a task that they ultimately do not prefer that strongly is interesting considering that math task has been shown to be preferred equally by men and women in the past (Jones, 2003). Perhaps the lowered preference of this task by women could be rooted in a self-image protection motive in which these individuals create a ready made excuse for a potential poor performance. For instance, in the event that a woman who chose the math alternative did not perform well on the task, she could potentially say, "Well, the reason I didn't do well on this math task, even though I chose it, was because I didn't prefer it that strongly." Such a self-protective mechanism could potentially buffer their self-esteem. Unfortunately, given that the instructional set manipulation did not produce the predicted differences on this measure, it is unclear of what impact, if any, that stereotype threat has

on the task preference of men and women. Therefore, future research will be needed before true impact of threat on task preference can be understood.

The results of Study 2 failed to produce any evidence that individual differences in performance motivation were related task choice. Neither independently nor interactively were the motivations to avoid work, to performance better than one's peers, or to avoid perceptions of ineptitude able to effectively influence the task choices of women and men in this study. These findings suggest that whereas individual differences in achievement motivation served as an important moderator of stereotype threat effects on performance-based outcomes, they appear to be less influential upon the task choices made by threatened individuals.

And finally, both Studies 1 and 2 had the over-arching goal of trying to clarify the mediational mire surrounding stereotype threat as it relates to women. Studies 1 and 2 did not demonstrate that gender or self-doubt activation, threat vs. challenge perceptions, self-reported anxiety, self-esteem, motivation, nor overall task expectancies, mediated the effect of stereotype threat on task performance or task choice. And the failure to detect a significant gender X instructional set interaction on the aforesaid measures despite using both implicit and explicit measurement was also quite puzzling. For instance, Studies 1 and 2 failed to replicate prior research examining threat effects using implicit measurement (Brown & Josephs, 1999; Steele & Aronson, 1995) despite using materials that were either identical or slight variations of the materials used by these researchers. And even when I did find significant results using these measures, these data were often

inconsistent with findings reported in previous research—in Study 2, I found that threatened women actually exhibited lower levels of gender activation than men under similar situational constraints, whereas women and men in the control condition did not differ in their activation levels. Given that these implicit measures were presented directly after the threat manipulation, it is unlikely that the failure to replicate prior research was because these items were too delayed. A more simplistic possibility is that these measures were tapping into a different construct in the current context. Suffice it to say that the relationship between threat and stereotype activation in women remains unclear.

## STUDY 3

### Method

#### Design and Participants

This experiment took the form of a 2 (misattribution cue: *present, absent*) X 2 (self-affirmation opportunity: *present, absent*) between participants factorial design. All participants were presented with the *gender differences* instructional set and math performance measure described in Study 1. I recruited 44 female undergraduates from the University of Maryland to participate in this experiment in exchange for course credit during the spring semester of 2004. The data from 2 participants—one participant in the self-affirmation opportunity: *present*/misattribution cue: *absent* condition and a second

participant in the in the self-affirmation opportunity: *present/misattribution cue: present* condition—were excluded because it was determined that these participants did not take the performance task seriously.<sup>12</sup> This left a total of 42 women randomly assigned to four experimental conditions. The ethnic breakdown of the sample included 22 European Americans, 6 African Americans, 3 Asian Americans and 11 participants self-identified as ‘Other’. The most prevalent majors were Psychology (50%), Education (12%), Letters & Sciences (10%), Criminology (7%), and Undecided (7%). The mean VSAT, QSAT, and College and High School GPAs were 583, 607, 3.3, and 3.7, respectively. The mean math-DIM score for the sample was 3.2 indicating that participants were both skilled in and identified with the domain of mathematics.

## **Procedure**

The procedure for Study 3 was similar to that of Study 1 with the lone exceptions of (1) how the experimental manipulations were introduced and (2) the manipulation checks use to measure the effectiveness of these manipulations.

### ***Malfunctioning Computer Cover Story and Misattribution Manipulation***

Upon arrival to the lab, all participants were met by a male experimenter who provided them with same cover story described in Study 1. Participants were seated in a lab room that contained two computers and, for one-half of the participants—those in the *misattribution cue—present* condition—the second computer was rigged to appear to be malfunctioning. For all participants in this condition, the rightmost computer was

presumably malfunctioning (e.g., the computer screen displayed a looped static image), whereas the leftmost computer was in perfect working order (for a visual depiction of the orientation of these computers, please refer to Figure 4). Participants were then seated at the leftmost computer and presented with the misattribution cue manipulation.



*Figure 4.* Visual depiction of each computer's orientation.

These participants were further informed that the lab had been experiencing networking problems on some of its computers and that a technician from the on-campus computing center would be servicing the affected machines later in the day. In addition, these participants were urged not to touch the malfunctioning computer or its peripherals while completing the upcoming tasks as instructed by the service technician. Despite this ostensible shortcoming, participants were told that they were expected to make the best of

the situation. These instructions were critical given that optimal manipulation of the misattribution cue hinged on being able to convince participants that one of the computers was not functioning properly. And finally, participants in this condition were informed that participants who completed the experiment earlier in the day had indicated that malfunctioning computer had made them feel anxious while completing the tasks.

The remaining half of the participants—those in the *misattribution cue—absent* condition—were not given this cover story and completed all of the subsequent tasks in virtually the same experimental context as their malfunctioning computer counterparts. For these participants, the second computer was turned off and did not appear to be malfunctioning in any way. These participants were not provided with any information regarding the reactions of prior participants to the presence of the malfunctioning computer. I assumed that by manipulating the presence (or absence) of the misattribution cue in this manner, I would be able to assess the extent to which participants would misattribute any arousal to the external stimulus—that is, the malfunctioning computer.

### ***Self-Affirmation Manipulation***

Upon receiving the misattribution cue manipulation, all participants were presented with the *gender differences* instructional set and then confronted with two initial tasks followed by a math task. The first task was enclosed within an envelope labeled ‘Task 1’ in the same manner as described in Study 1. This task served as the self-affirmation manipulation with one-half of the participants having the opportunity to affirm an aspect of the self-concept in a domain that was highly related to the domain of

mathematics—the general academic domain. Using a free-format response measure, participants were asked to describe their academic accomplishments by writing a brief paragraph about their GPA. The instructions stated:

“The average college student in America has a GPA of 3.0 on a 4.0 scale and an 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. Your essay need not be particularly long and there is no time limit for this task. However, please write no more than 150 words.”<sup>13</sup>

Given that this manipulation appeared to have a high degree of relevance to the mathematical domain, I assumed that stereotype threat would be effectively reduced in this condition. I further reasoned that this manipulation would be unlikely to heighten demand for these participants since the domain of this affirmation was not specific to mathematics.

The remaining half of the participants did not have to complete this paragraph and, upon opening the folder for the first task, discovered a document that stated, “You do not need to complete Task 1, PLEASE PROCEED to Task 2.” These participants were remained unaware that the first task consisted of an essay writing exercise and simply advanced to the second task without having an opportunity to affirm the self.

Upon completion of the first task, all participants were presented with a second task, labeled ‘Task 2’, that was identical to the gender and self-doubt activation measures described in Study 1. After completing this measure, participants then completed the same pre-test, performance, and post-test measures that were described in Study 1. The lone exceptions were that two manipulation checks were included at the end of the

experiment to examine the effectiveness of the misattribution cue and self-affirmation manipulations.

*Misattribution cue manipulation check.* Participants completed a single item designed to examine the effectiveness of the misattribution cue manipulation. They were asked to recall what they were told regarding how other participants had reported feeling about the presence of the malfunctioning computer in this experiment. This item had three response choices including (1) it made them feel anxious, (2) no such information was given to me, or (3) I do not remember.

*Self-affirmation manipulation check.* Participants also completed a single item designed to examine the effectiveness of the self-affirmation manipulation. Participants were asked to indicate what they were told to do on the first task using one of three response alternatives. Participants either indicated that (1) they were asked to write about (their) GPA, (2) that (they) did not have to complete Task 1 and were instructed to proceed to Task 2, or (3) that (they) did not remember. Upon completion of this item, the computer program stopped and participants were probed for suspicion, debriefed, and dismissed.

## **Results**

### **Manipulation Checks**

*Misattribution cue manipulation check.* To verify the effectiveness of this manipulation, participants were asked to recall what they were told regarding how other



participants had reported feeling about the presence of the malfunctioning computer. As expected, with the exception of 1 participant in the *misattribution cue—absent* condition that failed the manipulation check, I found that participants across conditions had relatively little difficulty in recalling this information,  $\chi^2(2, n = 42) = 27.90, p < .01$ . In the *misattribution cue—present* condition, almost all of the participants (86%) were able to correctly recall that the presence of the malfunctioning computer had reportedly made participants feel anxious while completing the upcoming tasks. The remaining three participants in this condition reported that they had either not been given information (9%) or that they did not recall it (5%). Similarly, in the *misattribution cue—absent* condition, all but a single participant (5%) were able to correctly identify that they had either not been given (76%) this information, or that they had failed to remember it (19%). These data suggest that the misattribution cue manipulation was successful in allowing participants to correctly decipher this information.

*Self-affirmation manipulation check.* To verify the effectiveness of the self-affirmation manipulation, participants were asked to recall what they had done for the initial task. As predicted, I found that all participants across self-affirmation conditions had correctly recalled whether or not they had been presented with this information,  $\chi^2(1, n = 42) = 42.00, p < .01$ .

## **Task Performance**

Once again, the math performance measure (identical to the one presented in Study 1) proved to be sufficiently challenging, but not impossible, with the percentage of

participants in Study 3 getting a particular item correct ranging from 38% to 81%. I then subjected this measure to a 2 (misattribution cue) X 2 (self-affirmation opportunity) ANOVA.<sup>14</sup>

Only a significant self-affirmation opportunity main effect emerged,  $F(1, 38) = 4.29, p < .05$ . This result was consistent with Hypothesis 3, in that women in the *self-affirmation—present* condition ( $M = 5.35, SD = 1.93$ ) performed significantly better on the task when compared to their *self-affirmation—absent* condition ( $M = 4.14, SD = 1.86$ ) (Cohen's  $D = 0.64$ ) counterparts. No other significant main effects or interactions emerged, all  $p$ 's  $> .25$ . The failure to find a significant misattribution cue main effect was inconsistent with Hypothesis 3.

*Reaction time (per item)*. I also tested whether these manipulations had an impact on the reaction times of participants after applying a  $\log(10)$  transformation to these data. Once again, response times on the performance measure were recorded, transformed, and then averaged. After deleting four participants from this analysis due to their inability to complete all nine items within the allotted time frame, I subjected these data to a 2 X 2 ANOVA. However, no significant main effects or interactions were detected, all  $p$ 's  $> .13$ . This finding was consistent with Hypothesis 4.

### **Potential Mediators: Implicit Measures**

*Gender activation*. Although scores on this measure exhibited a positive skew ( $skewness = 0.89$ ), I subjected these data to a 2 X 2 ANOVA. This analysis failed to yield any significant main effects,  $F_{\text{misattribution cue}}(1, 38) = 0.21, p = .65, F_{\text{self-affirmation opportunity}}(1,$

38) = 1.32,  $p = .26$ . However, a statistically reliable misattribution cue X self-affirmation opportunity interaction did emerge,  $F(1, 38) = 4.75, p = .04$ . I examined the simple effects of this interaction using planned comparisons involving the removal conditions and the condition that offered no removal opportunity—that is, the condition presumed to evoke threat. Means, standard deviations, and sample sizes per cell for each comparison are presented in Table 4.

Table 4

*Mean Number of Gender Word Completions as a Function of Self-affirmation and Misattribution Opportunity*

Self-Affirmation Opportunity	Misattribution Opportunity					
	<u>Present</u>			<u>Absent</u>		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
Present	1.00 <sub>+</sub>	1.16	10	0.30 <sub>a,+</sub>	0.48	10
Absent	0.73	0.65	11	1.18 <sub>b</sub>	0.98	11

*Note.* Means sharing a different subscript are significantly different,  $p < .05$ . + indicates that means are significantly different at marginal levels,  $p < .10$ . Higher means indicate a greater number of gender completions.

Although unexpected and inconsistent with Hypothesis 4, planned comparisons revealed that having an opportunity to affirm the self and to misattribute anxiety to an external source did not lead to a significant reduction in gender activation ( $M = 1.00, SD$

= 1.16), when compared to participants who did not receive either removal manipulation ( $M = 1.18, SD = 0.98, t(19) = 0.39, p = .70$ ). However, having only the opportunity to affirm the self ( $M = 0.30, SD = 0.48$ ) significantly lowered gender activation when compared to participants who did not receive either removal opportunity ( $M = 1.18, SD = 0.98, t(19) = 2.57, p = .02$ ). The latter finding was consistent with Hypothesis 4. Pairing a self-affirm opportunity with a misattribution opportunity did lead to a marginally significant trend for higher levels of gender activation ( $M = 1.00, SD = 1.16$ ), when compared to participants who merely received the self-affirmation manipulation in isolation, ( $M = 0.30, SD = 0.48, t(18) = 1.76, p = .09$ ).

*Self-doubt activation.* Similar to the previous measure, a slight positive skew ( $skewness = 1.04$ ) was evident in these data. Despite the positive skew, I subjected these data to a 2 X 2 ANOVA which failed to yield any significant main effects or interactions, all  $p$ 's > .75. Once again, these results did not support Hypothesis 4.

### **Potential Mediators: Self-Report Measures**

*Task motivation.* I subjected participants' responses on the task motivation measure to a 2 X 2 ANOVA. As expected, no significant main effects or interactions were detected, all  $p$ 's > .75. These data suggest that sheer motivation does not explain the increased performance experienced by women given the opportunity to affirm the self, when compared to those who were not afforded this opportunity. This finding was consistent with Hypothesis 4.

*Expectancies (overall).* Participants' overall performance expectations were

subjected to a 2 X 2 ANOVA. However, no significant main effects or interactions emerged, all  $p$ 's > .57. Consistent with Hypothesis 4, these data suggest that pre-task performance expectancies do not explain the increased performance experienced by women given an opportunity to affirm the self, when compared to participants who were not given this opportunity.

*Threat vs. challenge perceptions.* I examined participants' responses to the threat vs. challenge measure which, similar to studies 1 and 2, was found to only exhibit a modest level of internal consistency,  $\alpha(42) = .51$ . I then subjected these data to a 2 X 2 ANOVA which failed to reveal any significant main effects or interactions emerged, all  $p$ 's > .19. This finding did not support Hypothesis 4.

*State Anxiety.* I subjected participants' scores on Spielberger state anxiety inventory,  $\alpha(42) = .96$ , to a 2 X 2 ANOVA. Only a marginally significant main effect of self-affirmation opportunity emerged,  $F(1, 38) = 3.90, p = .056$ . Women in the *self-affirmation—present* condition ( $M = 2.32, SD = 0.73$ ) reported experiencing significantly lower levels of anxiety than their counterparts in the *self-affirmation—absent* condition ( $M = 2.71, SD = 0.55$ ). This outcome supported Hypothesis 4. No other significant main effects or interactions emerged, all  $p$ 's > .31.

*Self-Esteem.* Participants' scores on the Rosenberg self-esteem scale,  $\alpha(42) = .85$ , were subjected to a 2 X 2 ANOVA. However, no significant main effects or interactions were detected, all  $p$ 's > .40. This finding supported Hypothesis 4.

*Task Confidence.* I averaged participants' task confidence ratings per item and

subjected them to a 2 X 2 ANOVA. Although, no significant main effects emerged, all  $p$ 's > .22, a marginally significant misattribution cue X self-affirmation opportunity interaction was detected,  $F(1, 38) = 2.94, p = .09$ . I conducted t-tests to examine the simple effects of this interaction. For women provided with a misattribution cue, task confidence was significantly higher when these participants were also provided with a self-affirmation opportunity ( $M = 6.32, SD = 1.18$ ) as opposed to when this opportunity was not afforded ( $M = 5.02, SD = 1.49$ ),  $t(19) = 2.20, p = .04$ . Amongst participants who were not given an misattribution cue, there were no appreciable difference in task confidence perceptions irrespective of whether these participants had an opportunity to affirm the self ( $M = 5.36, SD = 0.94$ ) or whether no such opportunity was provided ( $M = 5.58, SD = 1.85$ ),  $t(19) = 0.34, p = .74$ . In addition, there was no appreciable difference in the task confidence perceptions of participants who received both removal strategies ( $M = 6.32, SD = 1.18$ ) when compared to those who received neither strategy ( $M = 5.58, SD = 1.85$ ),  $t(19) = 1.08, p = .29$ . This finding did not support Hypothesis 4.

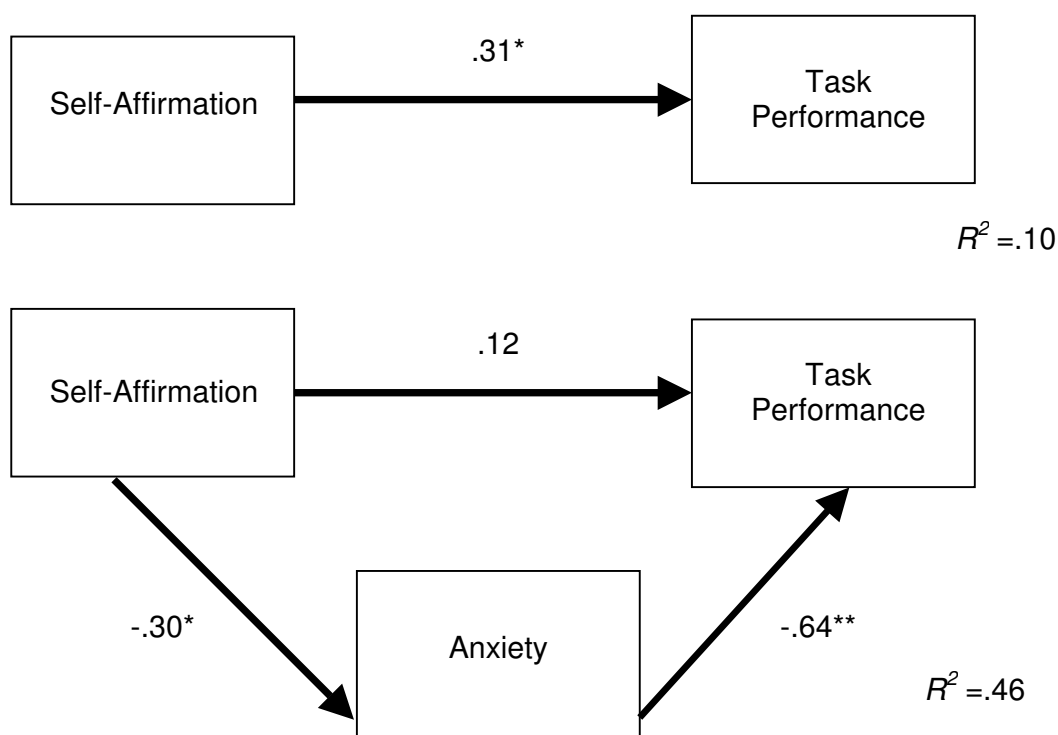
### **Mediational Analysis**

I examined whether anxiety was driving the differences in performance by including this variable in a regression model using the statistical approach outlined in Study 1—given that this was the only potential mediating variable that yielded even a marginally significant self-affirmation opportunity main effect. If the significant main effect of self-affirmation opportunity on task performance found earlier was either partially or fully mediated by anxiety, then these variables would have to meet the four

criterion of the mediational process.

To satisfy the first step of this process, I examined whether there was significant relationship between the self-affirmation opportunity main effect and task performance by entering the self-affirmation opportunity variable into a regression equation predicting task performance. As expected, the model was significant,  $R = .31$ ,  $F(1, 40) = 4.31$ ,  $p = .04$ , which indicated that having a self-affirmation opportunity was associated with increased task performance. In the second step, I examined whether the self-affirmation opportunity variable was correlated with self-reported anxiety. Indeed I found a significant association between these variables,  $R = -0.31$ ,  $F(1, 40) = 4.00$ ,  $p = .05$ , which demonstrated that having a self-affirmation opportunity was associated with lower levels of self-reported anxiety. In the third and fourth steps of this process, I entered both the anxiety and self-affirmation opportunity variables, respectively, into a regression model predicting task performance. Not only was the overall model successful in predicting task performance,  $R = .68$ ,  $F(2, 39) = 16.89$ ,  $p < .01$ , but I also discovered a statistically reliable negative correlation between anxiety and task performance,  $\beta = -0.64$ ,  $t(42) = -5.17$ ,  $p < .01$ . However, after statistically controlling for anxiety in this model, the once significant self-affirmation effect,  $\beta = .31$ ,  $t(42) = 2.08$ ,  $p = .04$ , was now reduced to non-significance,  $\beta = 0.12$ ,  $t(42) = 0.98$ ,  $p = .33$ . Although the overall model including both variables did a better job of predicting performance than the model only including the self-affirmation manipulation—as evidenced by a significant change in the proportion of variance explained by the model,  $R^2_{\text{change}} = .37$ ,  $F_{\text{change}}(1, 39) = 26.69$ ,  $p < .01$ —closer

inspection of the more inclusive model demonstrated that anxiety fully mediated the effect of self-affirmation on task performance. An additional Sobel test analysis revealed that there was a marginally significant reduction in the direct path from the self-affirmation opportunity variable to task performance after statistically controlling for self-reported anxiety,  $t(42) = 1.86, p = .06$ . This result suggests that I have found some evidence of anxiety as a mediator of performance in this study. Figure 5 presents the mediation of the self-affirmation effect on task performance by anxiety.



*Figure 5.* Mediation of the self-affirmation effect on task performance via anxiety. Note that the significant direct path from the interaction to task performance was reduced to non-significance when the effect of the mediator was statistically controlled.  $R^2$  values reflect the proportion of variance explained by each model, respectively. \* =  $p < .05$ , \*\* =  $p < .01$ .



Given that the anxiety measure was assessed after task performance, it is possible that the self-affirmation manipulation influenced task performance, which in turn produced differences on the anxiety measure. To rule out this plausible alternative account of these data, I conducted a mediational analysis with self-affirmation opportunity predicting anxiety and task performance serving as the mediator variable. I found that the self-affirmation variable significantly predicted anxiety (Step 1),  $R = .30$ ,  $F(1, 40) = 4.00$ ,  $p = .05$ . In addition, I found that the self-affirmation manipulation was also significantly correlated with task performance (Step 2),  $R = .31$ ,  $F(1, 40) = 4.31$ ,  $p = .04$ . When task performance scores were entered into this regression equation (step 3) including both self-affirmation opportunity and anxiety, the independent effect of performance on anxiety was significant,  $\beta = -0.64$ ,  $t(41) = -5.17$ ,  $p < .01$ , whereas the previously significant effect of self-affirmation opportunity on anxiety,  $\beta R = .30$ ,  $F(1, 40) = 4.00$ ,  $p = .05$ , was no longer statistically reliable (step 4),  $\beta = -0.10$ ,  $t(41) = -0.82$ ,  $p = .42$ . A subsequent Sobel test analysis revealed a significant difference in the direct path from the self-affirmation manipulation to anxiety after statistically controlling for task performance,  $t(41) = -1.92$ ,  $p = .05$ . These findings demonstrate that task performance fully mediated the effect of the self-affirmation manipulation on anxiety perceptions and that this alternative model can not be ruled out as a plausible account of these data.

## Discussion

Study 3 presents several findings with respect to the impact of two potential reduction strategies on the performance of women under stereotype threat. First, Study 3

revealed that having an opportunity to affirm the self, prior to completing a math test, improved the performance of threatened women. More specifically, I found that women given an opportunity to affirm to the self performed significantly better than women not given a self-affirmation opportunity. However, having only the opportunity to misattribute arousal did not lead to increased performance when compared to women who had neither removal opportunity.

Although I discovered that the impact of self-affirmations on performance was mediated by anxiety in Study 3, the finding that an equally plausible model with anxiety and performance serving as outcome and mediator variables casts doubt upon the fit of the initial model to these data. And given that the latter model can not be ruled out due to the temporal sequencing of these variables, I encountered some of the same mediational ambiguity that has been described in prior research (Jones & Stangor, 2003). However, what does remain clear is that, in terms of the alleviation of stereotype threat via self-affirmation, anxiety does appear to play a role in this process.

One finding worth noting was the general ineffectiveness of the misattribution manipulation. Similar to prior research (Stone et al., 1999; Study 2), I found a significant effect of this manipulation on the manipulation check—although Stone et al. also demonstrated this effect but in the opposite direction of what was to be expected. However, I failed to find a significant main effect or interaction involving the misattribution manipulation on task performance. The failure of this manipulation to produce differences on the Spielberger state anxiety inventory—which could be viewed

as a way of assessing whether the misattribution manipulation successfully manipulated the perceived anxiety source—suggests that although participants may have perceived differences in terms of the source of this arousal, they may not have detected the variation in this manipulation. In addition, the failure to produce differences on this measure suggests that women in the *misattribution cue—present* condition may not have actually (mis)attributed their arousal to the malfunctioning computer. Their failure to do so may have stemmed from the manipulation being too subtle to produce the intended alleviation effect. Alternatively, it may have been or that the Spielberger state anxiety inventory was either too insensitive or too delayed to detect differences in anxiety. However, this explanation seems less plausible given the internal consistency and widespread generalizability of this measure and given that fact that no differences were detected on the implicit self-doubt activation measure as a function of the misattribution manipulation (which preceded the performance task). Given the null results produced by the misattribution paradigm in the Study 3 and the inconclusive findings produced in other research using this framework within the threat literature (e.g., Stone et al., 1999; Study 2), suffice it to say that the merits of this procedure for reducing stereotype threat—whether used in a classic sense or not—remain unclear.

## General Discussion

Prior research has demonstrated that the activation of negative group-based stereotypes can depress the performance of women and minorities in the academic domain (e.g., Spencer et al. 1999; Steele & Aronson, 1995). Study 1 replicated this effect

by demonstrating that the performance of women, when compared to men, could be exacerbated by merely informing them that an upcoming math test had produced gender differences. When this instrument was described as having produced no gender differences, women and men performed equally well on the task.

Reviews of the stereotype threat literature (Jones & Stangor, 2003; Wheeler & Petty, 2001) have also concluded that the mediation of the stereotype threat outcomes remain unclear. With regards to women, prior research has produced relatively few studies that have found a significant mediational link between stereotype threat and performance. Outside of evidence derived from Spencer et al. (1999; Study 3)—which found marginal support for anxiety as a mediator of threat outcomes in women—there has been little clarity added to the mediational mire surrounding this phenomenon. Study 1 was able to generate support for task confidence perceptions as a potential mediator of the threat-performance relationship. However, this finding should be interpreted with some degree of caution for two reasons. First, the initial mediational analysis in Study 1 revealed that such perceptions are more likely to explain the boost effect experienced by men (Walton & Cohen, 2003) than they are to explain the performance deficits experienced by stereotype threatened women. Second, a plausible alternative account of this mediational model was found to fit these data as well as the initial model. Thus, the mediation of both threat and stereotype lift for women and men, respectively, remains unclear—although, it appears that task confidence perceptions do play a role in stereotype lift as it relates to men.

To this point, the examination of non-performance based dependent measures has

been relatively ignored in this burgeoning literature. Study 2 addressed this research question directly by examining the impact of stereotype threat on task choice. I found a provocative trend in the task choice behavior of women who were more likely to select a proofreading task over a math task under control conditions, whereas the opposite choice pattern emerged for threatened women. The task choice behavior of men appeared to have remained consistent across conditions. Although intriguing, these findings are speculative and further research will be needed before it can be determined whether threat effects generalize to non-performance based outcomes.

To date, only a limited number of studies have focused on ways to effectively reduce the impact of threat within the experimental context. Although several experiments have examined the merits of reduction strategies that redefine the situation as non-threatening (e.g., Steele & Aronson, 1995) or that diffuse responsibility (McIntyre, Paulson, & Lord, 2003), relatively few studies have examined the utility of misattribution processes (in the classic sense) and self-affirmations as a means to buffering women from stereotype threat. Amongst the studies that have had the explicit goal of alleviating threat effects, almost all of them have examined the impact of a single removal strategy in isolation—that is, as opposed to assessing the merits of employing multiple removal strategies. Study 3 examined the impact of multiple reduction strategies on the effects of threat and found that self-affirmations were particularly effective in removing the effects of threat in women. By allowing women to affirm the self, prior to completing a math task, these individuals performed significantly better on a math task than women who did not have an affirmation opportunity. Having an affirmation opportunity was also

associated with decreased anxiety, which in turn, was correlated with increased performance. And although an initial mediational model revealed that the change in the direct path from self-affirmation to performance proved to be marginally significant after statistically controlling for anxiety, this model suggested that self-affirmations can be used to effectively reduce anxiety and to increase the task performance of threatened women. However, given that an equally plausible alternative model was also found to fit these data, there some degree of ambiguity remains as to the specific role of anxiety in this mediational chain.

### **Implications for Stereotype Threat Theory**

Collectively, these present research expands our understanding of the stereotype threat phenomenon in several important ways. First, stereotype threat appears to influence the math performance of women when compared to their male counterparts. This finding replicated prior research (e.g., Spencer et al. 1999) although the underlying mechanism(s) of this process remain elusive.

Second, the present research examined the process that underlies the boost effect experienced by men on performance-based tasks. I found that this effect appears to be heavily rooted in the task confidence perceptions of men under these situational constraints. The current mediational model demonstrated that men who received information about negative-out group stereotypes increased their perceptions of task confidence, which in turn, led to an increase in task performance when compared to their female counterparts. However, it should be noted that an alternative model also found

that task performance mediated the threat-confidence relationship, which could also explain these data. What remains clear, is that task confidence perceptions appear to play a role in this process.

Third, the present research adds to our understanding of the impact of stereotype threat on non-performance based measures, such as task choice. Although speculative, the current findings suggest that threat may operate differently across different types of tasks. More specifically, there appeared to be trend which indicated that the activation of stereotype threat in women may actually lead to an approach tendency, as opposed to an avoidance process, when these individuals select tasks in a stereotyped domain.

### **Implications for Educational Environments**

The current findings also have implications for those in educational settings and in public policy and can be used as an aid in setting research priorities and selecting interventions that will most likely buffer stigmatized individuals from the burden of stereotype threat. For instance, the present findings underscore the importance of understanding how prevalent stereotypes can have a profound influence on the test performance of women and stigmatized group members. Therefore, by adopting initiatives, policy, and curriculum that address how tests are presented to students, the potential for stereotype threat effects to influence performance can be greatly reduced.

A second implication of the current research is to illuminate the merits of including self-affirmations in the educational context. I found that by simply having women affirm the self-concept, prior to completing a math exam, I was able to

significantly improve their performance when compared to women who were not given this option. Although it remains unclear as to whether self-affirmations mediate the threat relationship, it is apparent that they are involved in the experience of stereotype threatened participants at some point. These findings suggest that an adoption of this strategy within the educational context could be a relatively inexpensive and yet powerful way to ameliorate the performance of women and stigmatized group members. However, what appears to be critical is the placement of these affirmations. It appears that this reduction strategy may be most powerful when offered before an exam.

### **Future Directions for Stereotype Threat Research**

Although the present research has expanded our understanding of the stereotype threat phenomenon, further research is needed in several important areas. First, future research is needed to explore the impact of threat on other types of dependent measures aside from task performance. Outside of the current research, the impact of this phenomenon on non-cognitive measures (e.g., sports-related tasks) has only been examined in a modicum of empirical studies (Stone et al., 1999). Clearly research on the influence of stereotype threat on non-performance based outcomes, particularly those with relevance to more applied settings (e.g. public speaking), would be an interesting area worth exploring.

A second direction for future research is to further explore the role of achievement motivation in stereotype threat effects. Given that relatively few empirical studies have actively examined the influence of this disposition on threat outcomes, the potential



utility of this variable remains unclear. And although Study 1 and Study 2 attempted to examine the moderational impact achievement motives on stereotype threat effects, neither study provided conclusive evidence regarding its potential impact on threat outcomes. Perhaps, using a more reliable index of achievement motivation could also be helpful in aiding those interested in pursuing this area of research.

A third direction for future research is in the area of devising ways to alleviate the impact of stereotype threat. Given that the effect of stereotype threat has been well documented (Jones & Stangor, 2003), future researchers may be well served by exploring the ways in which stereotype threat effects can be effectively reduced. Whereas several studies have found that by not characterizing a task as diagnostic of ability, one can remove the deleterious effects of threat (Steele & Aronson, 1995), the current findings suggest that by simply allowing women to affirm the self prior to completing a math test, the effects of stereotype threat can be effectively alleviated. Although this strategy appears promising, further research is needed to discover other ways of reducing this situational predicament from the cognitions and behaviors of women and stigmatized group members.

A fourth direction for future research is to further examine the disidentification tenet of stereotype threat theory. The present thesis is one of the few empirical studies that examines this process, albeit indirectly, by measuring the choice behavior of women and men under stereotype threat. Although suggestive, I found a trend demonstrating that women under stereotype threat appeared to select tasks in a potentially threatening domain. This trend would challenge the current formulation of the disidentification

hypothesis in terms of its implication that women would be likely to avoid tasks in a stigmatized domain. However, future research is needed to further understand this process before any definitive conclusions can be reached.

Finally, future research will need to examine the effect of stereotype threat in real world contexts. Although, many empirical studies have failed to examine the impact of threat in applied settings, Steele and Davies (2003) have estimated that lab studies may actually *underestimate* the effect of threat when compared to what is experienced in the real world. Such a prospect is intriguing, but undoubtedly speculative, given the minimal number of applied studies that have been conducted within the threat literature and given the inconclusive nature of the results produced by this sample (Jones & Stangor, 2003). However, what is clear is that future research will need to examine threat outside of the lab as a way to aid policy makers, research scientists, and educators in the prediction, control, and prevention of this phenomenon.

## FOOTNOTES

<sup>1</sup> Some researchers have noted that standardized exams such as the SAT do not utilize representative samples given that most individuals who complete the exam are prospective college students (Neisser, 1998). Other researchers have criticized the validity of such examinations on the grounds of their various shortcomings when applied to women (i.e., that the SAT tends to over-predict for men and under-predict for women—high school grades are presumed to be more predictive for the latter group [Gross, 1998]).

<sup>2</sup> This performance improvement does tend to dissipate after adolescence. Some researchers (Jenks & Phillips, 1998) have reasoned that this dissipation effect could be due to a transformation of the social context from one that is more like that of European Americans—as experienced during their pre-adolescent years—to one that is more like that of other minorities.

<sup>3</sup> Neisser (1998) supports this contention by examining Flynn’s synopsis of longitudinal scores posted on the Raven’s Progressive Matrices measure (which is believed to capture *fluid* IQ, as opposed *crystallized* IQ—the latter refers to knowledge that is acquired over time). He suggests that the observed performance increases are unlikely to be due to biological processes given how rapidly they have emerged. Therefore, the notion that intelligence is fixed phenomenon seems less tenable in light of these findings.

<sup>4</sup> Note that this performance difference only emerged on the difficult items. There were no significant differences detected amongst the condition means on the moderate and easy items.

<sup>5</sup> However, these findings do not represent the majority of stereotype threat studies (Jones & Stangor, 2003).

<sup>6</sup> The domain identification measure (DIM) is composed of 16 items and three sub-scales that measure math, English, and general academic identification. The DIM’s sub-components have been shown to be internally consistent ( $\alpha$ ’s = .93, .90, and .75, respectively) with both the math and English sub-components also having been demonstrated to remain consistent over time ( $r = .89$ ,  $r = .56$ , respectively; Smith & White, 2001). However, the general academic identification component has been found to be deficient in its ability to remain stable over time (test re-test  $r = .26$ ). In the studies reported herein, I was only interested in the math subscale of the DIM which is composed of 10 items and is assessed on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). A typical item on this subscale is, “Mathematics is one of my best subjects” (item 2), and the scale is believed to capture the extent to which participants

value their performance on math-related pursuits.

<sup>7</sup> I decided to retain all participants that self-identified as ‘Asian American’ in all three studies ( $N = 10, 7, \text{ and } 3$ , respectively). Within the threat literature, it is customary to exclude these participants from consideration, given the possibility that the negative stereotype associated with their gender, may interact with the positive stereotype associated with their race. However, I decided to retain these participants after analyzing these data, both with and without Asian American participants, and discovering that relatively little changed in terms of the direction and magnitude of the experimental effects. Therefore, despite the concerns described herein, I decided to include these participants in all analyses across all three studies.

<sup>8</sup> My rationale behind using implicit measurement was threefold. First, although empirical support for either mechanism as a single explanatory mediator has not been definitively provided (e.g., Jones & Stangor, 2003), both stereotype activation and anxiety are assumed to play a pivotal role in the threat-performance relationship. Second, the use of implicit measurement has been known to reduce the likelihood of self-presentational concerns influencing responses on self-report measures (e.g., Greenwald & Banaji, 1995). Therefore, using this form of measurement may not only prove to be more sensitive in the detection of subtle differences if they do exist—especially if threat mechanisms operate at a pre-conscious level—but it may also be one of the most effective means of tapping into such phenomena, while simultaneously circumventing self-presentational concerns. Third, implicit measures have only been included in three published threat studies to my knowledge. Although the results of these studies were mixed, dismissing the utility of such measures—given the small sample of studies employing them—may be somewhat premature especially given their success in other domains of stereotyping and prejudice research (e.g., Greenwald & Banaji, 1995).

<sup>9</sup> Blascovich et al. (2001a, Study 2) have reported inconsistent findings regarding using self-report measures to assess perceptions of threat vs. challenge. For instance, these authors noted that when participants were confronted with a stigmatized partner—assumed to trigger perceptions of threat—they indicated that the task was more competitive and that they had exerted more effort than individuals paired with a non-stigmatized partner. However, stigmatized partners were rated as more industrious and likeable than their non-stigmatized counterparts. Such contradictions in the judgments of these participants were considered to be reflective of “compensation” strategies and were to be expected given the limitations associated with using self-reports to examine phenomenon such as stigma (p. 260). It should be noted that responses to the items designed to assess perceptions of threat vs. challenge appeared to be uninhibited by such self-presentational concerns, whereas partner ratings were clearly influenced. Thus, given the limitation of not using psychophysiological measures to assess threat vs. challenge reactivity in this thesis, it appears that relying upon self-reports seems quite appropriate in

this instance despite the potential self-presentational concerns. And although acquiring measures of threat and challenge using psychophysical measurement would be optimum—in a converging operations manner—such physiological reactivity should be evident via self-report as the Blascovich et al. study reported herein demonstrates.

<sup>10</sup> Only three participants were unable to finish the task within the 15-minute time limit. These participants were allowed to complete the measure, but only after their progress up to that point was noted. These data were analyzed both with and without these participants and none of the experimental effects were altered in terms of their direction or magnitude in either analysis. Therefore, I retained these participants in all of the subsequent analyses—with the lone exception of the reaction time measure. In addition, given that these participants were the only ones who did not finish the performance measure within the time limit, I decided not to conduct a separate analysis for accuracy—i.e., the number of items correct divided by the number of items attempted.

<sup>11</sup> Given the ineffectiveness of the gender and instructional set variables to produce any significant main effects or interactions on the Rosenberg self-esteem scale in Study 1, I decided not to include this measure in Study 2. I also failed to include a measure of task confidence perceptions since this study employed a non-performance based primary dependent measure.

<sup>12</sup> I excluded the data of two participants from further analyses based on the assumption that these participants did not take the performance task seriously. The first participant provided the same response set for several items (e.g., 1 and 2, 3 and 4, 5 and 6, and 7-9). Similarly, this participant provided a confidence score value of one for more than half of these items and posted reaction time scores for three items that were either extremely brief or equal to zero. The second participant exhibited the same type of response pattern on the performance measure as described above. In addition, this participant stated, “I had no idea what I was doing [on the math task]...,” which suggested that this participant also did not take the task seriously.

<sup>13</sup> I used data from a fall 2003 mass testing administration to generate reasonable values for mean HSGPA and SAT scores that would serve as a threshold for the self-affirmation manipulation. The means and standard deviations for this sample were  $M_{\text{HSGPA}} (N = 485) = 3.64, SD = 0.5$  and  $M_{\text{SAT}} (N = 479) = 1232, SD = 143$ , respectively. I used these average values as a threshold and chose what I believed were modest values— $M_{\text{HSGPA}} = 3.0$  and  $M_{\text{SAT}} = 1000$ —that almost everyone in our sample would have likely exceeded, but that were still plausible given the University of Maryland’s admissions criterion. I expected that participants in the *self-affirmation* condition would likely have HSGPA and SAT scores that far exceeded these modest values. This expectation was critical to the success of this manipulation given that participants in our experiment not only viewed their math performance as an important aspect of their self-concept (as evident in their math-DIM scores), but that they were also likely to view the exercise of

discussing of their GPA as self-affirming—to the extent that their scores exceeded what was presumed to be the statistical average on these measures. To the extent that their scores exceeded these values, the self affirmation manipulation was likely to be strengthened. After examining the SAT and GPA information of the sample in Study 3, I found that only 3 participants posted combined SAT scores below the 1000 point threshold (950, 950, and 980, respectively). One of these participants posted a college GPA at the 3.0 threshold, whereas the second participant posted a GPA just below this threshold (2.7). The final participant failed to post their GPA which suggest that for almost all of the participants (save the final two participants described above) the self-affirmation opportunity in Study 3 was potentially a self-affirming experience.

<sup>14</sup>Four participants were unable to finish the task within the 14-minute time limit. Once again, I allowed them to complete the performance measure, but only after I recorded their progress at the 14 minute mark. I analyzed these data, both with and without these participants, and none of the experimental effects changed in direction or magnitude in either analysis. Therefore, as in the earlier experiments, I retained these participants in all of the subsequent analyses except for those on reaction time measure. In addition, I decided not conduct a separate analysis for accuracy since these four participants were the only ones who did not finish the performance task within the allotted time.

## APPENDICES

### ***Appendix A - Consent Form***

Project Title: Problem Solvers

I am 18 years of age or older at the present time.

I am willing to participate in a research activity being conducted by Paul R. Jones at the Graduate School, University of Maryland College Park, Department of Psychology.

In this study, I will be asked to complete several tasks on a computer. I will also be asked to provide some information about my experience during the task.

The data that are gathered in this study will be treated confidentially. The data will be stored by a code number, and only the project director will have access to the master list that links participants' names and code numbers. The master list will be kept in a locked file cabinet.

There are no known risks associated with participating in this study. I understand that the benefits of this study are not intended to help me personally, but rather that the investigator hopes to learn more about the problem solving process.

I am free to discontinue participating at any time without penalty.

At the end of the study I may have my data withdrawn from the study without penalty.

I am free to ask questions. If I have questions at a later time, I may contact Paul R. Jones at (301) 405-5921 or via email at [pjones@psyc.umd.edu](mailto:pjones@psyc.umd.edu).

If I have any further questions I am free to contact the Chair of the Human Subjects Committee, Harold Sigall, via e-mail at [sigall@psyc.umd.edu](mailto:sigall@psyc.umd.edu).

---

My signature below attests to the fact that I have **read and understood** the above statements and that I have voluntarily agreed to participate in this study.

---

Signature of participant

Date

---

Printed name of participant

Date

Project Director: Paul R. Jones

Principal Investigator: Charles Stangor, (301) 405-5921, [stangor@psyc.umd.edu](mailto:stangor@psyc.umd.edu)

## **Appendix B - List of Critical Items Used on the Gender and Self-Doubt Activation Measures**

### *Word Fragments Used as a Measure of Gender Activation*

#### Gender Activation Items

1. \_\_\_ER (Gender)
2. MA\_\_ (Math)
3. \_\_\_AN (Woman)
4. \_E\_\_LE(Female)
5. SI\_\_\_\_\_ (Sister)
6. \_\_LE (Male)
7. TO \_\_\_ (Token)
8. \_\_\_MAL (Normal)
9. \_\_\_AGE (Average)

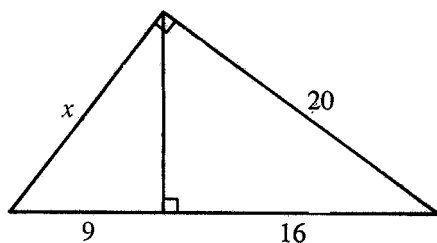
### *Word Fragments Used as a Measure of Self-Doubt Activation*

#### Self Doubt Activation Items

1. LO \_\_\_ (Loser)
2. DU \_\_ (Dumb)
3. SHA \_\_ (Shame)
4. \_\_\_ERIOR (Inferior)
5. FL \_\_\_ (Flunk)
6. \_AR\_ (Hard)
7. W\_\_K (Weak)



**Appendix C - Items Used on the Performance Task in Studies 1 and 3**



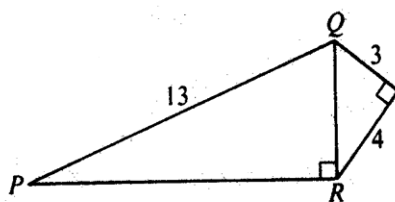
What is the value of  $x$  in the figure above?

- (A) 12
- (B) 12.5
- (C) 15
- (D)  $9\sqrt{3}$
- (E) 18

- 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.

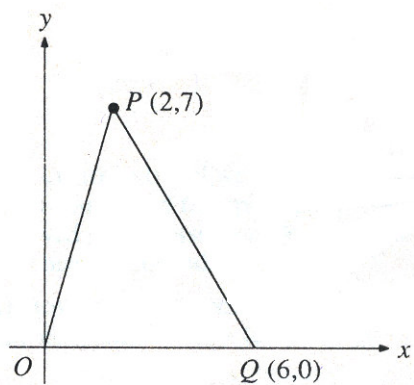
Column A

Column B



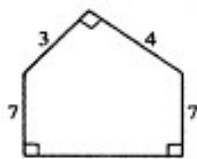
The perimeter of  
triangle  $PQR$

36



The area of  $\triangle OPQ$  in the figure above is

- (A) 6
- (B) 12
- (C) 14
- (D) 21
- (E) 42



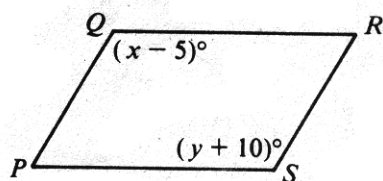
What is the perimeter of the pentagon above?

- (A) 21
- (B) 26
- (C) 28
- (D) 31
- (E) 41

- 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



$PQRS$  is a parallelogram.

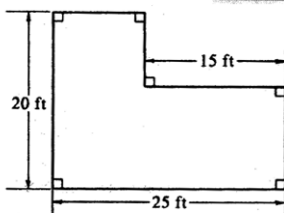
$x$

$y$

- 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

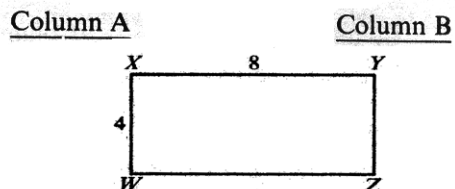


The figure represents the floor of a certain room.

The area of the floor

350 square feet

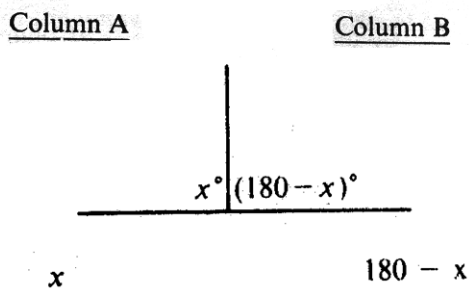
- A if the quantity in Column A is greater;  
 B if the quantity in Column B is greater;  
 C if the two quantities are equal;  
 D if the relationship cannot be determined from the information given.



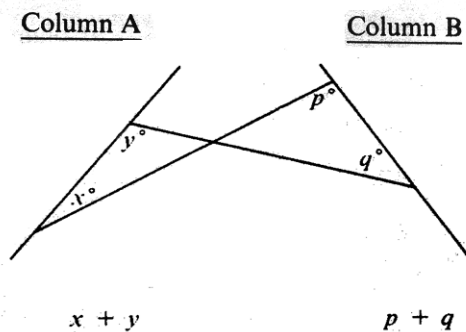
The area of a square region with a perimeter equal to the perimeter of rectangular region  $WXYZ$

36

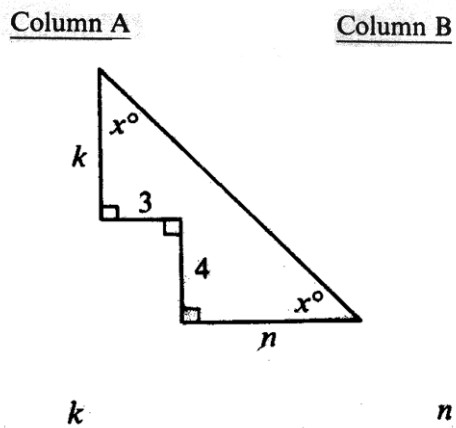
- A if the quantity in Column A is greater;  
 B if the quantity in Column B is greater;  
 C if the two quantities are equal;  
 D if the relationship cannot be determined from the information given.



- 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.



- 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.



### **Appendix D - Achievement Motivation Measure**

Using the following scale, please indicate the number that best describes how much you agree with each of the statements below.

1	2	3	4	5
Strongly Disagree	Moderately Disagree	Neither Disagree or Agree	Moderately Agree	Strongly Agree

1. \_\_\_\_\_ It's important to me that the students in my classes think that I am good at my school work.
2. \_\_\_\_\_ An important reason I do my school work is so that I don't embarrass myself.
3. \_\_\_\_\_ I like school work best when it really makes me think.
4. \_\_\_\_\_ I want to do as little school work as possible.
5. \_\_\_\_\_ I want to do better than the other students in my classes.
6. \_\_\_\_\_ The reason I do my school work is so my professors don't think I know less than others.
7. \_\_\_\_\_ An important reason why I do my work in school is because I want to get better at it.
8. \_\_\_\_\_ I want to get out of having to do school work.
9. \_\_\_\_\_ I would feel successful in college if I did better than most of the other students.
10. \_\_\_\_\_ The reason I do my work is so others won't think I'm dumb.
11. \_\_\_\_\_ I do my school work because I'm interested in it.
12. \_\_\_\_\_ I want to do things as easily as possible so I won't have to work very hard.

## REFERENCES

- Ad Council & Girl Scouts of the USA, (2004). Girls go tech: It's her future. Do the math. Retrieved March 4, 2004, from the Girl Scouts of the USA Web site: <http://www.girlscouts.org/girlsgotech/adcouncil.html>
- Aronson, J., Fried, C. & Good, C. (2002). Reducing the Effects of Stereotype Threat on African American College Students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38*, 113-125.
- Aronson, J., Lustina, M. J., Good, C. 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 Social Psychology*, *35*, 29-46.
- Aronson, J., Quinn, D. M., & Spencer, S. J. (1998a). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 83-103). San Diego: Academic Press.
- Aronson, J., Steele, C. M., Salinas, M. F., & Lustina, M. J. (1998b). The effect of stereotype threat on the standardized test performance of college students, *Prejudice* (pp. 403-419).
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, *84*, 191-215.
- 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.
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, Vol. 6. New York: Academic Press.
- 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*(12), 1262-1264.

- Benbow, C. P., & Stanley, J. C. (1984). Gender and the science major: A study of mathematically precocious youth. In M. W. Steinkamp & M. L. Maehr (Eds.), *Women in Science*. Greenwich, CT: JAI press.
- Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B. & Kowai-Bell, N. (2001a). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social Psychology*, 80(2), 253-267.
- Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social “facilitation” as challenge and threat. *Journal of Personality and Social Psychology*, 77(1), 68-77.
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001b). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, 12, 225-229.
- 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(2), 246-257.
- Boykin, A. W., & Toms, F. D. (1985). Black child socialization: A conceptual framework. In H. McAdoo & J. McAdoo (Eds.), *Black Children*. Beverly Hills, CA: Sage Publications.
- Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology*, 33, 267-285.
- Callahan, C. (1991). An update on gifted females. *Journal for the Education of the Gifted*, 14, 284-311.
- Cantor, J. R., Bryant, J., & Zillman, D. (1974). Enhancement of humor appreciation by transferred excitation. *Journal of Personality and Social Psychology*, 30, 812-821.
- Chandler, M. (Writer, Director). (1999). Secrets of the SAT [Television series episode]. In M. Chandler (Producer), *Frontline*. Virginia: Public Broadcasting Service (PBS).
- Cotton, J. L., (1981). A review of research on Schachter's theory of emotion and misattribution of arousal. *European Journal of Social Psychology*, 11, 365-397.



- Dess, N. K. (2001, Spring). Statistics: Grin big and bear it! *The APAGS Newsletter*, 13 (2), 20.
- Dickstein, L. S., & Kephart, J. L. (1972). Effect of explicit examiner expectancy upon WAIS performance. *Psychological Reports*, 30, 207-212.
- Dolly, J. P., Bell, M. E., Reynolds, A. B., & Saunders, J. C. (1979). The influence of sex and race on the test scores of research subjects exposed to research purpose information. *The Journal of Psychology*, 103, 61-65.
- Dressler, W. W., Bindon, J. R., & Neggers, Y. R. (1998). John Henryism, gender, and arterial blood pressure in an African American community. *Psychosomatic Medicine*, 60, 620-624.
- Dutton, D. G. & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, 30, 510-517.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040-1048.
- d'Ydewalle, G., Swerts, A., & De Corte, E. (1983). Study time and test performance as a function of test expectations. *Contemporary Educational Psychology*, 8, 55-67.
- Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, 11, 135-172.
- Educational Testing Service. (1994). GRE: Practicing to take the general test (9<sup>th</sup> Edition). Princeton, NJ: Author.
- Educational Testing Service. (2002). *Sex, race, ethnicity, and performance on the GRE General Test*. Princeton, NJ: Author.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row Peterson.
- Fleming, J. S., & Courtney, B. E. (1994). The dimensionality of self esteem: II. Hierarchical facet model for revised measurement scales. *Journal of Personality and Social Psychology*, 46, 404-442.
- Flynn, J. R. (1998). IQ gains over time: Toward finding the causes. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures*. Washington, D.C.: American Psychological Association.

- Fried, C. B., & Aronson, E. (1995). Hypocrisy, misattribution, and dissonance reduction. *Personality and Social Psychology Bulletin*, 21, 925-933.
- 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.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4-27.
- Graves, J. L., Jr. (2002). The misuse of life history theory: J. P. Rushton and the pseudoscience of racial hierarchy. In J. M. Fish (Ed.), *Race and intelligence: Separating science from myth* (pp. 57-94). Mahwah, NJ: Lawrence Erlbaum.
- Gross, S. (1998, July) *Participation and Performance of Women and Minorities in Mathematics*. Department of Educational Accountability, Montgomery County (Maryland) Public Schools.
- Halpern, D. F. (1989) The disappearance of cognitive gender differences: what you see depends on where you look. *American Psychologist*, 44, 1156-1158.
- Harris, M. E., & Hwang, L. C. (1973). Helping and the attribution process. *Journal of Social Psychology*, 90, 291-297.
- Herrnstein, R. E., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life*. New York: The Free Press.
- James, S. (1994). John Henryism and the health of African-Americans. *Culture, Medicine, and Psychiatry*, 18, 163-182.
- James, S. A., Hartnett, S. A., & Kalsbeek, W. (1983). John Henryism and blood pressure differences among black men. *Journal of Behavioral Medicine*, 6, 259-278.
- James, S. A., Strogatz, D. S., Wing, S. B., & Ramsey, D. L. (1987). Socioeconomic status, John Henryism, and hypertension in blacks and whites. *American Journal of Epidemiology*, 126, 664-673.
- Jencks, C., & Phillips, M. (1998). *The black-white test score gap*. Washington, DC: Brookings Institution Press.
- Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39(1), 1-123.

- Jones, P. R. (2003, April). Generating a task choice measure for stereotype threat research. Unpublished manuscript, University of Maryland, College Park, MD.
- Jones, P. R., & Stangor, C. (2003). The effects of activated stereotypes on stigmatized and non-stigmatized individuals: A meta-analysis of the stereotype threat literature. Unpublished Manuscript, University of Maryland, College Park.
- Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology*, *57*, 469-480.
- Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review*, *57*, 54-73.
- Jussim, L. & Fleming (1996). Self-fulfilling prophecies and the maintenance of social stereotypes: The role of dyadic interactions and social forces. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes & Stereotyping*. New York: The Guilford Press.
- 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* (pp. 233-265). New York: McGraw-Hill.
- Kolata, G. B. (1980). Math and sex: Are girls born with less ability? *Science*, *210*(12), 1234-1235.
- Leyens, J. -P., Désert, 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*, 1189-1199.
- Liben, L. S. (1978). Performance on Piagetian spatial tasks as a function of sex, field dependence, and training. *Merrill Palmer Quarterly*, *24*, 97-110.
- Liu, T. J., & Steele, C. M. (1986). Attributional analysis as self-affirmation. *Journal of Personality and Social Psychology*, *51*, 531-540.
- Lynch, S. M. & Graham-Bermann, S. A. (2000). Woman abuse and self-affirmation: Influences on women's self-esteem. *Violence Against Women*, *6*(2), 178-197.
- Marks, B. T. (2000). *Stereotype threat, stereotype obligation, and the intellectual test performance of African Americans and European Americans*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

- Mayer, D. M. & Hanges, P. (2003). Understanding the stereotype threat effect with “culture free” tests: An examination of its mediators and measurement. *Human Performance, 16*(3), 207-230.
- Mendes, W.B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during interactions with White and Black men. *Personality and Social Psychology Bulletin, 28*, 939-952.
- 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.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat thought salience of positive group achievements. *Journal of Experimental Social Psychology, 39*, 83-90.
- McKay, P. F., Doverspike, D., Bowen-Hilton, D., & Martin, Q. D. (2002). Stereotype threat effects on the raven advanced progressive matrices scores of African Americans. *Journal of Applied Social Psychology, 32*, 767-787.
- Midgley, C., Kaplan, A., Middleton, M., Urdan, T., Maehr, M. L., Hicks, L., Anderman, E., & Roeser, R. W. (1998). Development and validation of scales assessing students' achievement goal orientation. *Contemporary Educational Psychology, 23*, 113-131.
- National Science Foundation (1996). *Women, minorities, and persons with disabilities in science and engineering: 1996*. Arlington, VA: Author.
- Neisser, U. (Ed.). (1998). *The rising curve: Long-term gains in IQ and related measures*. Washington, D.C.: American Psychological Association.
- 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.
- Osborne, J. W. (1995). Academics, self-esteem, and race: A look at the underlying assumptions of the disidentification hypothesis. *Personality and Social Psychology Bulletin, 21*(5), 449-455.
- Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and gender differences in achievement? *Contemporary Educational Psychology, 26*, 291-310.

- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Rosenberg, M. (1989). *Society and the adolescent self-image* (rev. ed.). Middletown, CT: Wesleyan University Press.
- Rosenthal, R., & Jacobson, L. F. (1968). Teacher expectations for the disadvantaged. *Scientific American*, 218, 19-23.
- 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.
- Schachter, S. (1964). The interaction of cognitive and physiological determinants of emotional state. In P. H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior*. Stanford, California: Stanford University Press.
- Schachter, S., & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional states, *Psychological Review*, 69, 379-399.
- Schachter, S., & Wheeler, L. (1962). Epinephrine, chlorpromazine, and amusement. *Journal of Abnormal and Social Psychology*, 65, 121-128.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80-83.
- Smith, C. E. (2002). *Noncognitive variables in predicting academic performance of African American students*. Unpublished doctoral dissertation, Walden University.
- Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review*, 16(3), 177-206.
- Smith, J. L., & White, P. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational & Psychological Measurement*, 61(6), 1040-1057.

- Spencer, S. J., & Steele, C. M. (1992, August). *The effect of stereotype vulnerability on women's math performance*. Paper presented at the Psychological Mediators of Academic Underachievement and Intervention Programs, Symposium conducted at the 100th Annual Convention of the American Psychological Association in Washington, D. C.
- 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. (Ed.). (1972). *Anxiety: Current trends in theory and research*. (Vol. 1). New York: Academic Press.
- Spielberger, C. D., & Diaz-Guerrero, R. (Eds.). (1976). *Cross-Cultural Anxiety*. Washington, D.C.: Hemisphere Publishing Corporation.
- Stangor, C., & Carr, C. (1997). *Influence of solo status and task performance feedback on expectations about task performance in groups*. Unpublished Manuscript submitted for publication, Maryland, College Park.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating Stereotypes Undermines Task Performance Expectations. *Journal of Personality and Social Psychology, 75*(5), 1191-1197.
- Stangor, C., & Sechrist, G. B. (1998). Conceptualizing the determinants of academic choice and task performance across social groups. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 105-124). San Diego: Academic Press.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261-302). San Diego, CA: Academic Press.
- Steele, C. M. (1992). Race and the schooling of Black Americans. *The Atlantic Monthly, April*.
- 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 and Social Psychology, 69*(5), 797-811.

- Steele, C. M., & Aronson, J. (1998). Stereotype threat and the test performance of academically successful African Americans. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp.401-427). Washington, D.C.: Brookings Institution Press.
- Steele, C. M., & Davies, P. G. (2003). Stereotype threat and employment testing. *Human Performance, 16*, 311-326.
- Steele, C. M., & Liu, T. J. (1983). Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology, 45*, 5-19.
- Steele, C. M., Spencer, S. J., & Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology, 64*, 885-896.
- Steele, S. (1998). *A dream deferred*. New York: HarperCollins, Publishers.
- Steinberg, L., Dornbusch, S. M., & Brown, B. B. (1992). Ethnic differences in adolescent achievement. *American Psychologist, 47*(6), 723-729.
- Stone, J., Sjomeling, M., Lynch, C. I., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology, 77*(6), 1213-1227.
- Wainer, H., & Steinberg, L. S. (1992). Sex differences in performance on the mathematics section of the scholastic aptitude test: A bidirectional validity study. *Harvard Educational Review, 62*, 323-336.
- Walters, A. M., Shepperd, J. A., & Brown, L. M. (2003). *The effect of test administrator ethnicity on test performance*. Manuscript submitted for publication.
- Walton, G., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456-467.
- Wheeler, S.C., & Petty, R.E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*, 797-826.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461-481.
- Wicherts, J. M. (2004). *Stereotype threat research and the assumptions underlying analysis of covariance*. Manuscript submitted for publication (American Psychologist). University of Amsterdam.

- Wiesenfeld, B. M., Brockner, J., Petzall, B., Wolf, R., & Bailey, J. (2001). Stress and coping among layoff survivors: A self-affirmation analysis. *Anxiety, Stress, and Coping, 14*, 15-34.
- Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269-274.
- Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology, 29*, 703-709.
- Zillmann, D. (1971). Excitation transfer in communication-mediated aggressive behavior. *Journal of Experimental Social Psychology, 7*, 419-434.
- Zillmann, D. (1972). The role of excitation in aggressive behavior. *Proceedings of the Seventeenth International Congress of Applied Psychology, 1971*, Editest, Brussels.