UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DOES MANIPULATING STEREOTYPE THREAT CONDITION CHANGE PERFORMANCE GOAL STATE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

CECIL MAX SIMMONS
Norman, Oklahoma
2010

DOES MANIPULATING STEREOTYPE THREAT CONDITION CHANGE PERFORMANCE GOAL STATE

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

BY

Raymond B. Miller, Ph.D., Chair

Barbara A. Greene, Ph.D.

H. Michael Crowson, Ph.D.

Teresa K. DeBacker, Ph.D.

Ryan P. Brown, Ph.D.

For my family

ACKNOWLEDGEMENTS

This work has taken many years to complete, and could not have been finished without the grace of God who has, without exception, made a way for me to move forward when I could see no way myself.

I would like to thank Dr. Raymond Miller who has been the very essence of patience and encouragement for me over the years. His positive attitude, advocacy, availability, knowledge, and willingness to help and listen have been invaluable. I am also deeply indebted to the other members of my committee; Dr. Barbara Geene, Dr. Teresa DeBacker, Dr. Ryan Brown, and Dr. Michael Crowson. I always felt that my committee had my best interest in mind. I know that this is not always the case. This makes me even more aware of how fortunate I have been to work with all of them.

Many thanks to those at Oklahoma City Community College who have given me encouragement, and the time, needed to complete this work. Specifically I would like to thank Dr. Tamara Carter, Dr. Sonya Williams, Ms. Cheryl Wolfe, Ms. Kim Do, Ms. Daisy Mitchell, Dr. Felix Aquino, and Dr. Brenda Harrison for their efforts and understanding. It is also important for me to acknowledge the College as a whole for allowing me the opportunity to take extended professional leave. Without doubt this time away from my primary duties allowed me to complete this work.

Thanks to Dr. Nancy Matthews for allowing me to work with her students over the course of almost two years; to Dr. Chris Walker for hours of conversation and

collaboration that crystallized many things for me; and to Dr. Bryan Duke whose dissertation has been a companion and model.

Special thanks Kathy Tlapak, Janis Simmons, Joe Simmons, Thomas Tlapak, Carrie Nail, Kristin Nail, and George Clonts (my family) who have prayed for and been patient with me.

Then there are those who have borne the extraordinary burden of actually living with me during all this. Zinger, Madison, Odie, Pete, and Sam have given me unquestioning loyalty and love. You can't beat that. To Caleb and Eden, my wonderful children, thank you so much for being there for me at all times and for giving me more joy than you can know until you have your own kids. I think you are about to get an upgraded, much less cranky, dad.

And finally to my wonderful wife Valarie; thank you for your understanding, patience, help, prayers, kindness, ability to remember, and ability to forget. There is no way that this pursuit, and this document, could have been done without you. Whether in your role as Dr. Tlapak-Simmons, who knew just what to expect at each step of the process and helped me to understand myself, or as Valarie who lets me know that I am loved, you were there all the time in just the right way. I love you. I am a truly blessed man.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS……..iv

LIST OF TABLES…….vii

LIST OF FIGURES……...viii

ABSTRACT……..ix

CHAPTER

I. INTRODUCTION AND LITERATURE REVIEW 1
- Stereotype Threat 3
- Motivation 18
- Synthesis of Stereotype Treat and Achievement Motivation 25

II. RESEARCH METHODS 40
- Study questions 41
- Participants 42
- Instruments 42
- Protocol 46

III. RESEARCH FINDINGS 50
- Data Preparation 50
- Factor Analyses and Scale Reliabilities …….. 51
- Descriptive Statistics……...55
- Preliminary Statistical Tests……56
- Female Only Threat State Analyses……57
- Testing for Gender Based Stereotype Threat Phenom…..……..58
- Tests of Research Questions……...63
- Other Findings ……75

IV. DISCUSSION 83
- Stereotype Threat, Domain Identification and Achievement Motivation .. 83
- Measuring Performance Approach with Negative and Positive Valences . 88
- Threat State, Mathematics Domain Identification and Motivation 89
- Future Research 106
- Conclusion 107

REFERENCES 111

APPENDICES 118

LIST OF TABLES

Table

1 The Hierarchical Model of Approach and Avoidance Achievement Motivation with Demonstrated Antecedents and Affect…….22

2 Reliability and Scale Statistics for Motivation Factors/Sub Scales……..54

3 Normality Statistics for Measured Quantities……..…….....56

4 Math Self Efficacy Scores for Females…….....57

5 Descriptive Statistics for Math Test Analyses…….……62

6 Pairwise Comparisons for Math Test ScoresCovariate…….…...63

7 Tests of Between-Subjects Effects for Threat State and Mathematical Domain Identification on Motivational Scales (Significant Results)…….....66

8 Dummy Coding of Threat State Variable for Regression Analysis…….73

9 Standardized Regression Coefficients for Mediation Analyses of Mastery Score on Threat State and Math Score…….75

10 T-test for Significant Differences in Motivation Sub Scale Means…….78

A1 Rotated Factor Matrix for Identity Scales……...……..118

A2 Rotated Factor Matrix for Motivation Scales……..119

A3 Item Statistics for Motivation Sub Scales (Four Factor Model)……..……....120

B Descriptive Statistics for Math Test Score (for analysis where 24 was not used for missing ACT Math Scores)…….121

E1 Descriptive Statistics for Males Only for PAP+…….....146

E2 Descriptive Statistics for Males Only for Mastery……...147

E3 Descriptive Statistics for Males only for PAV…….…....148

LIST OF FIGURES

Figure

1 Effect of Identification on Performance in STT Situations (from Aronson et al. 1999)……...12

2 The Stereotyped Task Engagement Process Model (Smith 2004)……....31

3 Conceptual Model of Psychological Processes Underlying Stereotype Threat and Standardized Math Test Performance (Ryan & Ryan, 2005)……..….33

4 Scree Plot for Achievement Motivation Items…….....53

5 Adjusted Scores for Males and Females with Threat States (ACT Math Scores Covaried…….60

6 PAP+ Scores by Math ID Quartile and Threat State……..….68

7 PAP- Scores by Math ID Quartile and Threat State……..…..…68

8 Math Test Scores by Math ID Quartile and Threat State…….....69

9 Mastery Scores by Math ID Quartile and Threat State……71

10 PAV Scores by Math ID Quartile and Threat State……...…76

11 Scores for Motivation Subscales for Both Threat States……..……..77

12 Scores for Motivation Subscales for Math Identification Quartile……..…...79

13 Scores for Motivation Subscales for Math Identification Quartiles for the Enhanced Threat State……..….…….…80

14 Scores for Motivation Subscales for Math Identification Quartiles for the Reduced Threat State……..…….....81

15. The Stereotyped Task Engagement Process Model (Smith 2004)……....101

16. Conceptual Model of Psychological Processes Underlying Stereotype Threat and Standardized Math Test Performance (Ryan & Ryan, 2005)…….....102

17. Mediation of Stereotype Threat by Situation Dependent Achievement Goals……..……..…...104

ABSTRACT

This study tested whether the Stereotype Threat effect is mediated by achievement goals, in particular performance-avoidance goals. Threat level was altered before a difficult math test to observe how the endorsement by females of various achievement goal dimensions was affected. 222 people (96 females) in a pre-calculus class at a Mid-Western university participated. It was anticipated that females in the higher threat condition would show significantly more performance avoidance (PAV) endorsement than females in the lower threat condition; and that PAV endorsement would mediate the threat state – math test performance interaction. Analysis confirmed the presence of the stereotype threat effect with females in the high threat group under-performing on a math test compared to males. Women in the low threat group showed no such difference. MANOVA revealed that females in the high threat condition endorsed mastery goals at a significantly higher level than females in the low threat state. Endorsement of mastery goals mediated the threat state –math performance relationship. There were no significant results for performance avoidance. Female participants in the high threat group exhibited a different pattern of response for performance approach (positive valence (PAP+)) measures than those in the low threat group when level of mathematical domain identification was taken into account. The high threat group showed a marked increase in PAP+ endorsement as domain identification increased. Those in the low threat group had essentially constant PAP+ endorsement over domain identification. Additionally, negatively valenced performance approach achievement goal endorsement was measured for the first time.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

A continuing concern for educators and psychologists alike is the fact that performance in certain academic domains seems to be related to some degree to uncontrollable group membership. Examples of this are African American under-performance in academics in general and the relative under-performance of women in the domains of mathematics and mathematically related fields. A gap of up to a third of a letter grade between Blacks and Whites at the college level is reported by Kane (1998). Studies have shown that sex related gaps in mathematical scores may have decreased during the K-12 years, (Hall, Davis, Bowen & Chia, 1999). However, other studies find that women are still underrepresented in mathematical domains in college and beyond (National Science Foundation, 2000). This finding mirrors the result of a meta-study conducted by Hyde, Fennama & Lamon (1990) that shows the performance gaps between males and females grow throughout postsecondary education. Studies continue to show that tests such as the ACT and SAT underpredict the achievement in mathematics for women (Bridgeman & Lewis, 1994; Wainer & Steinberg, 1992). This is of concern since college admission and financial aid opportunities are based, at least in part, on the scores on these tests.

In this era of high stakes testing in common education it is important to understand how the stress of these exams affects different populations who may have to contend with factors beyond academic competence (e.g., stereotypes about race and gender). For example, while girls' grades tend to be as high or higher than boys' in classroom

settings their performance on standardized tests lags behind their male counterparts by 17-18 years of age (Willingham & Cole, 1997). This is particularly noticeable on questions that measure more advanced mathematical reasoning skills that are needed to continue into mathematics intensive degree programs in college. (Ryan & Ryan, 2005).

Two essentially independent lines of research which deal with both long term and short term achievement in academic domains are stereotype threat theory (Steele, 1997) and achievement goal theory (Dweck & Leggett; 1988, Elliot & Church, 1997; Elliot & McGregor, 2001). While stereotype threat (STT) research deals specifically with the underperformance of stigmatized groups (often on standardized tests), achievement goal theory (AGT) deals with issues surrounding successful learning and performance over the entire spectrum of academic (and other) endeavors. STT researchers have had little success in explicating the mechanisms that underlie poor performance in situations where stigma is present (Smith, 2004). However, achievement goal research has successfully shown that specific cognitive and affective behaviors result from the adoption of specific goal states. It has been suggested (Smith, 2004 and Ryan & Ryan, 2005) that performance deficits seen in stigmatized groups in certain situations could be due to the adoption of a specific (maladaptive) achievement goal structure. This study seeks to explore the hypothesized relationship between stereotype threat and achievement motivation.

Stereotype Threat

Stereotype Threat (STT) is a phenomenon described by Claude Steele and others (Steele, 1997; Steele & Aronson, 1995; Spencer, Steele, & Quinn, 1999; Brown & Josephs, 1999; Osborne, 1999; Blascovich, Spencer, Quinn, and Steele, 2001; Inzlicht & Ben-Zeev, 2003; Davies, Spencer & Steele, 2005; Brodish & Devine, 2009) in which members of stigmatized groups, who feel that their work in the stigmatized domain is personally important, under-perform when placed in challenging situations. Those affected do not need to believe the stigma, and therefore have not internalized it. What matters is the awareness of the fact that the stereotype is well known and is applicable to them in the specific situation in which they find themselves (Steele & Aronson, 1995). The challenging nature of the task makes failure, and therefore confirmation of the stereotype, a real possibility. This hypothesis is a possible explanation for underperformance data that goes back as far as 1964 (Katz, Epps, & Axelson) indicating that for African Americans simply changing the supposed context of the test changes performance. In the case of the Katz study African Americans were told that an IQ test was a measure of hand-eye coordination. This simple manipulation lead to a significant improvement in performance.

Interestingly, ideas about negative stereotypes regarding gender and race are in place by 11-12 year of age (Aboud, 1988; Ruble & Martin, 1988). This is coincident with the time that African American males begin to disengage from academics (Osborne 1999).

It is believed that the effect of stereotype threat is more pronounced when the stigmatized person is identified with the domain in question (Steele, 1997; Osborne,

Walker & Rausch, 2002; Smith, 2004). It is postulated that this is due to the fact that those identified with the domain have a personal need for success, and have a stake in disproving the stereotype (Steele, 1997).

Academic domain identity is a dimension of self – identity. It's influence is tied to how central good results within academics (or specific areas of academic study) are to a person's self –esteem. Domain identification in academic areas has been linked to greater endorsement of learning and mastery goals, intrinsic valuing of academics and self –regulation (Osborne & Rausch, 2001). All of these are indicative of higher achievement and persistence in academic endeavors. Those who are identified with a domain in academics are likely to have an active self – regulatory process in which good performance increases self – esteem (and domain identification) while poor performance results in negative consequences for self – image. In specific instances of poor performance this process/cycle will inform a person that he or she needs to try harder, stay focused, spend more time studying, or in some other way attempt to improve performance to alleviate the negative affect caused by the prior performance. If poor performance within the domain becomes chronic the individual may dis-identify with the domain entirely to protect his or her self – esteem.

Lack of domain identity disables the feedback/self regulatory process described above; so that poor performance does not have an impact on self – esteem (Steele 1997; Osborne & Rausch, 2001). Indicative of this Osborne (1997) found that African American high school students had higher self – esteem than White students despite poorer academic performance. Although this study did not directly measure domain

identity other research has found that African Americans (particularly males) are not domain identified with mathematics (Steele, 1992, see also Steele 1997).

As indicated above an individual's level of identification with academic domains is not constant, in fact it is fairly malleable and people seem to be adept at protecting self – image by emphasizing or de-emphasizing their domain identity as needed. Individuals are likely to emphasize the centrality of domains in which they or a group with which they identify have had good results; and de-emphasize domains for which personal or group results have not been seen as satisfactory (Crocker & Major, 1989; Major & Schmader, 1998). Maintaining domain identification is viewed as an important component of success as one encounters more difficult academic situations.

Domain specific stereotype threats therefore differentially affects students whose natural inclinations (high domain identification) would lead them into fields where they could serve as role models that debunk negative stigmas. It is, in fact a rather disheartening cycle. Those who could disprove the stereotype feel the most threat when engaging in these activities no matter how many times they have been successful at equally or less challenging tasks. Underperformance (compared to their own ability) has increasingly salient negative consequences to self-image, and in the end talented individuals drift away from the domain in question (see Steele, 1997).

Stereotype, when viewed from this perspective is threatening to its victims in a number of ways (Steele, 1997). The first is the thought that judgment of performance might be based on the stereotype and not on ability (a threat to the individual on *specific task evaluations*). The second threat is that the individual's performance will confirm the stereotype about the group (a threat to your *group*).

The third threat deals specifically with the concept of domain identification. Those who are identified have the skills and the interest to succeed in the area in question. Steele (1997) posits that identification is formed (at least in part) when a person has good results in the domain. This leads to good feelings about the domain, which leads to sustained achievement motivation. It has been found that those who are keenly aware of the stigma *and* are in danger of *disidentification* tend to be those seeking dominance or superiority over others as opposed to seeking mastery of the material.(Josephs, Newman, Brown, & Beer, 2003).

In Steele's view identification is not reached without help from external sources. The identified person must feel accepted by those who are already working in the area, and must feel useful in the domain. Obviously the idea of a negative stigma regarding an individual's group in a particular setting could have a negative impact about how they perceive the chances of operating effectively in the domain. Lack of support from those whom you see as peers or role models would only exacerbate this worry. In the end these concerns constitute a threat to the individual's *identity* (Steele, 1997*)*.

Among the more important predictions of the stereotype threat theory is that cumulative effects will cause those in the stigmatized group who are identified with the domain to be more likely to show deficits in performance the longer they remain in the domain (Steele 1997). This is the proposed mechanism that will eventually cause some talented people to lose their motivation, and leave the domain.

Although the effect was originally proposed with respect to under-performance of African Americans in academics (Steele & Aronson, 1995), stereotype threat seems to be common; affecting other stigmatized groups, including but not limited to, women

in mathematics (Spencer, Steele & Quinn 1999; Brown & Josephs, 1999), the economically disadvantaged (Croizet & Claire, 1998), white men in mathematics (Aronson et al., 1999), and women's leadership aspirations (Davies et al., 2005).

The negative effect of implicit or explicit stereotype situations has been observed for many groups in many domains, is robust, and it is fairly easy to manipulate. Researchers have been able to invoke (or revoke) the stereotype threat effect with relatively small manipulations in the laboratory setting. Often, the effect does not need to be intentionally activated since participants assume the stigma is there unless told otherwise (Steele & Aronson, 1995). Indeed, a change of just a few words in a long series of instructions is enough to produce or negate the effect (Osborne & Simmons, 2002; Steele & Aronson, 1995). Participants who are told that the exam they are about to take is not a test of ability (and that they will therefore not be compared to others); or that the exam has never been shown to "prefer "one group over another show marked improvement in performance. Indeed performance "gaps" of long standing such as African American under-performance on standardized tests (Steele & Aronson, 1995), and women's under-performance in mathematics (Spencer et al., 1999; Brown & Pinel, 2003) essentially disappear under simple manipulation of this type.

Generally speaking those researching stereotype threat have demonstrated that performance gaps in various domains for members of stigmatized groups can be altered by changing the way that people view the task. In one of the seminal works in the field Steele and Aronson (1995) performed five related experiments on students at Stanford University. Taken together these experiments show convincingly that by

presenting academic tasks as not being diagnostic of intellectual ability the statistically significant performance gaps between African Americans and White participants on the tasks disappeared. However, the gaps remained in the control group in which the participants were allowed to form their own conclusions about the nature of what the test was measuring. In some of these experiments no effort was made to make any situation more threatening than normal. The "high threat" group was told nothing about the diagnostic nature of the tasks. Underperformance occurred among those for whom the stereotype condition was not reduced. In one of the experiments adding a single line in the demographics section asking for the participant's race produced an enhanced stereotype threat condition. This is a fairly subtle manipulation, but it caused a dramatic drop in performance for the group that received that demographic sheet. Apparently stereotype threat is an ambient condition for those in the stigmatized group, and any reference to the stigmatized identity as a factor is enough to enhance the effect.

Spencer et al. (1999) conducted a further series of experiments in which they tested various hypotheses from stereotype threat theory on women in mathematics. Equal numbers of male and female students were selected from the freshman psychology pool at the University of Michigan. A key selection criterion was that the participants had to have completed at least one calculus course (but not have more than one year of calculus). This criterion was instituted to provide students who were "identified" in the domain of mathematics, since calculus is not required for most degrees. Two mathematics tests were used, the Advanced Mathematics portion of the GRE, and the Quantitative section of the general GRE. These tests differ considerably

in difficulty, so that good sophomore level math students would do well on one and find the other fairly frustrating. Women taking the more challenging test underperformed compared to their equally qualified male counter parts *unless* they were specifically told that the test had been shown to have no gender bias. In that case the performance gap disappeared. There was no performance gap on the easier test. When told that the harder test *had* been shown to have a gender bias the gap got significantly bigger. Male performance was not significantly affected by any of the manipulations. These results demonstrate a number of things. First, underperformance due to stigmatization is very context dependant. The **same** test elicited three different results from the point of view of a gap in performance between males and females depending on what, if anything, participants were told about the test. Second, the effect appeared for these relatively domain *identified female* students only when the test was *challenging*, and therefore threatening. When told that the test was gender sensitive the gap increased dramatically, presumably because the females were very cognizant of the fact that their difficulties would be seen as confirming the inability of females to do well on this material. It is worth noting that the test itself would not be overtly sexist in content, so that the only way those taking the test could explain the fact that women did less well would be that women are not good at the domain content.

Third, as with the study on African Americans, simply telling the participants that the test was not gender sensitive eliminated the effect. This is very interesting, since it implies that a lifetime of awareness of a stigma can be negated for a particular instance by very simple means.

These results taken together cast doubt on most genetic or hormonal explanations of under-performance of women in mathematics. If such explanation were plausible the performance gap should have stayed the same, particularly for the more difficult material.

Could it simply be that the higher threat situations invoke the effects of the internalization of stigmas that the participants have suffered under all their lives; but in general have been able to overcome? This has been a well respected view for a number of years (Allport, 1954; Cook, 1979; Gerard, 1983) In this view domain identification is less important than the internalization of the stereotype; as opposed to the idea espoused by Steele, Aronson and others that domain identification is necessary while internalization is not. To test this idea Aronson et al., (1999) sought to provoke a stereotype threat response (underperformance) in a group that is not usually associated with negative stereotypes, namely white males who are good at math. Students were recruited from the student bodies at Stanford and the University of Texas at Austin, for two separate but complimentary studies. Participants were required to be good at math (based on high Math SAT scores), white and male. Students were then given challenging math tests. In the Stanford portion of the study the high threat group was allowed two minutes to read a description of the phenomenon that Asians outperform Caucasians in mathematics. Those who had this information performed significantly worse than those who did not. This demonstrates that specific situational factors are important, even in the absence of long term exposure to stereotypical roles.

In the UT portion of the study participants were asked to fill out an instrument designed to measure domain identification. The top third were designated "highly identified", while the bottom third were labeled "moderately identified (considering the sample none could truthfully be labeled as being minimally identified). Both groups were then randomly assigned to either a high threat group (where information about Asian ability was provided) or a control group where no such information about Asian over-performance was provided (although they were told that this was a study exploring why different people performed differently in mathematics; which probably would have triggered the threat response in traditionally stigmatized groups). The results were interesting. In the low threat situation, the highly domain identified students did significantly better than their moderately identified comrades. In the high threat situation those who were moderately identified did significantly better than those who were highly identified; **and** did as well as the highly identified group did in the low threat situation (see Figure 1). This strongly supports the idea that it is identification with the domain that is key. Taken together the studies seem to confirm the position that underperformance is a situational effect caused by the immediate presence of the threat and high personal stakes due to identification.

Differential Effects of Threat with Strength of Domain Identity

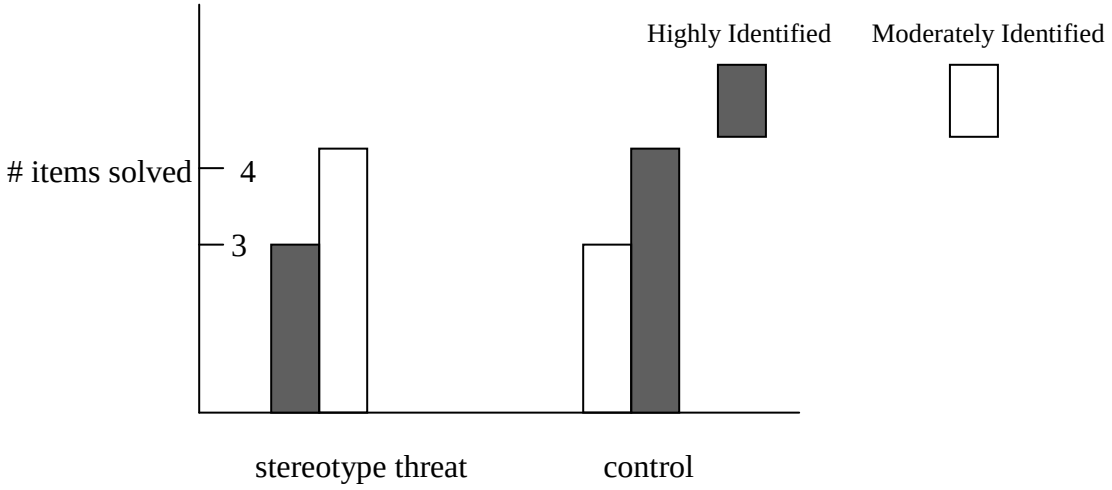


Figure 1. Effect of Identification on Performance in STT Situations (from Aronson et al. 1999).

To a large degree Stereotype Threat Theory grew out of data that indicates that stigmatized groups score noticeably lower on standardized tests than non-stigmatized groups. Steele (1997) goes on to note another disturbing trend; standardized tests **over-predict** subsequent college success for minorities, for women in technical and physical science areas, and women math *majors*. Steele concludes that something about the subsequent university environment causes this under-performance. This seems to conflict with the earlier citation from Ryan & Ryan (2005) in which it was found that standardized mathematics tests (SAT-M) under-predicted the achievements of women in subsequent college mathematics course. The devil seems to be in the details here. The difference is really in the groups being discussed. For female physical science, technical field, and mathematics *majors* the SAT-M overpredicts

subsequent success in course work. For women (of undifferentiated majors) the test underpredicts subsequent performance. The difference then seems to be a reflection of the level of domain identification of the groups. This sort of effect was noted earlier (Figure 1) when moderately identified men outperformed the highly identified in a high threat situation.

Mediators of the Effect

Steele & Aronson (1995) proposed that the threats caused by the knowledge of stereotype about a person's group in a domain they care about undermines performance through several, possibly interrelated, causes. The stigma may cause a distraction that affects performance; may result in lower expectations which reduce effort; and/or may cause anxiety, which inhibits cognitive functioning. In fact, the highly identified participants often studied in this line of research generally do *not* report a lack of confidence in the higher threat condition, *nor* do they report a higher incidence of distracting thoughts (Spencer et al., 1999). However, they do tend to reread problems more often, and in general work slower (a sign of anxiety) (Steele & Aronson, 1995; Aronson et al., 1999). Other work seems to show that high threat conditions affect cognitive load, disrupt working memory (Schmader & Johns, 2003) and reduce the ability of the afflicted to use problem solving strategies (Quinn & Spencer, 2001) as effectively as in reduced threat conditions. Additionally, Keller (2002) found that there is evidence that women doing math in a blatantly threatening situation tend to use more self-handicapping strategies.

There is little or no evidence supporting lack of effort or presence of distracting thoughts as causes of the effect. In fact, Aronson et al., (1999) reported that those in the high threat condition reported *more* effort overall although an ANCOVA analysis revealed that this did not serve as a predictor of score.

The presence of anxiety or apprehension is better supported in the experimental literature. It is worth emphasizing that the anxiety does not necessarily come from an individual doubting his or her own ability in the field, implying that the person had internalized the stereotype. It may simply be that the perceptions of others, or the possible impact the individual's performance may have on others in his or her group, causes the anxiety. Studies have shown, more directly than Steele & Aronson (1995), that there are indeed psychological anxiety markers (such as evaluation apprehension) present in situations where the stereotype threat is not explicitly debunked or is intentionally invoked (Spencer, Steele & Quinn, 1999; Blascovich et al., 2001). On the other hand, some studies have failed to find a link between anxiety and performance with threat condition (Aronson et. al., 1999). It seems likely that this may be due to the fact that in these studies anxiety was measured through "self-report" instruments. Also noteworthy is the fact that the group studied in Aronson et al. (1999) was white males. This group may not feel the same anxieties as those who have struggled with a negative stigma over their entire lives. It is also entirely possible that highly identified participants may not be willing to admit higher levels of anxiety in their domain than "usual", or don't, in fact feel such anxiety until they discover that the material is indeed challenging. Other studies have attempted to use physiological data to support the presence of increased anxiety in higher threat

conditions. Although the effect is subtle there have been studies (Osborne & Simmons, 2002; Blascovich, et al., 2001) that have shown physiological markers of anxiety such as significant changes in blood pressure, skin temperature, and skin conductance are present in higher threat situations. These studies seem to indicate that for people who are identified in a domain, the idea that they are being judged against a negative stereotype causes increased situational anxiety. This is hardly surprising.

While there is some evidence that performance anxiety may be an immediate cause of under-performance in challenging situations, there has been little success in identifying mediators other than anxiety, or indeed antecedent psychological constructs that might explain the source of the anxiety.

Intended effort and "general motivation" have been examined in the past, but there have been no statistically significant interactions between them and threat condition or performance (Steele & Aronson, 1995; Brown & Josephs, 1999; Brown & Josephs, 2000). Recently (since the inception of the present study) studies designed to test for links between threat state and specific goal dimenstions have produced significant results. Chalabaev, Sarrazin, Stone & Curry, (2008) found that individuals in a stereotyped group (female soccer players in Europe) who were operating in a threatening situation tended to adopt performance avoidance goals (although their performance was not affected). Brodish & Devine (2009) found that performance avoidance endorsement (and worry) mediated the effect of threat state on math scores for women.

No consistently significant effects have been found for self-perceived abilities or expectations of success (Brown & Pinel, 2003).

The general lack of consistent, well-defined mediators could be explained in a number of ways. First, while researchers in the field have looked for mediators since the beginning of the research (Steele & Aronson, 1995) the focus of the research was primarily on investigating the essential characteristics of the Stereotype Threat phenomena and on demonstrating performance differences caused by the effect in a broad range of domains. Mediational investigations, while performed, where not necessarily the focus of experimental design. The focus of research is now shifting (Brown & Pinel, 2003; suggested in Smith, 2004 and Ryan & Ryan, 2005; Chalabaev, et al., 2008; and Brodish & Devine, 2009) toward more sophisticated covariate analysis that examines underlying causes (particularly affects on achievement motivation).

Second, the phenomenon is clearly complex and is assumed to be situational. Both Smith (2004) and Ryan & Ryan (2005) have written review articles that attempt to summarize the complex, incomplete and often inconclusive or contradictory nature of the work done in finding mediators for the effect up until that time. In the end both conclude that the effect probably is not mediated by a single factor or construct (as suggested by Steele from the beginning). Multiple mediators are likely present in most situations and the dominant one (if there is one) is dependent on the specific situation and environmental cues.

The precise environment, circumstances, and set of cues found in a specific incident determine which of the different mediators (anxiety, self-handicapping, distracting thoughts, reduced (or "altered") motivation etc.) are triggered. Of course more than one of these may be involved, which would make measuring any of them

individually rather difficult, particularly with the kind of self-report instruments commonly used (Steele, 1997; Smith 2004; Ryan & Ryan 2005).

Smith (2004) and Ryan & Ryan (2005) both propose models that link stereotype threat and its observed effects to achievement goal theory. To quote Ryan & Ryan (2005); "The notion that individual perceptions of self and context lead to differential trajectories of cognitive engagement, affect, and performance is the essence of an achievement goal theory of motivation**"** (p. 57).

Lastly, the underlying cause may not be wholly psychological. Josephs, et al. (2003) performed a set of experiments which strongly indicate that testosterone (T) levels may be a mediator in the gender-math performance gap for domain identified subjects. Their findings show that males with high T do significantly better than males with low T in the high stereotype condition (but not the low threat condition). For females high levels of T are correlated with lower scores in the high threat case. Low T females are basically unaffected by the threat level. The authors link this result to other findings (Dabbs, 1998; Mazur & Booth, 1998; Scaramella & Brown; 1978) that correlate testosterone levels to participant concern about *dominance* and *status*. Further, the authors state that the general effect of T is to enhance feelings that *already exist*. Thus, males high in T who are aware that they are supposed to be better at math will see the test as a challenge, while females, who are aware that they are generally believed to be inferior in this domain will see the test as a threat. High T males will perform better, and high T females worse. Again, in the end the threat condition's effect is not about intended effort, but is about the motivational goal of the activity, in this case to enhance or protect status or dominance. This is in fact a performance goal

(Miller, Behrens, Greene, & Newman, 1993), and is supportive of the models proposed by Ryan & Ryan (2005), and Smith (2004).

Motivation

All motivation is not equal. To date there has been limited research done on the specific dimensions of the individual's motivational states before, during and after exposure to a situation in which they likely view themselves as a member of a stigmatized group, although there is evidence that motivational states are at least somewhat fluid and are sensitive to situational cues (Pintrich, 2000b; Simons, DeWitte, & Lens, 2000; Ryan & Ryan, 2005; Smith, 2007; Brodish & Devine, 2009).

Steele posits (1997) that identification is formed (at least in part) when a person has good results in the domain. This leads to good feelings about the domain, which leads to sustained achievement motivation. The implication here seems to be that mastery/learning and/or performance approach goals develop, but this is not explicitly stated. The fact that there is no real discussion of the different dimensions of achievement motivation is actually a bit disturbing, since “high motivation” does not always correspond to maximum effort, or even the need to do well.

Why perform at all, especially if the task is hard? What are the hoped-for consequences of performance? What are a person’s *goals*? In this case what are a person’s goals that help them achieve high levels of performance? Goal orientation/endorsement has been under investigation for some time (Nicholl’s (1984); Dweck (1986); Dweck & Leggett (1988); Pintrich & Garcia, (1991); Miller, Behrens, Greene and Newman (1993); Pintrich & Schunk (1996); Miller, Greene, Montalvo,

Ravidran, & Nicholls, 1996; Greene & Miller (1996)) There are a number of postulated goal categories; mastery, performance, future (e.g. perceived instrumentality), and pleasing an important other. All of these are possibilities in circumstances commonly associated with the stereotype effect.

A person who is mastery oriented is ultimately interested at being good at (or knowing a lot about) a particular thing for intrinsic reasons. Their satisfaction comes from eventual mastery of the behavior/domain. Mastery achievement goals focus on development of competence through task mastery (Elliot (1999); and Elliot & McGregor (2001)).

A person who is performance oriented is ultimately interested in how they do (or appear to do) compared to others (Dweck (1986); Elliot & Church (1997)). These goals focus on demonstration of competence. Elliot & Church (1997) further explicated performance goals by discussing and demonstrating approach and avoidance aspects (valences) for performance. Those whose goals are to learn faster and score higher in comparison to those around them and are eager to undertake the task are said to have performance approach (PAP) goals. On the other hand, those who want to avoid the implications of performing badly in front of (or compared to) others and seek/manufacture some excuse for poor performance or avoid comparison all together are said to be pursuing performance avoidance (PAV) goals.

Each type of goal has been linked to certain antecedent beliefs about the nature of learning/intelligence and to consequences for behavior/affect during learning and demonstration of learning (particularly when challenged). See Table 1 (page 22) below. In fact one of the main reasons that positive and negative valences were

introduced into achievement goal models was to account for the wide disparity in behavior and affect seen in those reporting simply "performance goals" (Elliot, 1999).

The antecedents associated with a particular situation lead to an inclination toward approach (positive) or fear of failure (negative), which, along with situational imperatives determines the way that competence is defined (in terms of mastery or compared to others) and ultimately the various self-regulatory strategies displayed. When the "selection" of performance goal structure is well matched to the circumstances the self-regulatory strategies are simple, and likely to be effective (as defined by the individual). Examples would include mastery endorsement in a situation where approach and mastery is the priority; or PAV in a situation where avoidance is called for and possible. Complex self-regulation is also possible particularly when there is a mismatch between valence and action. For instance it is possible to fear failure and yet feel the necessity to approach a performance, and do well. It is also possible that a person could be striving to attain or retain mastery level ability out of fear of incompleteness or of demonstrating declining ability. Negative affect and less than optimal performance are to be expected in such a situation (Elliot, 1999; Elliot & McGregor, 2001).

A theoretical construction with two dimensions such as valence and competence comparison also allows for discussion that includes the possibility that any/all goal structures are present in the same individual at the same time to some degree (Elliot & Church (1997); Elliot (1999); Elliot & McGregor (2001). Previous to these formulations (e.g., Dweck (1986); Nicholls (1984)) achievement goal theory did not

account for valence and mastery and performance orientations were implicitly viewed as fully independent entities.

Refinements to the trichotomous model (Table 1) have been suggested. Elliot & McGregor, 2001 suggested a full 2x2 model for achievement goals. This model has two dimensions (mastery or performance), which indicate how the individual perceives the way outcomes will be judged (competence as absolute or relative (to oneself or to others)); and two means of engagement -valences (approach or avoidance) which indicate whether the person is trying to "succeed" or trying to "avoid failure". In the resulting model one could be:

Mastery approach (MAP) in which competence is measured with respect to content (absolute) and the means of engagement is positive.

Mastery avoidance (MAV) in which competence is measured with respect to prior ability and the means of engagement is negative (a person is fearful of losing ability or knowledge).

Performance approach (PAP) in which competence is measured with respect to others and the means of engagement is positive (the person is eager to perform).

Performance avoidance (PAV) in which competence is measured with respect to others and the means of engagement is negative (the person is fearful of not performing well compared to others and wants to avoid meaningful performance).

Table 1.

The Hierarchical Model of Approach and Avoidance Achievement Motivation (Elliot & Church, 1997) with Demonstrated Antecedents and Affect (Elliot & McGregor, 2001).

Goal	Motivation	Antecedents to Goal Structure	Affect When Challenged
M	Achievement	High competency expectations; Work mastery; self-determination; Need for achievement; Class Engagement	Effort; Persistence; Mastery; Expectation of success; Increased deep processing
PAP	Achievement	High competency expectations; Competitiveness; Fear of Failure; Need for Achievement	Effort; Persistence; Competitiveness; Fear of failure; Increased surface processing; Good overall performance
PAV	Avoidance	Low competency expectations; Fear of Failure; Low self – determination; Belief in fixed nature of ability (entity theory)	Fear of failure; Affect; Anxiety; Distraction/Poor self – regulation; Lack of effort; Increased surface processing, decreased deep processing, decreased overall performance

This model seems plausible (and nicely symmetric). The studies reported in the 2001 paper show that the four constructs are measurable and have specific sets of antecedents and consequences. The MAV construct was the "new" piece in what had been a three piece model (M, PAP and PAV, see Table 1 above). It was examined in detail. The antecedents were found to be ; fear of failure, low self – determination, belief in ability as a constant (entity theory), and competence focus. The negative consequences were found to be associated with disorganization, worry, anxiety and emotionality. These are distinctly negative with regard to performance; however the study also indicated that those who endorse MAV in a specific circumstance are likely to adopt MAP or PAP goals in subsequent activities.

Conceptually MAV is a mix of the traditional mastery structure and PAV. Indeed it shares antecedents and consequences with both. MAV is rather conceptually difficult in some respects. The traditional definition of mastery goals (as Elliot points out in the 2001 work) is focused on development of competence. MAV does not seem to focus on development of competence. The focus is on the retention of competence (as defined by self –comparison of some sort or completion of a task that has been successfully done in the past). Examples cited by Elliot & McGregor include those who strive to avoid making any mistakes (perfectionists) or those in the latter part of their careers who focus on not performing worse than before. These examples are clearly negative in valence (avoiding something) but are not so clearly mastery in intent since they are focused on performance (finishing something perfectly, maintaining a level of competence on specific tasks).

Further, the two valences (approach and avoidance) do not necessarily map simply onto mastery and performance competence measures (as in the 2 x2 model above). For instance, someone in a PAP mode might be working hard to avoid failure or to make sure that others do not outperform them. This could be termed PAP with a negative valance (PAP-). Performance approach, where a person is trying to confirm that he or she is better than others (the traditional view of PAP), would be PAP with a positive valence (PAP+). In an overtly competitive endeavor this would be the difference in striving not to lose and striving to make sure to win/dominate. This PAP+/PAP- dichotomy was first hinted in Elliot and Church's 1997 article on hierarchical modeling of approach and avoidance achievement motivation. It was further explicated in Elliot's 1999 seminal work on the subject (in which he first suggests MAV as well). Clearly PAP- is a complex state, in which the person is approaching something out of fear. As mentioned above one would expect that task success would be less likely in this circumstance. The presence of both positive and negative valences in PAP implies that if the "balance" of positive and negative antecedents (or affect) changes (for instance in a situation where challenge is perceived as threat), a person could easily find herself shifting toward a more PAP– or even ultimately a PAV mode. The possibility of a PAP- modality does not fit well within the proposed 2x2 framework, although the existence of such a mode appears reasonable. To date, to my knowledge no attempt has been made to determine whether positive and negative PAP valences are separable and measurable.

As Elliot states, "Accumulating evidence indicates that persons process most, if not all, encountered stimuli in terms of valence and do so immediately and without

intention or awareness.", and that "this automatic valence-based processing is presumed to instantaneously evoke approach or avoidance predispositions."(Elliot & McGregor 2001, p. 502).

This fits well with situations where stereotype threat is present. In those cases cues (which are not necessarily blatant) seem to allow the individual to make intuitive, rapid, and perhaps unconscious decisions that change the way they approach a situation.

Synthesis of Stereotype Threat and Achievement Motivation

The complicated interplay of academic, social and self–image variables present in stereotype threat situations has been investigated in achievement goal theory literature. The adoption of a particular goal structure has been shown to result in the activation of a pattern of behavioral and phenomenological constructs (see Table 1 above and Ryan & Ryan, 2005). Both achievement goal theory and STT theory are concerned with performance differences in groups with similar backgrounds and experiences. Why does one perform well and another fail when they have similar abilities and backgrounds in the domain in question?

Approach or avoidance aspects (or positive and negative valences) of an individual's goal state are thought to be fluid, and based on situational cues that tell people whether they should fear failure (or not). The "definitional" aspect of the goal endorsement (e.g. mastery vs. performance) is probably somewhat more stable since in general, it reflects how a person perceives learning and its purposes. However, not all activities leave this interpretation of "learning" up to the person. Sometimes

activities are overtly performance (or mastery) in nature. Indeed Elliot & his co-authors have shown that performance and mastery aspects of goal selection are susceptible to experimental manipulation (Elliot 1999).

Standardized tests are overtly performance activities (they evaluate on a normative scale and the results are presented as percentiles). Some classrooms continue to be organized in a similar matter, wherein a person is "ok" as long as he or she is not dumber than average. These overtly performance situations are the very settings in which stereotype threat conditions seem to operate.

It is hard to see how someone with a generalized tendency to endorse performance avoidance could have made it to the point where they could be classified as "identified" with the domain. They should have withdrawn from the domain long ago (the ultimate form of avoidance). However a performance *approach* person (particularly with a positive valence) might make it, especially if they were very gifted in the domain. However, once they reached the point where things were not easy anymore, how would they respond? In academic domains eventually the realization dawns on a person that this is hard "even for me", that there is someone out there "smarter" than they are…they learn faster, and just seem to do better. At this critical juncture, highly identified, performance approach individuals will find themselves in a situation they care about, work hard at, and value success in, but will need to be able to explain why things are not as easy for them as they used to be or why someone else might be outperforming them. For someone who is status conscious the continual possibility of being seen as "losing their edge" or as being just average would be difficult to bear. In the overtly performance settings associated with stereotype threat

(e.g. standardized tests, upper-class and graduate school exams (particularly in mathematics and related fields) it seems reasonable that this set of circumstances could cause a person's means of engagement to shift toward avoidance rather than approach, and his or her overall state to be performance avoidance (or PAP- as an intermediate state) rather than mastery (or *PAP+).* That sort of sudden "shift" in motivation might well be enough to cause anxiety and affect performance. The negative performance would then affect the way the individual viewed the next such evaluation. The effect is cyclic, negative, and could lead to dis-indentification.

The path to disidentification may also be connected to distal goals (future goals) as opposed to immediate performance goals associated with the task at hand (although the two are linked). Elliot (1999) proposes that there are two important factors (antecedent types) that govern proximal achievement goal adoption; these are perceptions of competence and the individual's "underlying needs" which could include future goals. Miller & Brickman (2004) present a model of Future-Oriented Motivation and Self-Regulation in which they posit (among other things) that the perceived instrumentality (the degree to which a particular activity is perceived to be important to a distal but important goal) is a significant factor in whether the person operates under mastery or performance approach goals (with high instrumentality associated with mastery goal endorsement). Greene, Miller, Crowson, Duke & Akey (2004) show in a path analysis that the link between perceived instrumentality and goal type is significant (two-tailed) for mastery and nearly significant (reaches significance in a one-tailed analysis) for performance approach. Part of the underlying reasoning of this relationship is the idea that activities that are judged as highly

instrumental to the attainment of long-term goals carry with them a higher than usual self-evaluative (including efficacy) importance, and can affect the planning of proximal activities to reach the distal goal. Mastery goals and competence related activities and effort would result from active distal goals. Success in these activities reinforces the individual's choice of long term goals (and would presumably enhance domain identification). Failure calls into question those goals and the "plan" to get there. Domain identification (for those in school) is almost definitionally linked to future goals. "I want to be a mathematician" is not easily separated from "I want to get a degree in mathematics" for those on an academic path.

These scenarios seem reasonable for anyone who is attempting difficult tasks within their domain. Why should some groups be affected more than others? First, stigma is differential. One group (the stigmatized) sees the stereotype as a negative thing (whether they believe it or have even personally experienced it or not), while other groups are unaffected by it, unconscious of it, or perhaps even find it of some positive value. In the case of stereotypes of women in mathematics, males fit into the second group (as perhaps would females who were not acculturated to this stigma). These groups will view the impending activity differently; and therefore their "antecedents" to goal choice will differ, resulting in goals and/or magnitude of endorsement of those goals. This then influences the level and effectiveness of self regulation strategies (Pintrich 2000; Miller, et al., 1996).

There are also experimental findings from motivation literature that indicate certain groups are more prone to adopt avoidance valenced goals (PAV). These include women, ethnic minorities, and people from lower socio – economic

backgrounds (Elliot, 1999). All of these groups face negative social or academic stereotypes of the kind reported on in the stereotype threat literature. Clearly then the stigmatized group would seem more prone to suffer from negative affects in a threatening situation more often, and perhaps in a more personally meaningful way. The path to negative outcomes/frustration and disidentification discussed above is simply wider and steeper for members of those operating even under mild "disadvantages" within their chosen domain.

Many students experience the feeling that they are getting "dumber" as they continue to move deeper into their chosen field. The material becomes more difficult, the activities become more complex, and the peer group becomes more "elite". Why should this effect be more pronounced in individuals who are part of a group not normally associated with the domain? For those who are not stigmatized there is no worry that failure will really impact anyone beyond the individual (or their families), or for that matter effect anything beyond a limited part of what is defined as the self. Consider a white male in mathematics. Failure to achieve up to his potential is likely to be devastating, but no one blames it on being white, or male. His white-maleness was not in doubt. Nor does his failure mean that other; younger white males will be discouraged from doing math. It is really about him, and the domain. For a female in the same situation however, there is an added burden For her it isn't just her and the domain (which is bad enough), it is her, the domain, her identity as a female, and the impact of her performance on all those who know she has chosen this route. There is simply more pressure to perform (and perform well compared to others), or get out early.

Models Linking Stereotype Threat and Performance Avoidance Goals

The reasoning presented above leads to the following models proposed by Smith (2004) and Ryan & Ryan (2005). These articles, which both focus on PAV and its role in the stereotype threat effect both provide excellent reviews of the dysfunction in cognition and affective problems that result when a person is working under performance avoidance goals. To summarize, PAV leads to self-handicapping, poor study strategies (Elliot, McGregor & Gable, 1999), and minimal risk taking (not considered in STT research yet (Linnenbrink & Pintrich, 2000)). With the noted exception these are essentially the suspects that have been tested for (without much success) in STT experiments. It has also been shown that PAV changes feelings about interest (reduces them (Elliot & McGregor, 2001)), feelings about anxiety (enhances them (Elliot & McGregor, 1999; Middleton & Midgley, 1997)), and feelings about efficacy/confidence (reduces them, (Middleton & Midgley, 1997; Schunk & Pajares, 2001)). Again, these affective variables have long been implicated in the Stereotype Threat effect. There is therefore good convergence between suspected processes in STT and demonstrated ones in Achievement Goal research.

Smith's Task Engagement Process

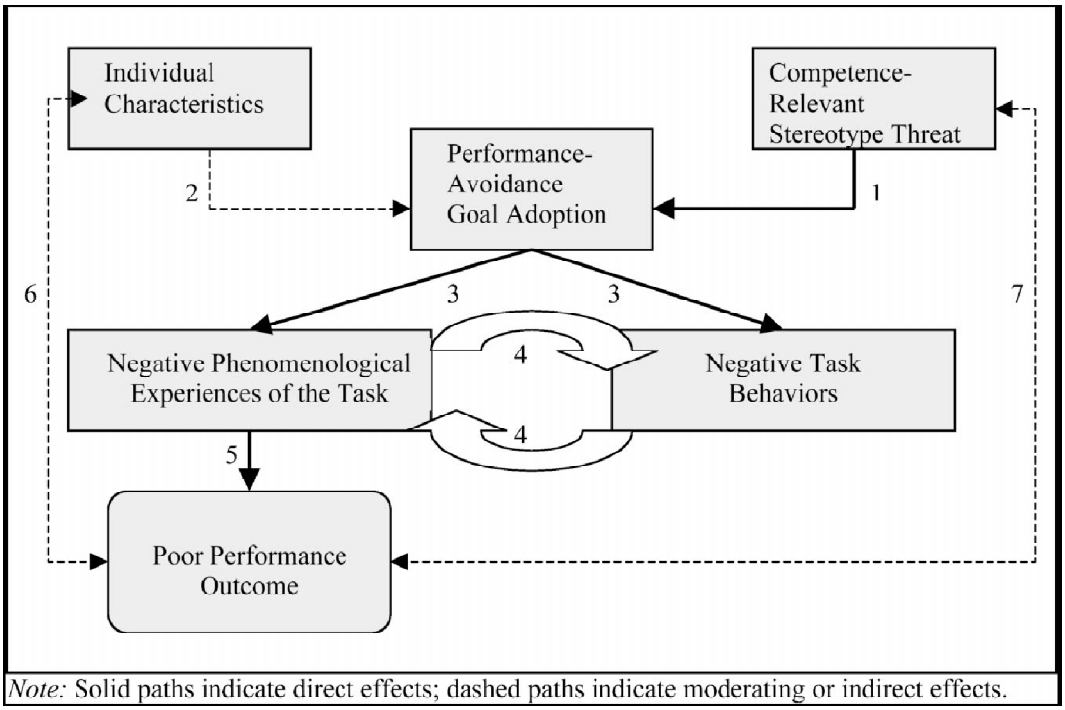


Figure 2. The Stereotyped Task Engagement Process Model (Smith 2004)

This model is a straight - forward representation of the process of goal state endorsement and the subsequent activation of specific self – regulation patterns, affect and behaviors discussed above. It is fairly simple in that it views PAV as the only direct route connecting STT to poor performance. Although this model is simple it is reasonable to postulate that situational cues about expected performance (knowledge of and salience of stigma; how it will be interpreted by others (stereotype threatens the individual and the group); and the challenging nature of the performance can lead people to adopt an achievement goal with some dimension of avoidance. In its mildest form the direct threat of the stigma could simply change the goal of the performance to

"not being outperformed" by the supposedly superior group (e.g. PAP-). In a more extreme case the avoidance could take the form of PAV, where the goal is to deflect personal responsibility for general poor performance. If these situational cues are causing a preference for goals with some manifestation of avoidance among stigmatized groups does debunking remove the threat of confirming the stigma and let people be PAP+? Can differences in any negatively valenced performance structure linked to stereotype threat condition be shown experimentally (particularly PAV as in the model)?

Smith (2004) discusses results of an interesting but rather weak preliminary experiment in which females in a math test situation that was threatening predicted that they would not do as well as the males predicted AND females were more likely to adopt PAV goals than males were.

The Ryan & Ryan (2005) model rests on the same basic set of observations that link Stereotype Threat effects and the results of PAV goals. It is in some ways more linear that the Smith model.

Ryan and Ryan Model of PAV Mediation of Stereotype Threat on Math Performance

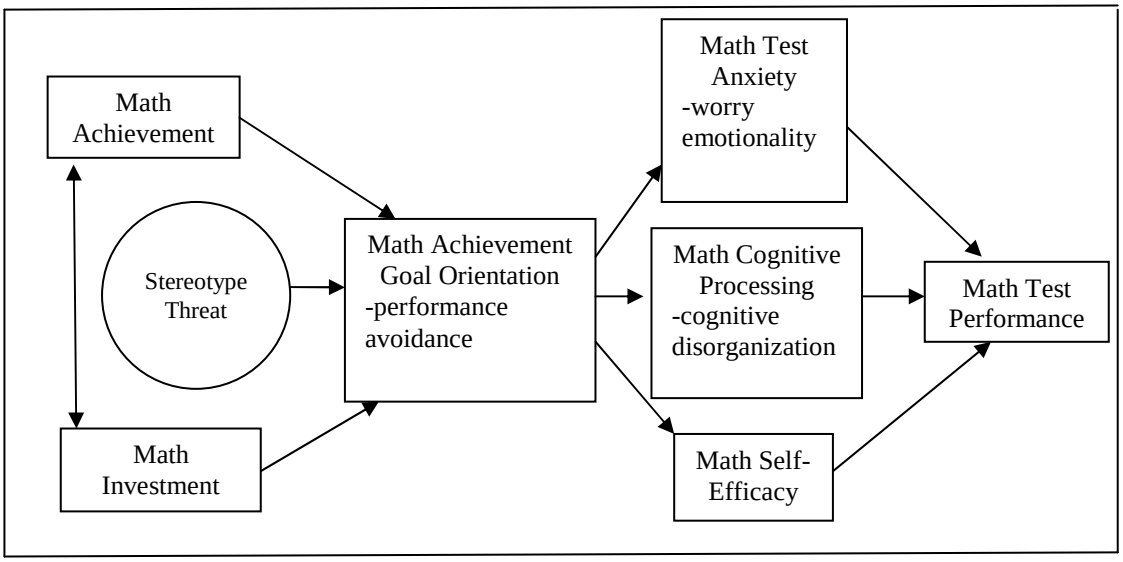


Figure 3. Conceptual Model of Psychological Processes Underlying Stereotype Threat and Standardized Math Test Performance (Ryan & Ryan, 2005)

As in the Smith model, a situation that is threatening from a stereotype point of view is not the only factor in determining the performance goal state. In both models individual characteristics as they relate to experiences in mathematics play an antecedent role. Of course from the view of Stereotype Threat research, the affect is dependent not only on the threat level, but also on the person's identification with the domain. So STT researchers might lump the "stereotype task" and the individual characteristics box(es) together to determine whether the person involved was susceptible to the threat.

Again it is of note that neither model proposed above includes the possibility of PAP-. This is due, almost certainly, to the fact that the PAP+ and PAP- have not been measured separately, and therefore differences in affect and behavior have not been

tested. It is also noteworthy that no mention of mastery avoidance (MAV) is made in the models although at least one study (Elliot & McGregor, 2001) had shown that MAV was measurable. This could be due to the somewhat problematic definition of MAV or to an assumption that the situations in which STT is relevant are always performance based and therefore would only affect performance goal structure endorsements.

Studies Linking Threat, Motivation and Performance

How does one view a difficult performance situation? The antecedent cues inherent in the type of activity, meaning of the activity as well as a person's past experiences determine her goals for that performance. Hopefully in the case of mastery oriented persons who find themselves in a manifestly performance situation their quest for mastery has in fact prepared them well for the situation and their performance endorsement will be strongly approach in nature.

Theoretical links between threat, motivation and performance seem clear. In the past few years models have been introduced that explicitly link stereotype threat to PAV goals and underperformance. Is there existing data to support the general concept and/or the specific models proposed by Smith and Ryan & Ryan?

In the field of athletic performance Wrisberg (1994) found that difficult situations trigger arousal, which leads to anxiety (as opposed to other kinds of affective arousal) when people feel that their skills are not up to the challenge. These emotional responses are statistically linked to meta-motivational states. There are two states: telic; serious goal oriented performance (performance goal) and paratelic (playful or

activity oriented…AKA mastery?). The noted affect seems to be less performance inhibiting in those who are in the paratelic state. These people see the difficulty as a challenge, while those classified as telic see it more as a threat, with more intense affective results that lead to distraction and loss of focus. While Wrisberg's study does link situational threat (seriousness of performance) to motivational states and results of activities it does not attempt to measure valence (approach or avoidance) and so, while suggestive it does not test the specific links suggested by the models presented by Smith (2004) or Ryan & Ryan (2005).

Smith (2006) has reported the results of a set of studies to investigate the links between stereotype threat and goal type. One study found that women in the higher stereotype threat condition reported a significantly higher endorsement of PAV goals than did women in the reduced threat state. A second study showed that women in the high threat state were statistically more likely to endorse PAV goals than were males in the same state, and that PAV endorsement mediated the effect of gender on math score expectancy (for the high threat case). No statistically significant results were noted for other goal types (PAP, M). While these results support the model presented by Smith in 2004 (see above) the studies have some drawbacks. First, the number of participants was small and there was no attempt to take inherent ability in the field into account. In the case of the first study where women were operating in high and low threat conditions the total number of participants was only 20. The second study involving males and females had a total of 13 female and 15 male participants. The statistical power of these studies is low and the chance for error is increased due to the low number of participants. Additionally the studies' methodology did not call for

participants to actually take a math exam. Rather they were asked to estimate their performance after looking at the exam. Thus, the stereotype threat effect was not clearly shown to be in operation since even though the expectancies were different there were no actual math scores to compare. Additionally, this methodology is very similar to that used to measure mathematical self-efficacy (Bandura, 1995) and it quite possible that mathematical self-efficacy is not an accurate measure of actual mathematical performance.

Smith, Sasone, & White (2007) conducted a series of experiments in which they tested various aspects of the relationships between stereotype threat and achievement goals. Interestingly the authors chose to focus on the omnibus "strength of achievement motivation" (rather than domain identification as suggested by stereotype threat theory) differentiating groups based on whether they were highly achievement motivated (based on survey results) or not. This study also intentionally chose women who had not yet declared a major in mathematics or related field (computer science) because "current computer science students may have already overcome any detrimental effects that stereotypes have on domain interest" (p 101). They also chose to focus on task interest rather than actual performance results, and although they manipulated the women in mathematics stereotype threat level the actual field from which they drew the task instrument was computer science. They found that among women who were assigned a PAV conditions those who had higher achievement motivation and who were in the higher stereotype threat condition showed less interest in the task. Additionally it was found that women higher in achievement motivation were more likely to adopt a PAV modality that women who were lower in

achievement motivation when in the higher threat state. While the results are interesting (particularly the spontaneous adoption of PAV mentioned above) and suggestive, the study has its drawbacks when viewed from the perspective of the model proposed by Smith in 2004 (Figure 2). The 2007 study's primary goal is to investigate the effect of PAV goal adoption on task interest in a stereotype threatened situation rather than on actual performance. It investigates the mediational effect of PAV connecting task absorption to interest, rather than the mediational effect connecting threat condition to performance, and it uses overall achievement motivation more or less in place of any indicator of domain identification (in fact choosing people outside the field who were not necessarily identified with it). While clearly overall achievement motivation and domain identification are related the two are not co-definitional. A person could be interested in a task without being identified with the domain. The focus of interest could lie outside of the "intended" domain. For instance a subject may have little interest in the field of computer programming but be interested in the task because of the novelty of the situation or an interest in psychology. Given that the study was done (intentionally) using participants who had not declared a major in computer science the above scenario is likely.

Chalabaev et al. (2008) showed that for female soccer players in Europe, an enhanced stereotype threat state (that women soccer players are not as athletic or technically capable as men) was activated. Relative to a control group that did not have this stigma activated women in the high threat groups showed a decrease in ability, and an increase in PAV endorsement relative to PAP endorsement (the measure was PAV-PAP score). However, this change in goal endorsement pattern

was not significantly related to actual performance thus failing to show a mediational relationship between threat state, PAV endorsement and performance.

Finally Brodish & Devine (2009) conducted an experiment on women in mathematics in which the mediational effects of PAV and worry on threat state and mathematics performance were tested. Women under threat and not under threat reported their test-related achievement goals and then took a difficult math test. Women who participated in the study were chosen from an introductory psychology pool based on evidence that they were good at mathematics (scores above 26 on the ACT math test (or equivalent)) and were aware of the negative stereotype regarding women in mathematics (as evidenced by their responses to questions regarding the importance of being a woman, and knowledge (not necessarily endorsement) of the stereotype). Assuming that math ability is congruent with math identification this selection procedure should have resulted in a group that was very sensitive to the effects of stereotype threat. There were 101 participants. Results showed that participants in the high threat group were significantly more likely to endorse PAV goals, and trended toward being less likely to endorse mastery goals ($p < .07$). Further PAV goals were shown to mediate the effect of threat state on subsequent math performance. This is a very compelling study, however it employed a technique of inducing the high threat state that both drew attention to the stigma in the heightened threat state and reframed the meaning of the performance by state. The high threat state was told that they were taking a test to measure their quantitative ability (and were reminded of the stereotype of women performing poorly compared to men in mathematics). The low threat state was told that the assessment was testing memory

capacity (and were not reminded of the negative stigma regarding women in mathematics). Given the supposed sensitivity of individuals to antecedent cues as they approach an activity it is possible that reframing the meaning of the activity could be affecting the goal selections of participant. Since no males were included in this study an analysis of this potential affect is not possible. A design that manipulated only the threat level would not have this complication.

CHAPTER 2

RESEARCH METHODS

A consistent finding in Stereotype Threat (STT) experiments is that long-standing gaps in academic performance are significantly reduced or eliminated by re-framing the meaning of the performance. (Steele, 1997). This effect has been observed for African Americans in general academic domains (Steele & Aronson 1995), and women in mathematics (Spencer et al., 1998; Spencer & Quinn 2001; Brown & Pinel, 2003). This re-framing to lower the threat of stereotype takes the form of an explicit statement that the activity is either: not going to be judged in such a way that the stereotype is a factor (see for example Steele & Aronson, 1995); or has never shown a result typical of the stereotype (for example Brown & Pinel, 2003).

Recently, it has been suggested that goal orientation models can be used to explain the stereotype threat phenomenon through the adoption of a performance avoidance (PAV) goals (Ryan & Ryan 2005; Smith 2004). These models suggest a high level of "fluidity" of goal endorsement. This is implied since the stereotype threat effect is clearly highly situation sensitive and the effect of the threat (if any) is easily manipulated. Goal endorsement (specifically approach or avoidance) must also be easily alterable if it is to explain part or all of the stereotype threat effect. An experiment where the level of threat is altered and subsequent strength of PAV responses are measured could be used to explore this possibility.

Study questions

1) In a situation where performance on a mathematical test is linked to stereotype threat does threat condition significantly affect the performance avoidance goal endorsement of female participants (but not male participants)?
2) Is performance avoidance endorsement a mediator in the stereotype threat effect on performance?
3) Does stereotype threat have an influence on endorsement of other goal types (M, PAP, PAP+, and PAP-)? If so is there a mediational effect between threat level and performance for that goal type?

The independent variables for this study were stereotype threat condition (STC), mathematical domain identification (math identification); gender identification; and sex.

Stereotype threat conditions were "enhanced" in which participants were told that the tests in question show gender differences with males doing better than females; or "reduced" in which participants were told that no gender difference has been noted for these tests.

The dependent variables were the participants' scores on a math test; a math self-efficacy test; and on motivational goal subscales for mastery (M), performance avoidance (PAV), performance approach with a positive valance (PAP+), and performance approach with a negative valance (PAP-).

Participants

Participants were selected from the student body at a university in the mid-west. Selection criteria were:

a) Mathematical background: All participants were enrolled Elementary Functions, a prerequisite for the calculus sequence for engineering and physical sciences. This was taken as an indication of participant identification with a mathematically related domain since a calculus course is not a General Education requirement in the state. Anyone enrolled in this course can be assumed to be either a major in a mathematically oriented field (engineering, mathematics, physics, computer science, etc) or simply intrinsically interested in mathematics.

b) Race: Everyone enrolled in the course discussed above was allowed to participate, however African Americans males were eliminated from the analysis due the confounding nature of known stereotypes regarding African Americans (particularly males) and academic performance.

Instruments

(All instruments can be found in Appendix D)

Math Test. A 20 item math test (D5) from Inzlicht & Ben-Zeev (2003), with questions taken from the mathematical portion of a GRE test guide (Educational Testing Service, 1994). These questions had appeared in previous Graduate Record Exams and were answered correctly by an average of only 36.6% of test takers. The questions were presented as a single test (20 minute time limit). This (or a very similar) method for generating exams for STT experiments is common in the literature

(Inzlicht and Ben-Zeev, 2003; Schmader & Johns 2003; Quinn & Spencer 2001; Keller 2002).

Achievement Goal Survey. An eighteen item Achievement Goal Questionnaire (D4) with fourteen items adapted from The Patterns of Adaptive Learning Scales (PALS) (Midgley et al., 2000). The items adapted from PALS (revised pool, updated in 2000) were mastery (five), performance avoidance (four), and performance approach (five). The performance approach items from PALS were all positively valenced. Participants respond on a five point Likert Scale where 5 was strongly agree and 1 was strongly disagree. The remaining four items were intended to measure performance approach with a negative valence. They used the same Likert Scale structure as the items from PALS and were generated based on the ideas presented by Elliot in his 1999 publication. The fundamental idea behind this goal type is that "the desire to avoid failure is strategically served by striving to attain success."(Elliot, 1999 p. 174). Scale items were developed based on the premise that the participant would work hard on the test due to a concern that they would not be able to perform as well as they might hope on the upcoming math test. The PAP- items presented to participants were:

a) I am concerned about performing poorly on this math test so I will work really hard at it.
b) My fear of not performing well on this test motivates me to try harder.
c) I will work hard on this math test because I am concerned that I can't do the problems.

d) I will try hard on this math test because I am worried that I might not perform well.

Threat Manipulation. Participants (in both threat states) were informed that their performance would somehow be compared to others (For example, directions on the goal survey for the enhanced threat group read in part, "*Standardized tests evaluate you based on how you perform relative to other students who took the same exam. The math test that follows this survey is one such exam that shows that men perform better than women.").* The directions for the reduced threat group read, "*Standardized tests evaluate you based on how you perform relative to other students who took the same exam. The math test that follows this survey is one such exam that shows women and men doing equally well."*

Instrumentation exists to measure the relative strengths of mastery approach and mastery avoidance; however, due to the fact that the mastery avoidance state is not well explicated (and to the overtly performance nature of the task) an instrument was chosen that does not differentiate between the two proposed sub-states.

Gender and Mathematical Identification Survey. A ten item instrument (D2) that measured identification with gender (four items) and the domain of mathematics (6 items) measured on a seven point Likert scale, with seven indicating strong agreement with the statement.

The gender identification items were adapted from an instrument found in Schmader (2002), which in turn was adapted from Luhtanen & Crocker (1992). The adaptation made for the current study was to make all items "positively" phrased. For instance, "My gender is *not* important to my sense of what kind of person I am." became, "My gender is important to my sense of what kind of person I am." This was done based on early results that indicated that the negatively phrased items scored did not score "oppositely" to similar items phrased in a positive way. (See statistics for this subscale reported in the Research Findings chapter below).

The mathematical domain identification subscale consists of items from two sources. Three items are from a math identity scale found in Smith, Morgan, & White (2005). The other three math identity items are from the Math Identification Questionnaire (MIQ) (Brown & Josephs, 2000). Statistics reported in the research findings section are for this combination subscale.

Math Self-Efficacy Questionnaire. A four item mathematics self-efficacy instrument (D3) entitled "Appraisal Inventory". This instrument consisted of mathematical problems that were very similar to those found on the math test. The instructions ask for participants to "Rate your degree of confidence by recording a number from 0 to 100 in each blank using the scale given below."; with 0 corresponding to "cannot do at all" and 100 corresponding to "certain can do". This methodology follows the *Guide for Constructing Self-Efficacy Scales* (Bandura, 1995).

Demographic Survey. A demographic survey (D6) was used to determine the participant's mathematical background, sex, race, age, high school attended, major, and in which section of the course they were enrolled (for the assignment of credit in class).

Standardized Math Scores. Scores on the math portion of the ACT or SAT were obtained. Participants gave permission for this on the informed consent form.

This study was experimental/causal comparative in nature since assignment to stereotype condition was random, but sex and identification with math domain are not assignable.

Protocol

Part 1) During a regularly scheduled class time prospective participants were briefed on the study by the investigator and had the opportunity to read the Informed Consent Form ICF (D1) for the study. They were told that participation would earn them extra credit in the class but that credit could only be earned if both portions of the study were completed. Those who choose to participate completed the gender and mathematical domain identification survey (titled "About Me") at that time.

Potential participants who were absent the day their class was visited but wanted to participate met with the investigator individually. These individuals received the same Informed Consent form and survey as those who participated in their regular class at that time.

In all cases the participants responded to the identification survey before they were randomly assigned to either the reduced or enhanced threat groups. All participants knew that the study was concerned with potential gender differences in standardized math tests since this information was included in the informed consent material.

Part 2) The second (and longer) portion of data collection occurred individually in a general use computer lab in the mathematics department. Participants who had completed part 1 (above) presented themselves at this site to complete the study. No appointment was necessary as the site was staffed by the investigator for many hours over the course of at least a week.

Participants were given a WebCT ID and password, and were assigned a computer station at which to work. The assignment of the ID - password combination randomly enrolled the participant in one of two "courses" in WebCT. Each "course" was in fact a different threat state for the study. The course ID number was used to track participants throughout the study and subsequent analysis. Course ID's are independent of any identifying participant information (e.g. name, address, university id number, social security number, etc.)

Participants logged into the appropriate site using the ID-password combination given.

During the initial instruction phase of this part of the study (after logging in to WebCT) the participants in each group were informed of the reputed gender bias or lack thereof (this was the only difference between the groups) of the upcoming math test. Participants completed the math self-efficacy survey (4 items) and then the achievement goal survey (18 items) at this time. The self-efficacy instrument exposed

the participants to the type of problems they would see; the level of difficulty of the items; and the time limit of one minute per problem as they would encounter on the math test. The goal survey immediately followed this exposure and again reminded the participants in each group of the purported gender bias of the math test ("women and men doing equally well" or "men perform better than women").

After completion of the goal survey participants initiated a test with 20 mathematics questions to complete in 20 minutes. Explanations/Directions appeared first, along with time limits. Information was once again provided to the participant telling them about the purported gender bias previously found in the test. The math test used has in fact been shown to have no gender bias.

. If the math test was finished in the allotted time the participants submitted the material. If timed out, automatic submission occurred. The math test was scored automatically by the software.

The manipulation of the two states took the following form. For the enhanced threat group the instructions for the test included *"In the past, males have been shown to perform better on this test than females."*. For those in the reduced threat group this read "*In the past males and females have been shown to perform equally well on this test*.". The rest of the instructions (and the test themselves) were identical outside of this difference (See Appendix D5).

All participants concluded the study by initiating and filling out the demographic questionnaire through WebCT.

The total time for participation was around 55 minutes (15 minutes in class and 40 minutes in the computer lab).

ACT or SAT math scores for participants were requested and obtained from the University Records Department.

CHAPTER 3

RESEARCH FINDINGS

Data Preparation

The data for this study were collected in sections of a pre-calculus class that is designed to prepare students for the calculus sequences taken by mathematics, physical science, and engineering students at a major mid-western university. There were 222 participants who completed all portions of the study. After eliminating African American males (due to the confounding influence of racial stereotyping) and people who did not follow directions (particularly on the efficacy instrument) or who clearly did not take the study seriously (those who took less than five minutes to complete the mathematics test) there were 206 participants left in the study. 110 were male and 96 were female. Of those 57 males and 48 females had been randomly assigned to the reduced stereotype threat state while 53 males and 48 females were in the enhanced threat state.

The data set used had a value of 24 inserted for any participant who was missing ACT and SAT scores (this was the average for both males and females). There were 25 of these. 10 were female and 15 were male. These participants were kept in the study to add power to the analysis beyond the "stereotype threat effect" test. A concordance table (Florida Department of Education, 2007) was used to convert SAT scores to ACT scores for those who only had SAT scores. In cases where both ACT and SAT were reported ACT scores were used.

The data were examined for outliers (-3>Z>3) for math score; ACT score; and scores on any of the various motivation, domain identification and efficacy scales. Only two participants fell into this category for any of the measured quantities. In both cases they had very high ACT math scores. They were kept in the analysis since they were outliers in only one of ten possible categories and represented less than one percent of the total sample. Additionally it should be noted that math test scores were used primarily to validate the presence of the stereotype threat effect in this study and exclusion of the two outliers did not materially change the conclusion reached about the presence of the effect.

See Appendix D for instrument content.

Factor Analysis and Scale Reliabilities

Gender and Domain Identification. Gender and mathematical domain identification items were presented to participants in a single instrument. Although all items had been used in previous studies a factor analysis was performed to determine if the items loaded onto individual subscales as expected. The factor analysis used alpha extraction and varimax rotation. Alpah extraction's factor model was designed to be used in scale construction and testing (Kaiser & Coffrey, 1965). Varimax rotation was used because it is known to produce an interpretable set of orthogonal dimensions. Results were as expected. Two very strong factors were present. Factor 1 had an eigenvalue of 3.6 and consisted of all the items (six) expected on the math identity scale. All items had factor loading at least .36. Factor 2 had an eigenvalue of

3.1 and consisted of all the items (four) expected on the gender scale. All items had a factor loading of at least .80 (See Table A1).

Both gender and mathematics identity subscales were analyzed for reliability for this sample. Both were statistically reliable (α = .92 and .83 respectively).

Efficacy. A reliability analysis was performed on the four-item math self-efficacy instrument (See Table D3, for instrument). Cronbach's Alpha was .74 for this sample. Alpha for "any item deleted" never exceeded that for the entire scale. This indicates that the scale is reliable.

Achievement Goal Motivation Subscales. There were four theoretical subscales for motivation included in the 18 item motivation measure. These were Mastery (M), Performance Avoidance (PAV), Performance Approach with a positive valence (PAP+), and Performance Avoidance with a negative valence (PAP-) (see Appendix D4).

Since, as far as I know, no one has ever tried to measure the theorized positive and negative valences for performance approach motivation scales a factor analysis of all the motivation questions was needed to see if, indeed, these questions acted like subscales (see appendix D4 for motivation items and theoretical subscale attributions) that were separable from each other and from PAV.

Factor analysis was performed for the entire sample, including both threat states and both sexes. Since the expectation was that the motivation instrument would produce several factors the Alpha Method of Factor Analysis (Kaiser & Coffrey,

1965) was used. This method's factor model was designed to be used in scale construction and testing. Varimax rotation was used because it is known to produce an interpretable set of orthogonal dimensions.

Three strong factors and one slightly weaker factor (with eigenvalue of slightly less than 1 (.969)) were produced by the factor analysis. The three factors were a bit muddled theoretically.

Factor 1 had 5 PAP+ items and 2 PAV items.

Factor 2 had all the M items but 2 PAP- items as well.

Factor 3 had 2 PAV and 2 PAP- items.

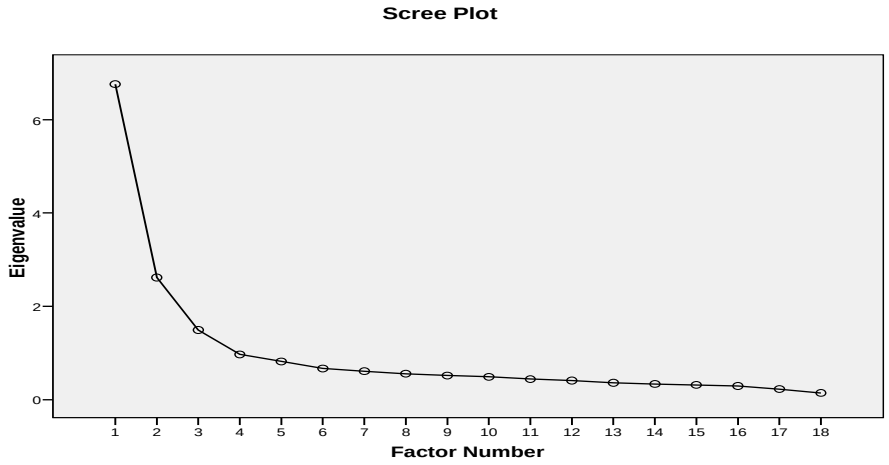


Figure 4. Scree Plot for Achievement Motivation Items

Given the strength of the fourth factor (eigenvalue of .969) and my interpretation of the break-point on the scree plot (Figure 4 above) the analysis was re-run with a forced fourth factor. The results were much more satisfactory from a theoretical perspective.

Factor 1 had all 5 M items, and no other types **(**See Table A2 for factor loadings). Item G7 made essentially no difference in the reliability of the scale (α was .85 with or without G7) so in the interest of parsimony, and to match other motivation scales in number of items it was not included.

Factor 2 had 4 PAP+ items (and no other types); the only PAP+ item that was not included in this factor was G5.

Factor 3 had all 4 PAV items (and no other types). G5 (alleged PAP+) also loaded in this factor, but its loading is relatively weak and it did not really belong in this factor theoretically so it was excluded.

Factor 4 had all 4 PAP- items (and no others).

Reliability analyses where then performed on the motivation scales as defined above.

Table 2.

Reliability and Scale Statistics for Motivation Factors/Sub Scales

Factor Name	Cronbach's Alpha	# of Items	Mean	Variance	Std. Deviation
Mastery	.85	4	14.03	15.190	3.897
PAP+	.85	4	13.30	14.469	3.804
PAV	.78	4	12.09	15.733	3.966
PAP-	.80	4	13.64	12.867	3.587

Further statistics (means, variances etc) for specific items are given in Table A3.

Overall G5 was the only item that did not fall in an anticipated (or predicted) category. It reads, *One of my goals is to perform like this test is easy for me*. None of the other purported PAP+ items have an implied level of difficulty (see item D4), rather they concern themselves with comparisons with how others will do on the items.

There was one item whose deletion from its subscale would have resulted in a slightly larger reliability for the scale. This was true if G1 were deleted from the PAP+ scale (α goes from .85 to .88 if G1 is deleted). However, the subscale was still quite reliable even if G1 was not excluded (see Table A3).

Items G10 and G16 (both purported to be PAV) load onto the PAV scale but have noticeable cross-loading on the PAP+ scale. This is not that unusual, and since both the PAP+ and PAV scales have good reliability as constructed these items were left in the scale they loaded onto most strongly (PAV).

In summary then the subscales used in subsequent analyses were:

Factor 1 - MASTERY (G4,G11,G13, G17)- *G7 omitted*

Factor 2 - PAP+ (G1,G9,G12,G15)

Factor 3 - *PAV* (G3,G6,G10,G16) – *G5 omitted*

Factor 4 - PAP- (G2,G8,G14,G18)

Descriptive Statistics

Results of descriptive statistical analysis for scores on various instruments (Math Self –Efficacy, Math Identification, Gender Identification, Performance Approach (total, positive valence and negative valence) Motivation, Performance Avoidance

Motivation, Mastery Motivation, ACT Math Test Score, and Math Test Score (for the instrument used in the experiment)) were performed. All instruments had acceptable levels of skew (between 1 and -1) and kurtosis (-1>k>1).

Table 3

Normality Statistics for Measured Quantities

	Math ID	Gender ID	Math SE	PAP+	PAP-	PAV	M	Math Test	Math ACT
N	206	206	206	206	206	206	206	206	206
Mean	30.17	19.17	68.04	13.23	13.64	12.09	14.04	4.45	24.02
Std. Error of Mean	.496	.444	1.29	.269	.250	.276	.271	.156	.207
Median	32.00	20.00	68.13	13.00	14.00	12.00	14.00	4.00	24.00
Std. Dev	7.11	6.34	18.50	3.85	3.59	3.40	3.89	2.25	2.97
Variance	50.63	40.66	342.3	14.85	12.87	15.73	15.12	5.04	8.82
Skewness	-.74	-.59	-.34	-.12	-.13	-.08	-.34	.75	-.39
Kurtosis	-.10	-.24	-.44	-.73	-.55	-.65	-.70	.39	.83
Range	33	24	93	15	16	16	16	11	19

Preliminary Statistical Tests

Gender, Threat State and Math Self – Efficacy. An ANOVA was run to explore differences in mathematical self-efficacy scores resulting from differences in gender, threat state or the interaction between gender and threat state. Self-efficacy

did not vary significantly based on gender, threat state or the interaction of gender and threat state.

Math Test Score and Math Self –Efficacy Correlations. Math self-efficacy was correlated with scores on the math test (Pearson Correlation of .274, $p < 0.000$), but was not correlated with gender or threat state.

Ultimately it was the female population that was of interest in this study so analysis (ANOVA) was performed on the female group. For the female group there was no significant difference in efficacy based on group membership (see Table 4 below).

Table 4.

Math Self Efficacy Scores for Females

THREATSTATE	Mean	Std. Deviation	N
Reduced (1)	69.07	18.950	48
Enhanced (2)	68.10	18.885	48

Female Only Threat State Analyses

Scores from the Mathematics and Gender Identification scales were placed in "quartiles". Quartile 1 scores were those in the bottom quarter of the scale; quartile two scores were the group between 25% and 50% of the maximum score; quartile 3

was between 50% and 75% of the maximum and quartile 4 scores were greater than 75% of the maximum score. This allowed for treatment of these scales as categorical variables for use in Analysis of Variance (see foot- note on p 64, and Appendix C for further discussion of this choice of methodology).

Math Test Score Differences by Gender Identification. An ANCOVA was performed where *math score* was the DV, threat state was the IV, and gender identification quartile score, and gender were fixed factors in order to test for interaction effects between identification scores and other independent variables. ACT math scores were co-varied to account for differences in mathematical ability.

Math score was not significantly different for gender identification quartile, or any of the main effects or interactions.

Math Test Score Differences by Math Identification. An ANCOVA similar to the one above was performed with mathematical identification quartile scores in place of gender quartile scores. There were no significant differences in math score for the main effects or interaction terms.

Testing for Gender Based Stereotype Threat Phenomena

Before directly addressing the research questions regarding achievement goal mediation of the threat state – math performance relationship it was necessary to show that the stereotype threat effect had been induced. The effect was indicated if females in the higher threat condition scored significantly lower than males while those in the

lower threat condition showed no significant difference when compared to males. ANCOVAs were performed with threat state and gender as independent variables (IV's), ACT score as the covariate, and math score as the dependent variable (DV).

For clarity in the following discussion the enhanced threat (for females) groups will be referred to as "gender different" and the reduced threat (for females) groups will be referred to as "gender neutral".

Initially the analysis was performed using only the participants who reported ACT Math Scores. The analysis showed that overall group membership approached significance, $F(3,175) = 2.418$ with $p = .068$. When comparing groups the gender different females were significantly lower scoring than the gender neutral males ($p < .014$) and almost so for the gender neutral females ($p < .081$). Note that the gender neutral females did not score significantly lower than the males in either threat state. In the presence of the threat females underperformed, but when the threat was minimized male and female performance was not significantly different. *That is the stereotype threat effect*. See Appendix B for more detailed information on this analysis.

Given the trends in the results above the same analysis was repeated with the value of 24 included for cases where no ACT or SAT score was reported. Twenty – four was the average for both males and females in the overall sample who did report scores. The results showed group membership as significant, $F(3, 206) = 2.646$ with $p=.05$, and significant or near significant differences between specific group membership (see pairwise comparisons (Table 6, page 63 below)). Once again the gender neutral males scored significantly higher than the gender different females ($p <$

0.013, mean difference of .954). The gender neutral males strongly trended toward being significantly higher than the gender different males (p <0.052, mean difference of .722). The gender neutral females' scores were not significantly different than the male scores but were nearly significantly higher than the gender different females (p < 0.071, mean difference of .724). See tables 5 and 6 on the following pages for statistical details and pairwise comparisons.

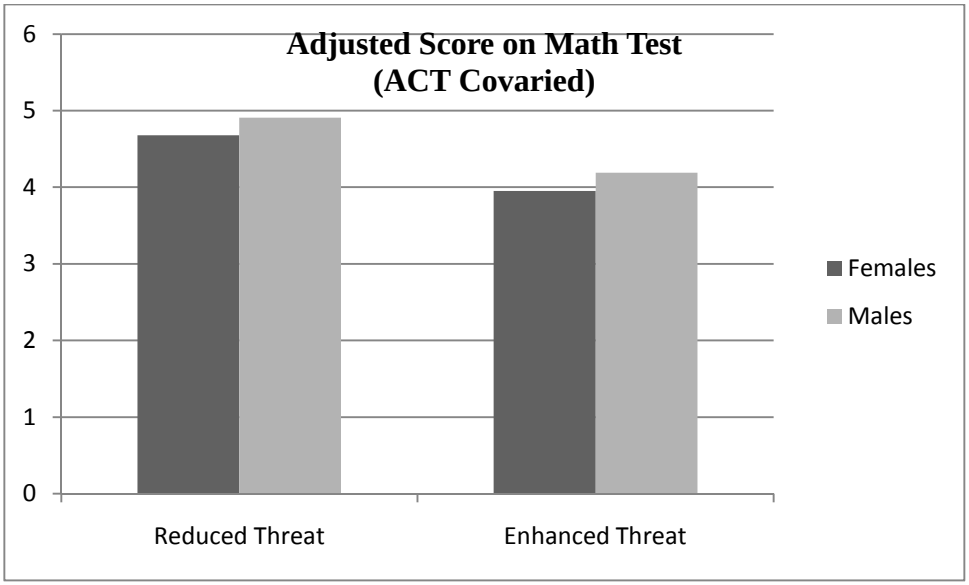


Figure 5. Adjusted Scores for Males and Females with Threat States (ACT Math Scores Covaried)

A third ANCOVA was performed for the same data set where the covariate was score on the mathematical domain identification instrument since the magnitude of the under-performance response to high threat situations is believed to be a function of domain identification. Results showed that group membership was significant F (3, 206) = 3.239 with p=.023.

As in the other analyses the gender neutral males scored significantly higher than the gender different females ($p < 0.08$, mean difference of 1.168). The gender neutral females' scores were not significantly different than the male scores but were significantly higher than the gender different females ($p < 0.012$, mean difference of 1.145). See tables 5 and 6 for statistical details and pairwise comparisons.

Overall then there was strong evidence for the presence of the stereotype threat effect as females in the gender different state underperformed compared to males in the gender neutral state (and to females in that state when math identification is covaried); while the females in the gender neutral state were not significantly outperformed by any group. This indicated the presence of the stereotype threat in mathematics for females and the previously seen (Inzlicht & Ben Zeev, 2003; Spencer et al., 1999; Brown & Josephs, 1999) amelioration of the threat effect when the performance was framed in a manner that debunked the default assumption of the stereotype's presence.

Table 5.

Descriptive Statistics for Math Test Analyses (24 used for missing ACT scores)

Group	Raw Mean	Raw Std. Deviation	Adj. Means (**ACT cov.)**	Std. Error (**ACT cov**.)	Adj. Means (**Math ID cov**.)	Std Error **(Math ID cov.)**	N
Females (GN)	4.88	2.582	4.68	.281	4.87	.320	48
Males (GN)	4.89	2.335	4.91	.257	4.90	.293	57
Females (GD)	3.73	1.888	3.95	.281	3.73	.320	48
Males (GD)	4.23	1.958	4.19	.266	4.23	.304	53
Total	4.45	2.245	4.43	.217	4.43	.309	206

Math ACT evaluated at 24.02. Gender Identity Subscale evaluated at 30.17

Table 6.

Pairwise Comparisons for Math Test Scores Covariate Analyses (24 used for missing ACT)

Dependent Variable: Math Test Score

(I)	(J)	Mean Diff. (I-J) (**ACT cov.)**	Std. Error (**ACT cov.)**	Sig. (**ACT cov.)**	Mean Diff. (I-J) (**Math ID cov**.)	Std. Error (**Math ID cov.**)	Sig. (**Math ID cov.**)
Females (GN)	Males (GN)	-.229	.380	.547	-.023	.434	.957
	Females (GD)	.724	.399	.071	1.145*	.452	.012*
	Males (GD)	.493	.386	.203	.648	.441	.143
Males (GN)	Females (GN)	.229	.380	.547	.023	.434	.957
	Females (GD)	.954(*)	.380	.013*	1.168*	.434	.008*
	Males (GD)	.722	.370	.052	.671	.423	.114

*significant at $p < .05$

Tests of Research Questions

Having established that the stereotype threat effect was in operation, and that the motivation scales were reliable measures for M, PAV, PAP- and PAP+, analyses that directly address research questions 1 and 3 could be done meaningfully. At this point interest in males, who were necessary in the preliminary analysis to establish the presence of stereotype threat effects ebbed considerably. It was the differences in female behavior/performance based on varying threat condition that was of

fundamental interest in the remaining analyses. Unless noted otherwise the analyses below were performed on the female cohort only. There were 48 females in each threat condition.

A MANOVA was performed with Threat State as an IV and Math Domain Identification (quartiles) as a fixed factor (IV) (for females only). The DV's were scores on the Mastery (M), Performance Avoidance (PAV), Performance Approach Positive (PAP+) and Performance Approach Negative scales (PAP-). [1]

In the analysis the math domain identification scores (which were actually continuous) were re-coded into quartile scores (1, 2, 3 & 4), as described previously, so that this variable became categorical and could be used in a MANOVA as an IV.

---

[1] A multiple regression analysis was also attempted in lieu of the MANOVAs discussed above (see Appendix C for discussion of the regression analysis). The results were essentially the same as those discussed below with the exception that the multiple regression technique did not indicate that there was the possibility of an interaction between math identification and threat state (which the MANOVA shows for PAP+). This, along with greater error associated with the multiple analyses run in the regression technique lead me to use the MANOVA approach even though the regression analysis preserved more of the variability of the math identification scale (and hence would have more power). Additionally the quartile approach follows Aronson et. al. (1999) who similarly divided groups of participants into more broadly defined aggregates based on their scores on a domain identification instrument (although they used thirds).

For the analysis, with all four motivation scales, threat state was significant (or very nearly so) (Wilks' Lamda = .055, F = 2.4); as was identification with the domain of mathematics (Wilks' Lamda = .000, F= 3.5) and the interaction term between domain identification and threat state (Wilks' Lamda = .034, F= 1.9).

Question 1: In a situation where performance on a mathematical test is linked to stereotype threat does threat condition significantly affect the performance avoidance goal endorsement of female participants?

The mean for the enhanced threat group was 11.12 while that of the reduced threat group was 11.92. This difference in PAV endorsement between females in the different threat states was not significant.

Question 2: Is performance avoidance endorsement a mediator in the stereotype threat effect on performance?

Since there was no statistically significant link between threat state and endorsement of PAV goals, performance avoidance could not have been a mediator of the threat state – math performance relationship.

Question 3: Does stereotype threat have an influence on endorsement of other goal types (M, PAP, PAP+, and PAP-)? If so is there a mediational effect between threat level and performance for that goal type?

Examining specific scales, the largest effect (from threat state) was seen for the mastery scale. The effect was significant (F (1,96) = 6.584 with p = .012) with the

enhanced threat group reporting higher levels of mastery achievement motivation. The mean for the reduced threat group was 12.81 while the mean for the enhanced threat group was 14.70. This was the only significant result for motivation scales for different substates. It is of interest to note that the enhanced threat group reported a higher level of mastery endorsement while scoring significantly lower on the math test that the low threat group.

Although not specifically mentioned in research question 3, differences in scores for various motivational states across math identification levels were included in the MANOVA described on page 64. Results of that portion of the analysis are discussed below; in part because of a clear interaction effect between mathematical identification level and threat state with respect to PAP+ endorsement (see below).

Both the Mastery and PAP+ terms were significantly affected by math identification, with higher mathematical domain identification scores resulting in higher reported levels of M and PAP+ motivation.

Table 7.

Tests of Between-Subjects Effects for Threat State and Mathematical Domain Identification on Motivational Scales (Significant Results)

Source	Dependent Variable	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[b]
Threat State	Mastery	1	81.2	6.58	.012*	.070	.718
Math ID quartiles	Mastery	3	124.8	10.12	.000*	.256	.998
	PAP+	3	39.7	3.29	.024*	.101	.734

*significant at $p < .05$

For the mastery scale $F(3,96) = 10.119$ with $p = 0.000$ for all female participants. In fact, endorsement of mastery goals increased significantly as math identification quartile increased for *both* the reduced and enhanced threat groups (see Figure 9, page 71). For the reduced threat state $F(3,48) = 6.227$ with $p = .001$. For the enhanced threat state $F(3,48) = 3.965$ with $p = .014$.

For the PAP+ scale $F(3,96) = 3.288$ with $p = .024$. This significant result was due almost entirely to changes in the PAP+ score in the enhanced threat state (see Figure 6 below).

Oddly, even though during the MANOVA analysis the interaction term between Math Identification and Threat State was significant there were no scales that were specifically significant when examined from a univariate perspective. However, looking at the motivation means for the cross term possibilities it was found that in every case, by the fourth quartile of identification the motivation score was *higher* in the high threat state than the reduced threat state**.** Additionally it wass clear from Figure 6 that for PAP+ there was an interaction (mentioned previously) where the enhanced threat group reacted much differently that the reduced threat group as math identification increased.

The following graphs show scores for the various motivation scales for both threat levels, by math identification quartile. The x-axis is the identification quartile, and the y-axis is the score on the motivation scale.

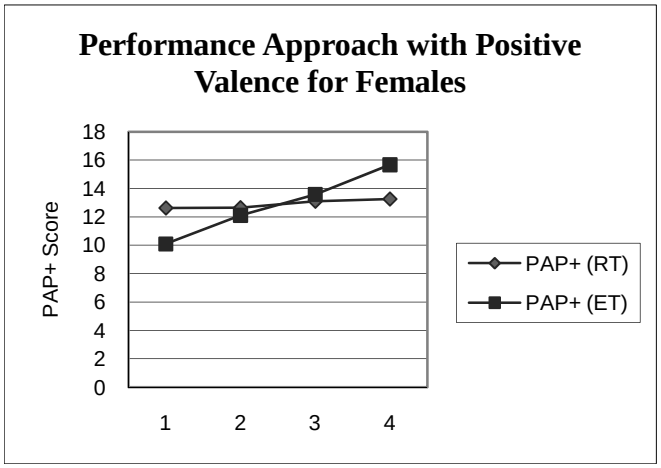


Figure 6. PAP+ Scores by Math ID Quartile and Threat State

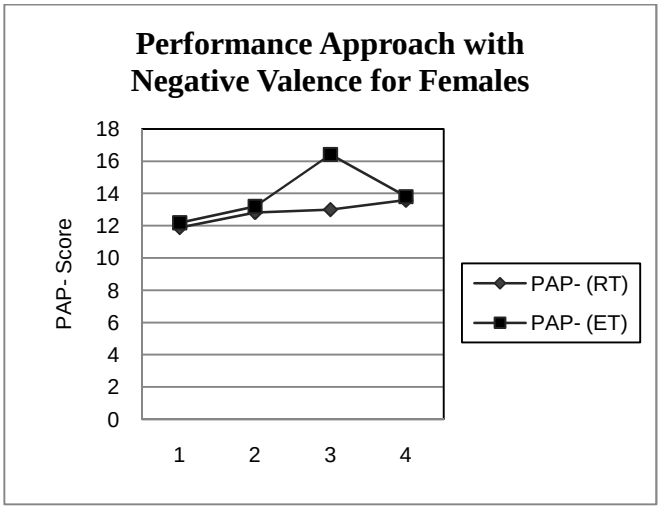


Figure 7. PAP- Scores by Math ID Quartile and Threat State

From Figure 6 it appears that there was an interaction effect between threat level and identification with the domain of mathematics for PAP+. Figure 6 also shows that the significant result for PAP+ with mathematics identification was due almost entirely to the enhanced threat group; even though the mean scores on PAP+ for the

two different threat groups were virtually the same. It was the interaction of threat state (enhanced) and the level of math identification that led to differences in the performance approach (+) scores.

Specifically then, the low threat group did not show significant differences in PAP+ score as math identification varied; but the high threat group did. $F(3,48) = 5.693$ with $p = .002$. This was in keeping with stereotype threat literature; which reports that in the high threat situation people with lower levels of domain identification are not particularly affected but the more highly identified participants are (Aronson et al., 1999). In the reduced threat state there was little variability in PAP+ motivation no matter how identified the person was with the domain.

Scores on PAP+ for first quartile reduced threat vs. first quartile enhanced threat were not significantly different ($p<.166$). The same held for the fourth quartile ($p<.097$). With higher numbers of females (more analytical power) it is possible that these differences would have reached statistical significance.

Interestingly, although in the high threat state increasing identification was associated with higher PAP+ score, the same was not true for the score on the math test (see Figure 8 below). In general the math test scores showed no significant difference over identification quartile (although there was a significant difference due to threat state). The patterns of response were intriguing however since they were not simple (e.g. increasing score with increasing identification) with the highest scores occurring in the third quartile but they were the same for both threat states. This pattern may be an indication of an effect similar to that found by Aronson et al. (1999) where moderately identified female participants (second out of three thirds) scored

significantly higher on the math test in the higher threat state. They took this as an indication of the differential effect of stereotype which causes the most identified to feel most threatened in the STT situations. It may be that the overtly stated nature of the experiment (to see why some tests favor males) caused the highest scoring females in either state (in the present study) to suffer some performance deficits.

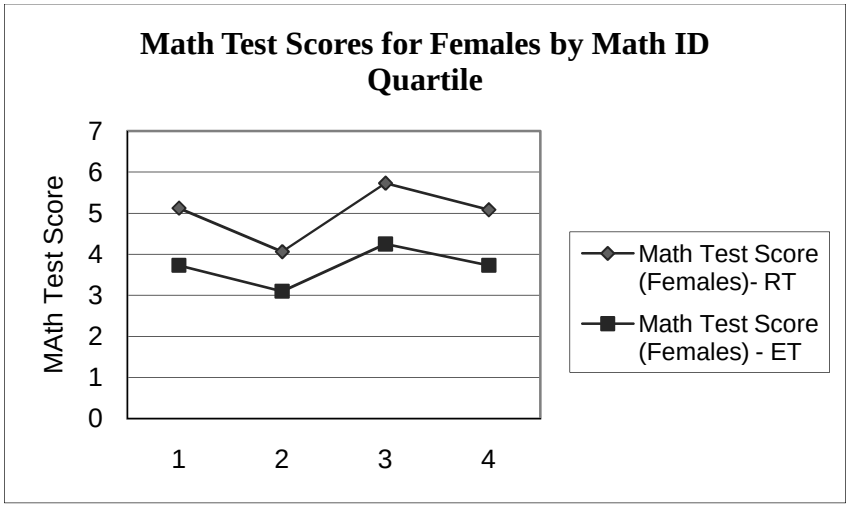


Figure 8. Math Test Scores by Math ID Quartile and Threat State

Analyses were run to determine if there were any correlations between math identification and ACT score, Math Test Score, or Math SE. None of the correlations were significant, although math identification with Math SE came close with $p < .052$.

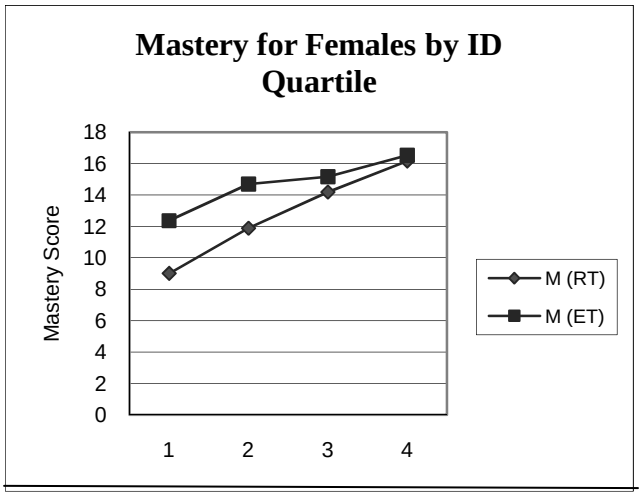


Figure 9. Mastery Scores by Math ID Quartile and Threat State

For mastery scale scores the plot (Figure 9) shows that the scores started off quite differently for the lower quartile participants (with high threat reporting much higher scores on mastery motivation) but the difference became smaller and smaller for the more highly identified subjects. It seems reasonable to postulate that the most highly identified would have reported high levels of mastery motivation no matter what the threat condition, and since the scale is not open ended (maximum of 20) the scores on the mastery scale should have approached one another for the more highly identified.

Therefore generally speaking being in the enhanced threat state induced females to report higher levels of motivation (particularly mastery in the case of threat state alone, but also in PAP+ when math identification and threat state are considered) than those in the reduced threat state. These higher levels of motivation were in the state that produced lower math scores.

Research question 3 was also concerned with the possible mediational effects of threat state manipulation on other dimensions of performance motivation (besides PAV). Mastery goals and performance approach (positive valence) goals were both found to vary significantly with respect to threat state (mastery score) and/or math domain identification (M score and PAP+ score). Mastery and PAP+ were examined for mediation effects (of threat state/ math identification and math identification, respectively).

Following Baron & Kenny (1986) a series of regression analyses were performed to test for mediation by mastery endorsement on the threat state – math performance relationship.

For threat condition and mastery (M) score the following regressions were performed.

1) Math score with threat condition
2) Math score with mastery score
3) Threat condition with mastery score
4) Math score with threat condition and mastery score entered simultaneously

If math score and threat condition; and math score and mastery score; and threat condition and mastery score were all significantly related then the conditions for mediation would have been met. If, when math score, threat condition and mastery score are entered simultaneously in the regression, the relationship between threat state and math score became non-significant and mastery score retained its significant relationship with math score; then the conclusion is that mastery state was a mediator

for stereotype threat and math performance. If the variables were still significantly related, but the effect was significantly less than before the inclusion of mastery score then reported mastery endorsement score would have been a partial mediator of the effects of threat condition.

Since threat state was categorical, effect coding had to be used. This created one new variable (since the categorical variable had only two states).

Table 8.

Dummy Coding of Threat State Variable for Regression Analysis

	EFF 1
Enhanced State	-1
Reduced State	+1

This made the enhanced state the base group. This agreed with previous studies that have shown that when no information was given individuals susceptible to the effect operated as if the threat was present (Steele & Aronson, 1995). Comparisons for significant differences were done with reference to the grand mean. Comparisons are not reported for the base group.

Regression of threat condition and math score showed that $R = .249$ and $p < .015$. Thus the conclusion is that threat condition accounted for a significant amount of the variance in math score (6.2%).

Regression of math test and mastery score showed $R = .286$ with $p < .005$. The conclusion here is that mastery motivation accounted for a significant portion (8.2%) of the variance in math test score.

Finally a regression of mastery score with threat state was performed. This analysis showed R= .230 with $p < .024$. So threat state and mastery score shared a significant amount of variance.

At this point the effects of threat state and mastery score on math score were tested together. A simultaneous regression, where threat state and mastery score were entered in the same step, and math score was the dependent variable was performed. The simultaneous regression showed that β for threat state drops from .248 ($p = .015$) when it (threat state) was considered alone to .193 ($p = .058$) when it was entered along with mastery score ($\beta = .242$, $p = .018$). The reduction in the amount of variance accounted for by threat state and the fact that the variance accounted for by threat state was no longer significant ($p = .058$) when entered with mastery score supports the conclusion that mastery motivation was a mediator of the effects of threat state on math score.

Note that the coding of the effective variable for Threat State made the reduced threat state positive and the enhanced state negative, which effects the sign on β (see Table 9 below) , making it the opposite of what might be intuitive (where enhanced > reduced).

Table 9.

Standardized Regression Coefficients for Mediation Analyses of Mastery Score on Threat State and Math Score

	Standardized Regression Coeff. (β)
Analysis 1 (Math Score with Threat State)	.25 (p = .015)*
Analysis 2 (Math Score with Mastery Score)	-.29 (p = .005)*
Analysis 3 (Threat State with Mastery Score)	-.23 (p = .024)*
Analysis 4 (Math Score with Threat State and Mastery Score)	
Threat State	.19 (p =. 058)
Mastery Score	-.24 (p= .018)*

*significant at $p < .05$

Considering mathematical domain identification as an independent variable (rather than threat state) a similar analysis for mediation of M and PAP+ on the domain identification - math test score relationship was attempted. However, regression analyses showed that math identification did not account for a significant amount of variance in math score (neither did PAP+). This result ruled out mediation by mastery or PAP+ scores between domain identification and math score.

Other Findings

Although performance avoidance (PAV) scores did not show a significant change with threat state it may be informative to look at some of the results for PAV and other motivation subscales in more detail.

Clearly there was no real pattern of response (see Figure 10 below) as far as increasing mathematical domain identification was concerned. Although it did not reach statistical significance (compared to the reduced threat state) it is of interest to note the sudden increase in PAV score between the third and fourth quartiles (of math identification) for the enhanced threat state.

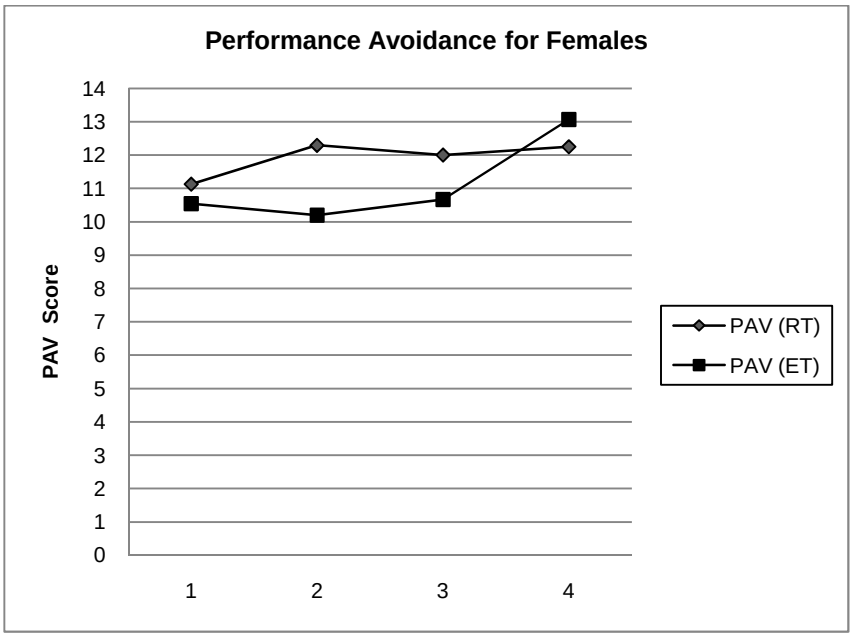


Figure 10. PAV Scores by Math ID Quartile and Threat State\

Since all the motivation subscales had four items and a maximum score of 20 it might be of some value to compare the strength of response for each scale in each threat condition.

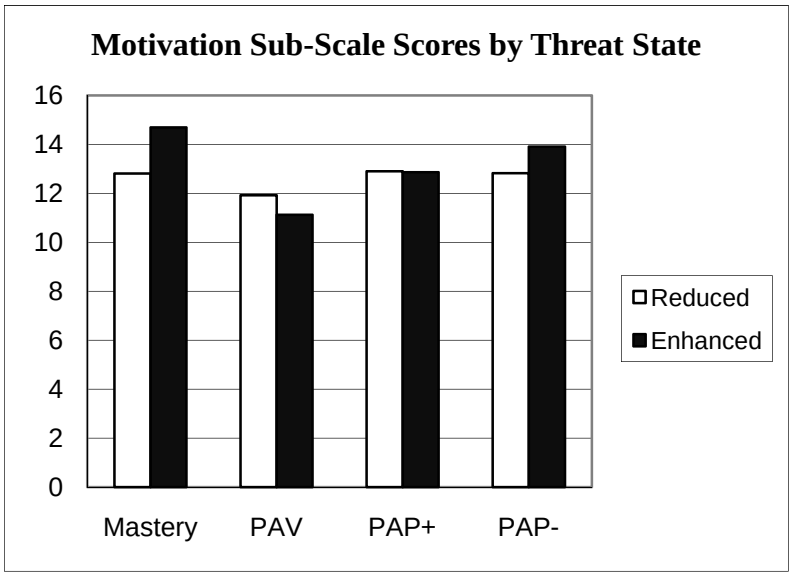


Figure 11. Scores for Motivation Subscales for Both Threat States

There were some trends of note here. First, the scores were very similar (in the 12-13) range for most of the groups. The exceptions were the mastery scale score for the enhanced threat group; the performance approach scale with negative valence for the enhanced threat group; and the performance avoidance scale for the enhanced threat group (really scores for PAV under either condition seemed a little low).

A t-test was performed to determine if any of the motivation scores in either state were significantly different from the average subscale score (13.01). The results are in Table 10.

Table 10.

T-test for Significant Differences in Motivation Sub Scale Means

	t	df	sig (2-tailed)	Mean Difference
PAP+ (Reduced)	-.23	47	.821	-.11
PAP+ (Enhanced)	.21	47	.837	.12
PAP- (Reduced)	-.19	47	.851	-.11
PAP-(Enhanced)	1.8	47	.084	.95
M (Reduced)	-.01	47	.994	-.005
M (Enhanced)	3.83	47	.000*	1.85
PAV (Reduced)	-1.72	47	.092	-.98
PAV (Enhanced)	-2.94	47	.005*	-1.71

*significant at $p < .05$

The mastery score increase in the enhanced threat case was, in fact, one of the statistically significant results of the study and the t-test confirms that it was significantly higher than the average motivation subscale score.

The scores in the enhanced threat group for negatively valenced performance approach were interesting. This was the second highest scoring group, and the t-test indicates that the difference between the mean for that score and the overall average subscale score (.953) might have reached significance with a larger number of female participants.

The performance avoidance group did not behave at all as expected, and not like the other groups. The scores for both threat conditions for this group were lower than the means of the other scales; and this was the only scale in which reported endorsement of motivation went *down* in the enhanced threat group. In fact the t-test shows that this difference was significant for the enhanced threat group.

Overall it seems that the enhanced threat group was highly motivated, and in the many of the "right" ways. They wanted to demonstrate mastery of the material and were eager to try hard on the test (although it was because they were worried about their performance), while being less interested in avoiding looking bad. And yet they scored lower on the math test.

Of course threat condition was only one of the factors that contributed to stereotype threat effects. Math domain identification was also of interest.

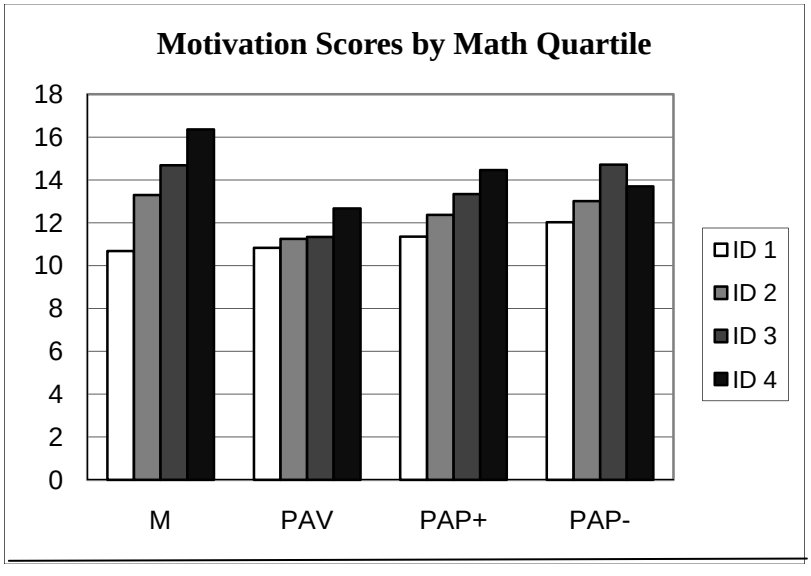


Figure 12. Scores for Motivation Subscales for Math Identification Quartile

The significant differences with math identification are clear for mastery and PAP+ from this figure. Both showed steady increases as math identification increased.

Figures 13 and 14 (below) combine the data represented in Figure 12 and earlier line graphs such as Figure 9 to show how the scores on the motivation scales varied over mathematics identification *and* threat state.

Figures 13 and 14, taken together illustrate the previously noted interaction effect between threat state and mathematics identification on motivation scores (particularly PAP+, which reaches significance). In the reduced threat state there was little variability in motivation scores from scale to scale (except M) or over math identification; while in the enhanced state there was considerable variability in score, particularly over math identification quartile.

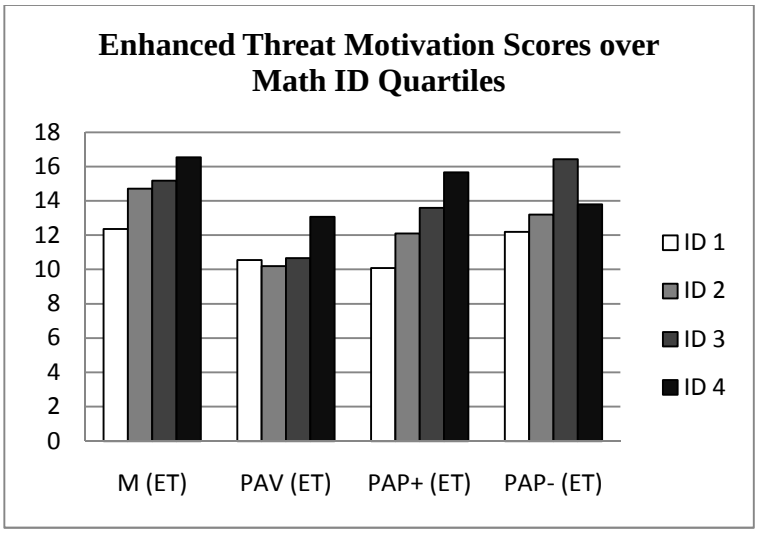


Figure 13. Scores for Motivation Subscales for Math Identification Quartiles for the Enhanced Threat State.

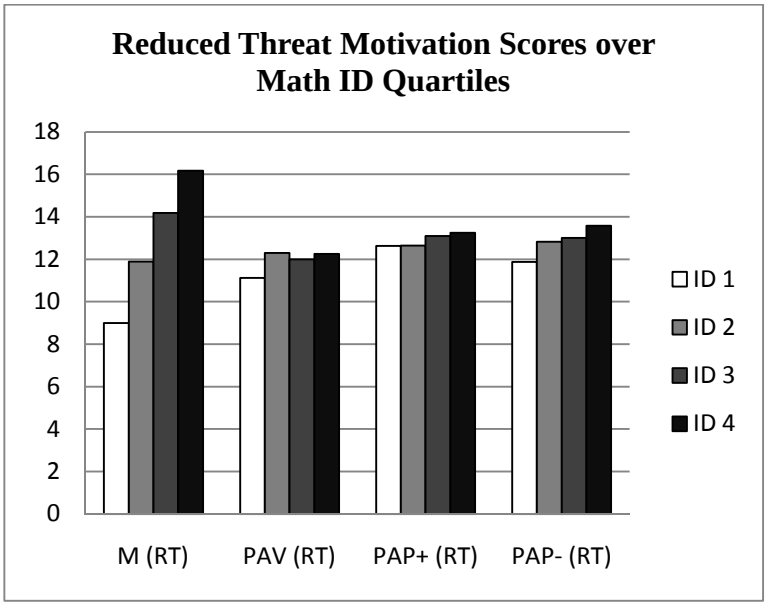


Figure 14. Scores for Motivation Subscales for Math Identification Quartiles for the Reduced Threat State.

Figure 13 is of further interest in that it shows that the threatened females who were less identified reported scores that showed less of a tendency toward avoidance motivation when they knew they would be compared to males.

Males, mentioned parenthetically in question one of this study, have not been addressed thus far in terms of threat state/domain identity/goal choice interactions. Males (particularly since African American males have been excluded) have not been negatively stigmatized with respect to the academic domain of mathematics. There was no real reason to expect that they would react to manipulations of threat state in a way that would make them useful as control group for females. Their self-image regarding mathematics (given that they have not had to labor under a stigma for years)

would be different than that of females. However as a check, analyses parallel to those described above were run for males. As expected the two groups had little in common in how they reacted to the manipulation of threat state. Males had significant results where females did not, and vice-versa. In only one circumstance did both sexes show significant motivational responses for the same threat state or math identification situation. This occurred for PAP+ motivation changes seen over domain identification quartile. As with females, the males in the group told that "females do as well as males" had higher scores on the PAP+ scale and showed a pattern of increasing score as math identification increased. Unlike females however the males who reported higher levels of PAP+ endorsement scored better on the subsequent math test than males in the other threat state. See Appendix E for more detailed results and discussion for males.

CHAPTER 4

DISCUSSION

Stereotype Threat, Domain Identification and Achievement Motivation

Stereotype Threat research has had difficulty explicating mechanisms for the reduction or elimination of underperformance of stereotyped groups after a simple statement that the behavior in question has not been shown to follow the stereotypical results. Susceptibility to the inherent threat associated with performance in a stereotyped domain clearly has three major components. First, the person must belong to the stereotyped group (whether or not they accept to the stereotype) and second they must be identified with the behavior (or intellectual domain) for which the bias is present. A stereotype involving women in mathematics does not concern all women. It does concern women who, to some degree, feel that skill in mathematics is a part of their self-image. The magnitude, or even presence of, the effect of a stereotype should therefore be dependent on the degree of centrality that "math" holds for a person. Thirdly, the actual situation in which they find themselves must be challenging so that they feel the pressure of possible underperformance.

The present study was concerned with the following research questions:

1) In a situation where performance on a mathematical test is linked to stereotype threat does threat condition significantly affect the performance avoidance goal endorsement of female participants (but not male participants)?

2) Is performance avoidance endorsement a mediator in the stereotype threat effect on performance?
3) Does stereotype threat have an influence on endorsement of other goal types (M, PAP, PAP+, and PAP-)? If so is there a mediational effect between threat level and performance for that goal type?

Stereotype Threat Condition and Motivation

It has been hypothesized that the stereotype threat effect (lower performance among threatened populations that is ameliorated by reducing the threat) is at least in part mediated by changes in the achievement goal structure (motivation) in response to the changing perceptions of threat (Smith 2004; Ryan & Ryan 2005). Theory has focused on the maladaptive behaviors associated with performance avoidance goal structure (Elliot & Church, 1999; Smith, 2004; Ryan & Ryan, 2005). Indeed in recent studies, Smith (2006), Smith & White (2007), and Chalabaev, et al. (2008) found that individuals in stereotyped groups who were operating in a threatening situation (in the domains of mathematics and soccer respectively) tended to endorse performance avoidance goals. In the case of the 2006 Smith experiment PAV mediated threat state and math performance for females. Each of these studies however had certain procedural, practical or theoretical problems that made them less than absolutely convincing (see chapter 1). More germane to the present study is the fact that Brodish & Devine, (2009), in an excellent study, found that performance avoidance endorsement (and worry) mediated the effect of threat state on math scores for women. The current study was expected to produce similar results, however no such

mediational relationship was found between threat state, PAV and math test score. Indeed there is no statistically significant connection between threat state and PAV endorsement at all for this group.

The reasoning behind the expected (and in other studies demonstrated) performance avoidance endorsement in a threatening situation is straightforward on the surface. Those in the high threat state should feel, threatened, and it would be reasonable to assume that the desire to avoid failure would increase. An increase in PAV motivation should lead to an increase in the maladaptive behaviors (such as poor self – regulation, lack of effort and increased surface processing) associated with avoidance goals, resulting in decreased performance. The performance deficit was there, but the difference in performance avoidance endorsement between threat groups was not. Why? Performance goals of all types arise from the idea that the results of an activity will be compared to the scores of others. This may not be the case for individuals in a study such as the current one. Participants were told that their identities and scores were confidential. Even the course instructor was only told whether the participant finished the study or not (and participants were aware of this). In this case it might be that there was no real threat of "personal comparison" to others. It is their groups' score that matters and is reported. Avoidance would do little to protect the individual and could do real harm to the group score.

The presence of stereotype threat is clearly situational since individuals are able to key on relatively small differences in environment or instruction that either enhance or reduce the level of threat. In the Brodish & Devine (2009) study (mentioned above and discussed in chapter 1) the threat/no threat conditions varied both in whether

gender differences were mentioned (yes in the case of high threat, no in the case of low threat) and in what the test itself was purported to measure ("quantitative" capacity in the high threat state and "memory" capacity in the low threat state). Could it be that the differences in PAV endorsement reported by Brodish & Devine are due to the reframing of the nature of the test; or as an interaction effect between the reframing of the meaning and the enhanced threat level due to a clear statement of gender significance? In the present study all participants were told that some math tests showed differences in score based on gender while others did not and the purpose of the study was to determine why this might be occurring. The manipulation was to simply tell the participant that the test they were taking had not shown these differences (reduced threat) or that it had (enhanced threat). The two groups therefore had the same preparation and the same "meaning" attached to the test.

In this study there was no "avoiding" membership in a group based on gender or race or some other uncontrolled personal characteristic (like threat state) and the meaning of the entire exercise was fundamentally the same for all participants. Participants knew that their specific scores would not be revealed, and were not even really in question. It was their group's score that was important. "Avoidance" was of no real use to them personally and would have been a detriment to their group. It seems unlikely that participants would actually "think" these things through, but it has been shown that people are very sensitive to the contextual clues in their environment (Elliot, 1999). So it could be that in this particular experiment the avoidance option was minimized by the circumstance of the performance and stated meaning of the study. Indeed, for this study it has been shown that the PAV scores were lower than

the other types of motivation scores (see Figure 10 and Figure 12). For the high threat group that difference was significant ($p = .005$) and for the low threat group the difference was nearly significant ($p = .092$) (see Table 10). These were the only significant (or near significant) results for performance avoidance endorsement. This result may have to do with the antecedents to the performance (as suggested above) or it may simply be that people were less willing to admit PAV goals. Whatever the case may be, it is intriguing that the PAV scores trended downward when the threat was enhanced, which is the opposite of the predicted result.

Since PAV scores were not significantly linked to threat state the second research question (involving PAV's mediational role in linking threat to performance) is moot. In this study performance avoidance goals did not mediate the effect of stereotype threat on math performance for females.

This having been said, the study (which included instruments that measured Mastery (M), Performance Approach with Positive Valence (PAP+) and Performance Approach with Negative Valence (PAP-)) did produce significant/interesting results that linked dimensions of motivation to threat state-math performance and/or identification with the domain of mathematics for females. Additionally, this study successfully showed that PAP-could be measured separately from both PAP+, and PAV. These results will be discussed in the remainder of this chapter beginning with the PAP- result.

Measuring Performance Approach with Negative and Positive Valences

Elliot & McGregor (2001) hypothesized the existence of two forms of performance approach achievement goals. However, no known attempt has been previously made to measure positively valenced and negatively valenced performance approach goal structures separately. A positive valence implies approach to the math items in questions with the intent of showing how good the individual is at the material (in comparison to others, or to show that the material is easy for the individual). This (PAP+) is "performance approach" as generally measured (see Elliot, 1999 for example). A negative valence indicates that the person is concerned that they may not do well on the material but intend to work hard (approach) to overcome that possibility.

Factor analysis in this study showed that the two subtypes of performance approach goal structure are separable from each other and from other goal types (see Table 2). This is an important and thought provoking result in general and for this study in particular in that anxiety and other maladaptive responses generally associated with PAV can reasonably be hypothesized as being present (although to a somewhat lesser extent) in the PAP- goal structure as well. The ability to isolate and differentiate between PAP- and PAV (or PAP- and PAP+) will give researchers more granularity in their measures and greater understanding of the goals adopted by people in performance situations. As discussed in more detail below PAP+, PAP- and PAV were adopted to different degrees and in different patterns in this study.

Threat State, Mathematics Domain Identification and Motivation

Threat State and Motivation. This study did show a main effect for mastery goal endorsement with threat state ($F(1,96) = 6.584$ with $p = .012$); with women in the enhanced threat condition self-reporting significantly more mastery orientation than those in the reduced threat condition. Indeed, it was shown that mastery goal endorsement mediated threat state and lower math performance in this experiment (see pages 70-74). The fact that this higher level of mastery goal endorsement was coupled with lower scores on the math test is somewhat counterintuitive; as one might expect higher levels of mastery goals (generally regarded as beneficial to persistence and effort) would result in higher scores.

How could higher levels of mastery goal endorsement lead to poor performance? If the participant frames learning and performance for "myself and my own edification" other goals such as comparisons to others, memberships in a group, and the "good of the group" are minimized in importance. In essence it appears that in this case mastery goals were operating as selfish goals. This rather maladaptive aspect of mastery orientation has been hinted at in the literature. Senko & Miles (2008) found that students' mastery endorsement was predictive of study habits. Students studied material that was personally interesting; which was not necessarily the same as material that was important for the course. This tendency was predictive of lower grades in class. While the Senko study was focused on study habits over a relatively long period of time, and the current experiment was focused on immediate goal endorsement in a very specific performance situation. The idea that mastery goals can lead to choices that are not helpful in performance situations is interesting. In the case

of the present study where time was a definite constraint (one minute per problem) it is possible that task persistence associated with mastery endorsement might not be such a good thing. In other words spending too much time on one problem before moving on is not a good strategy for this test (which is modeled after similar ACT and GRE tests). High scores on this test are almost certainly due to mathematical ability and good, focused, test taking techniques. That is the case for most mathematics tests however.

Why *should* higher levels of mastery goal endorsement occur for females in the enhanced threat state in this experiment? Perhaps if avoidance (the predicted response) is not a viable option, and yet it is possible that underperformance might happen, the next best thing (from a self-protection point of view) might actually be "mastery endorsement"; thus minimizing the importance of group membership and results.

There is a pre-supposition here that participants in the enhanced threat group are consciously or unconsciously anticipating the need to deal with underperformance of the group they are associated with. It should be recalled that in this experiment the participants were exposed to questions that were very similar in nature to those that they would see on the math test as part of a self-efficacy measure before completing the motivation questionnaire. Could it be that any type of differential goal structure between enhanced and reduced groups would serve the same "protective" role so that in one situation there might be a significant increase in performance avoidance endorsement while in another case some other dimension of motivation might be triggered? If this is the case the mediating role of some dimension of motivation (PAV, M or any other) is a reflection of the deeper, self-protective need. When the

participant discovers the challenging nature of the material is it possible that any enhanced motivation becomes a distraction? This would support the findings of Schmader & Johns, (2003) that associated enhanced threat conditions with high cognitive load.

While there was no significant main effect for performance approach with a negative valence (PAP-) from the MANCOVA calculations there was some indication that PAP- is important in the enhanced threat group. As reported in Table 7, there was a tendency toward higher PAP- scores for the enhanced threat state ($F(1,96) = 1.823$ with $p = .180$). While certainly not significant this might suggest a trend. This trend would be supported by the fact that in a subsequent t-test (see Table 10) the PAP- score for the enhanced threat state approached a significant difference (higher than the average) ($p = .084$) compared to the other motivation subscale scores. Neither of these results is necessarily scintillating on its own, but taken together they do indicate that participants in the enhanced threat group may be more likely to endorse motivation goals that stress approach and effort but did so out of a participant's concern that she might not do as well as others. This would seem to be a reasonable response to being informed that your group (females) has been shown to be outperformed on the math test you are about to take by another group (males). In fact this might be a more natural response than PAV given the situation. This relationship warrants further study.

In past studies where "performance approach" has been reported the items on the instruments used were phrased positively. In essence the PAP+ scale in this study is the PAP scale of other studies. Had the PAP- scales not been used in this study there

would have been no indication that there was any performance goal change associated with a negative valence with differing threat condition. This seems to indicate that the PAP- and PAP+ subscales may provide insight into situations that are not available with the generic PAP scale.

Domain Identification and Motivation. There is no uniform agreement regarding the measurement of domain identification for women in mathematics. In some studies the fact that a person was enrolled in a particular class was used as an indication of identification (c.f. Spencer et al., 1999). In at least one other study scores on standardized college entrance math tests was used (Brodish & Devine, 2009). Others have used specifically developed survey type instrumentation (Smith, 2006; Aronson et. al., 1999, Brown & Josephs, 2000, Schmader, 2002; the present study) In at least one case, (Smith et al., 2007) participants were chosen because they were early in their college experience and had not yet committed to pursuing a degree in math or a related field.

Initially domain identification for this study was intended to be linked to class enrollment. Women enrolled in a pre-calculus course were assumed to be at least partially identified with mathematics simply because there were other, less rigorous, mathematics courses that could be used to fulfill the requirements for graduation. However, in order to more specifically gauge identification with gender and mathematics, participants also responded to questionnaires regarding their identification with both their gender and the domain of mathematics. There are no significant results associated with the gender identification measure. There are

however significant results with respect to scores on the domain identification instrument. Participants were grouped into quartiles according to their score on the domain identification instrument. Those in the first quartile could be labeled as somewhat identified with mathematics while those in the fourth quartile are strongly identified. This basic methodology is also seen in Aronson et al. (1999) where the group was divided into thirds.

Endorsement of mastery goals increased significantly as math identification quartile increased for both the reduced and enhanced threat groups (see Figure 9, page 71). As discussed above there was also a significant difference in the means of the two threat conditions for mastery score.

Scores on the mastery subscale increased over math identification quartiles for both threat groups, although the scores converged at a little over 16 on a 20 point scale (see Figure 9). The big differences (that results in a significant finding) between the two threat states occur for the lower two quartiles. The convergence of scores for the upper two quartiles is most likely a ceiling effect where the most identified in either group were reporting strong mastery endorsement on a limited scale. The major difference between how the different levels of identification report their mastery score for the different threat states is mainly a matter of magnitude, with the reduced threat group reporting lower scores for each quartile.

Endorsement of performance approach with positive valence (PAP+) was significantly higher for the enhanced threat group than for the reduced threat group (see Figure 6, page 68). In fact the increase in score on the PAP+ over math

identification quartile was so large for the enhanced threat group that it alone leads to a significant result for the entire sample of females for math identification and PAP+. The mean scores for PAP+ for the two threat conditions are not significantly different (both are 12.9) so analysis by threat state does not reveal any significant difference in the PAP+ score due to threat condition. However it is clear that the significant result for PAP+ score over math identification grouping is a result of the different ways in which the threat groups responded as math identification increased. See figures 6 and 7, or 13 and 14.

This subscale measures the strength of response to the basic proposition that the goal of performance on this math test is to show that the participant wants to do better than others. The endorsement of this goal escalates for the enhanced threat group, as it did for both groups for the mastery subscale. However, the reduced threat group shows essentially no change for PAP+ by quartile. Being informed that the test had not showed a gender bias appears to have significantly reduced the competitive motivation to outperform others in this group's motivation profile. Significantly higher scores on the math test resulted. Is the reduction of threat causing participants to be less sensitive (or care less about) the type and difficulty of the work as it relates to their own self-identification with the domain? Their identification with the domain is just less important? The attenuation of the salience of this bit of self-identity then leads to better performance? So it seems.

For the negatively valenced PAP there is no such simple pattern, or statistically significant result over math identification, although the sudden increase for those at

the third quartile of identification who are in the enhanced threat state is noteworthy (see figures 6 and 7 or 13 and 14) as it indicates that this group may be more concerned with working hard to overcome possible difficulties than the other groups. It is possible that the more identified might have recognized how hard the material was while the less identified did not. The third quartile participants might be more concerned with this discovery than the most highly identified people in the fourth quartile of the enhanced threat group. However if this were true the third or fourth quartile folks in the reduced threat group might be expected to show a similar increase (they did not); unless the reduction in threat ameliorated this concern.

Is the third quartile group in the enhanced threat state a special group? The items in the PAP- scale are things like:

> I will try hard on this math test because I am worried that I might not perform well.

Items of this sort are, in fact aimed right at the underlying concern that stereotyped groups have. If this were so, it is interesting to speculate that they (the third quartile – enhanced threat group) are the most vulnerable to outside cues. Are they the group that is in danger of dis-identifying with the domain? Stereotype threat theory predicts that at each level of achievement some members of the stigmatized group drop out of the domain (Steele, 1997); which then reinforces the stigma for those who continue in the domain. It might be helpful to identify the group at risk. Is there further evidence that, for this population, it is the third quartile group who is at risk? Figure 8, (math test scores) also shows some anomalous behavior for the third quartile (and the second quartile as well). For both threat states it is the third quartile that scores the highest,

and the second quartile that scores the lowest. Unfortunately this does not support the basic stereotype threat prediction that those who are the most threatened in a particular circumstance (3^{rd} quartile) perform least well. At this point the high PAP- motivation scores for the third quartile – enhanced threat group appears to be a statistical anomaly, however this group might be interesting as the focus of future research.

To summarize; the differences in the enhanced and reduced threat females are:

1) Higher mastery endorsement among those in the enhanced threat condition.
2) A tendency toward greater endorsement of PAP- goals among those in the enhanced threat condition.
3) A clear pattern of increasing endorsement of PAP+ goals as domain identification increases for those in the enhanced threat condition. No such relationship exists for those in the reduced threat. There is no corresponding increase in math test score as domain identification increases for either condition.

Other findings include

1) An increase in mastery endorsement (for both threat conditions) as identification with the domain of mathematics increases.
2) Significantly lower scores (compared to other motivation subscales) for PAV for both conditions but particularly for the enhanced threat group.

Perhaps it might be more informative to think of the reactions (or lack thereof) of the reduced threat group. These are the people who are operating outside of the "assumed" presence of the threat, which has been shown to operate even when not explicitly activated (Steele & Aronson, 1995). Why do they perform better? The general trend toward less reported motivation is interesting of course, but the most striking thing is actually the lack of response for PAP+ across math identification levels for the reduced threat group. They are "steady" in their motivation to show that they can perform better than others, even in a situation where they have been told that the whole purpose of the study was to examine why males sometimes outperform females. The reported PAP+ motivation level for all identification levels for this group is, in fact, essentially the "background" motivational level for all subscales (see figures 12 and 13). The reduced threat groups' reaction to the situation (particularly when viewed across the range of identification with the domain of mathematics) is muted. They seem to be "missing something" that the enhanced threat group has. Yet the reduced threat group outperforms the enhanced threat group on the math test. Whatever the reduced threat group is missing seems to be an inhibitor of cognition and performance.

Cognitive Load

Cognitive load theory (CLT) rests on fundamental ideas about the way the brain acquires, stores, and retrieves information. Working memory (in which all conscious information processing occurs) is very limited in the number of elements it can manage (Paas, Renkl & Sweller, 2003). In complex situations, where many

interrelated elements must be processed simultaneously, work is possible because schemata (stored in long term memory) can be retrieved (if they exist for this type of activity) and used in working memory. Schemata are cognitive structures that combine multiple elements into a single structure. Algebraic manipulations would be an example. This skill consists of many related bits of knowledge about the nature of equalities, order of mathematical operations, and the concept of "inverse operations" (e.g. the inverse of multiplication is division). The skill is difficult to learn and requires close attention at first; but through practice solving an equation for a particular variable becomes quite easy (to the point of automation). It is a schema that when activated requires little attention. The activity goes from being high in cognitive load to being low in load. A person can, eventually, do or think about other things (like talk, or think about the next part of the problem) while manipulating variables in an equation. Difficult activities are difficult because they approach or exceed the carrying capacity (cognitive load) of working memory. Schema acquisition and activation is therefore key to performance of challenging tasks.

CLT posits at least three different kinds of cognitive load (Paas et al. 2003). Intrinsic load imposes demands on working memory based on the material itself. Extraneous load occurs when situational circumstances cause a person to devote cognitive resources to material that is unnecessary for the actual activity. Germane load is also situational, and may not be directly related to the specific problem, but it is information that helps in the retrieval and use of relevant schema. The three forms of load are additive; however extraneous and germane load tend to have little effect on task performance in situations where the intrinsic load is low.

Situations in which stereotype threat is relevant are by definition high in intrinsic cognitive load in that there must be a high enough level of difficulty associated with the task that people highly identified with the field feel challenged. Can increased endorsement of any type of motivational goal be considered extraneous load? That is, can increased motivation itself interfere with the retrieval and use of schema needed to perform a task (despite any affect or behavior that might ordinarily be considered "positive")? This seems counterintuitive, but when it is recalled that STT occurs during the actual performance of a task (as opposed to during learning or preparation for an activity) it seems more reasonable. "I *must* succeed *right now*" is often detrimental to actual performance.

Beilock, Kulp, Holt, and Carr (2004) investigated "choking" in mathematical problem solving. They found that pressure leads to underperformance in situations where participants were doing unpracticed problems under circumstances that made heavy demands on working memory. They further found that it was the distraction caused by thoughts *about the situation* (rather than changes in the way participants approached the problems) that mediated the underperformance. This finding would seem to indicate that for stereotype threat situations reduction in stigma related motivation (via explicit reduction of threat) might reduce cognitive load enough to improve performance. It is after all the reduced threat groups' performance that is "different". The ubiquitous presence of the threat, and any additional motivation associated with dealing with the threat, may explain why self-report surveys fail to show significant results so often. The threat-motivation is just part of the normal operating condition for the stigmatized person. The extraneous cognitive load due to

this extra motivation is just taken for granted, and it is the sudden absence of that load that is strange.

Models of Stereotype Threat and Achievement Motivation

This study was intended to investigate conceptual models where the effect of stereotype threat on performance is mediated by differing levels of performance avoidance goal endorsement. Figures 15 and 16 (below) represent models proposed by Smith (2004) and Ryan & Ryan (2005). While the two models differ in structure with regards to the antecedents of PAV adoption and the exact details and path structure regarding why PAV adoption would produce poor results, the central feature of both is PAV adoption. The present study failed to find any differential endorsement of PAV across threat conditions. In fact (as discussed previously) the only interesting findings that involved performance avoidance were that scores on that subscale were significantly (or nearly significantly) lower than the scores on other subscales for both enhanced and reduced threat women.

There was however a significant difference in mastery (M) endorsement across threat condition, and as shown previously (Figure 9, page 71) M score does mediate threat state and performance for this group.

These findings, along with the additional result that the threat condition groups' PAP+ subscale scores have differing functional responses when mathematical domain identification is taken into account, indicate that these models may need the following revisions.

Smith's Task Engagement Process

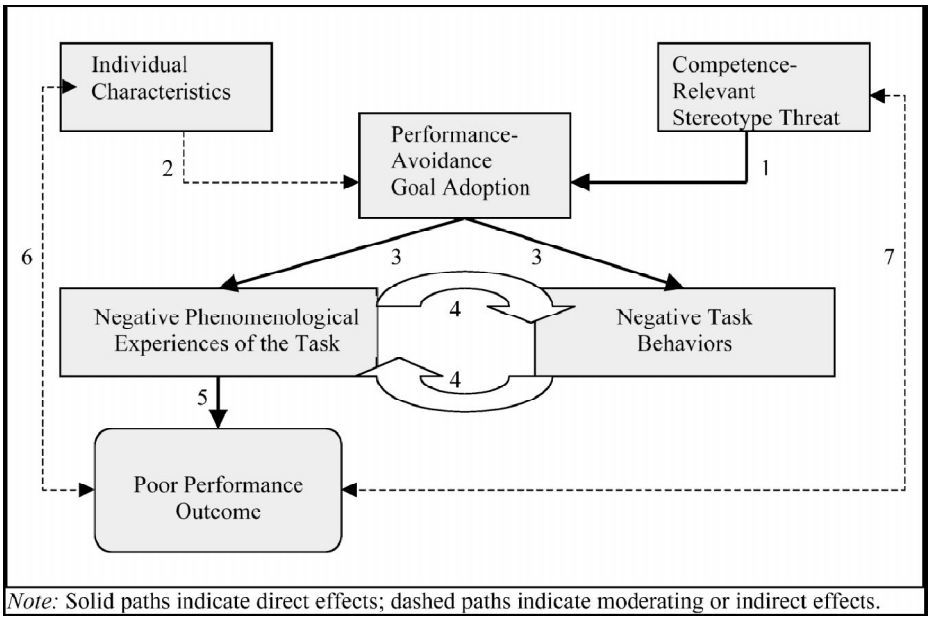


Figure 15. The Stereotyped Task Engagement Process Model (Smith 2004)

Ryan and Ryan Model of PAV Mediation of Stereotype Threat on Math Performance

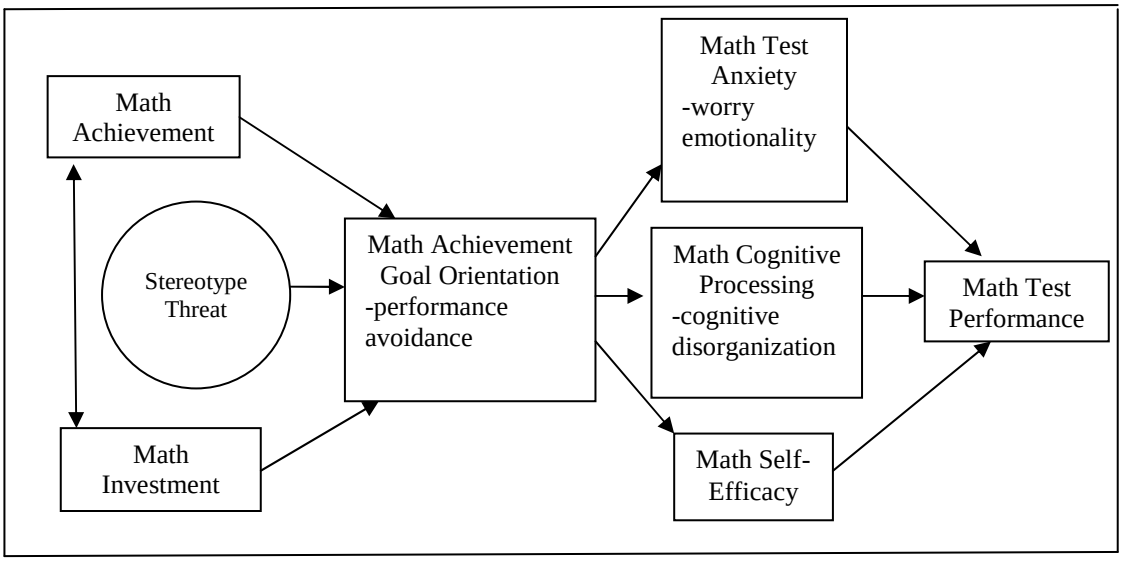


Figure 16. Conceptual Model of Psychological Processes Underlying Stereotype Threat and Standardized Math Test Performance (Ryan & Ryan, 2005)

The mediating effect of changes in achievement goal profile as a result of enhanced or reduced stereotype threat appears to be broader based than simply PAV goals alone. Increased motivation of any kind (in the case of this study M (and suggestive of PAP-), but in the cases of Chalabaev, et al (2008) and Brodish & Devine, (2009) PAV) seems to negatively impact performance in the threatened group. The proposed mechanism involves cognitive load, rather than simply the affect and behaviors associated with various motivational dimensions. This is implied in the Ryan & Ryan model but is only applied to PAV.

Negative behaviors, affect, phenomenological experiences, anxiety and self efficacy problems, while possible are not uniform results of the various forms of goal

adoption. This may explain the varying results of studies that attempted to measure these constructs in situations in which stereotype threat was important. Increases in cognitive load can occur with any form of motivation if the need" for motivation becomes evident.

The importance of domain identification should be explicit in the models since identification level, threat state and motivation, in this case PAP+, interact with one another. The Ryan model in effect addresses this through its emphasis on past math achievement and math investment. The Smith model addresses this more obliquely through the "Individual Characteristic" reference.

A revised model might look something like Figure 17 below.

While the focus of this model is on stereotype threat and its effects it also allows for the possibility of poor performance in the absence of a salient threat. Solid black lines indicate "multiplicative constructs" that is, if box A is connected to box B by a solid black line and A is zero then B is zero. Dashed lines indicate that the constructs are related, but not multiplicatively. So, if the task is not challenging, there is no relevant stigma present, or the person is not identified with the field stereotype threat is not activated and any effect on motivational goals must come from another route.

Model of Interactions of Stereotype Threat and Affects on Math Performance

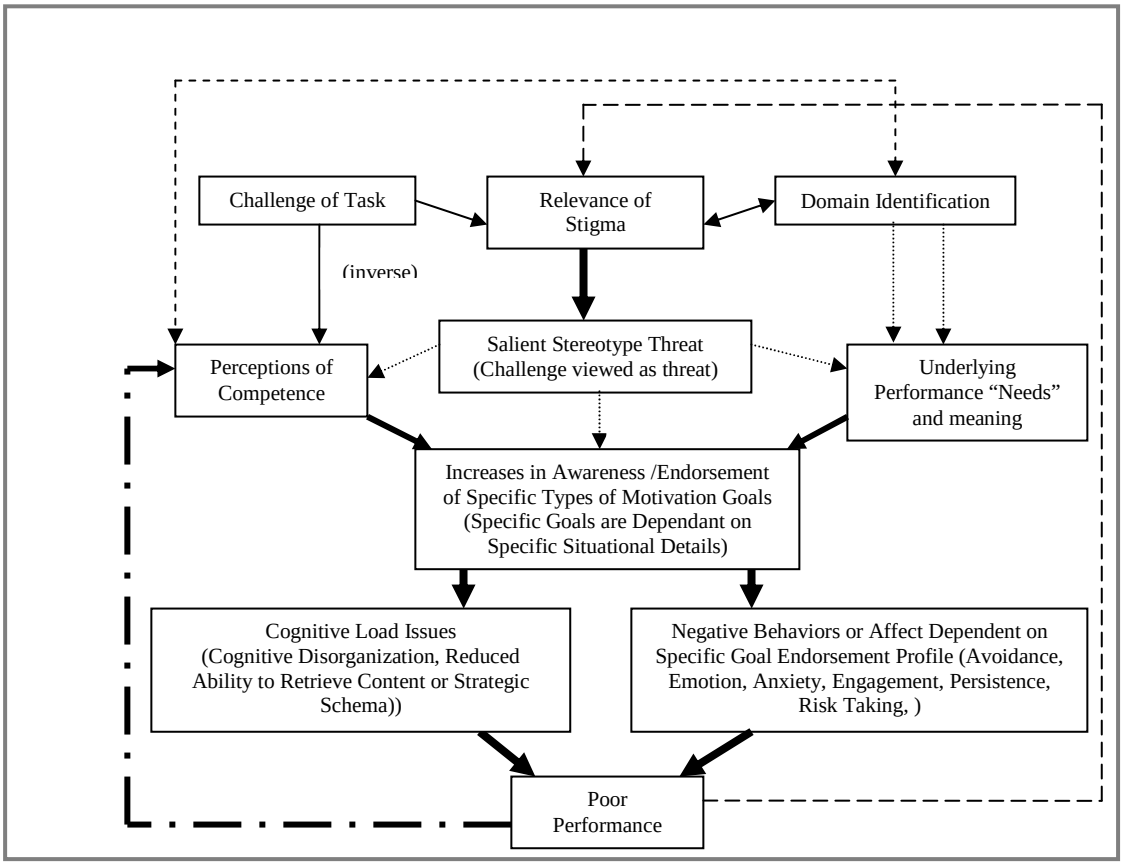


Figure 17. Mediation of Stereotype Threat by Situation Dependent Achievement Goals

However, if all three (challenge, stigma and domain identification are present) stereotype threat is activated, the situation is seen as threatening and motivational goals (or the awareness of "outcome needs") change. The model really emphasizes domain identification in that the construct is connected to choice and magnitude of motivational goals by three different interactions. It affects the salience of any stigma, perceptions of competence in the field, and underlying performance needs. However

the effects are not necessarily all bad. High domain identification should increase a person's perception of their task competence…while at the same time increasing the negative effects associated with stereotype threat. Thus, when asked, someone who is highly identified with a field will say they feel good about a performance even when the stigma is present. This is how the model attempts to cope with the "contradictory" nature of stereotype threat results.

When the antecedents to changes in the magnitudes of any of the various goal types include stereotype threat those changes produce maladaptive (negative) results, either through specific patterns of behavior associated with negatively valenced goals, or through the effects of cognitive load (as a person becomes "aware" of their motivation). Of course it is also possible for both paths of action to be activated at the same time.

This model differs from that proposed by Ryan and Ryan (Figure 16) in two ways. The first is that this model does not rely on the adoption of PAV goals solely. Rather any change in motivational state has the potential to impact performance in a negative way. The second difference is in the role of self-efficacy. In the Ryan model self – efficacy is directly affected by the adoption of PAV goals and plays a direct role in reducing performance. In the proposed model self – efficacy does not play this direct role. This is due to the fact that self-efficacy has not been shown to suffer in STT situations (in this and other studies). Additionally, in the proposed model self-efficacy could actually increase depending on the specific changes in goal structure, without a corresponding improvement in performance.

The model also indicates that poor performance will affect perceptions of competence (for later tasks), and ultimately domain identification.

Not shown are outside influences on the stigma's relevance such as dwindling numbers of females (contemporaries, role models) as a person progresses further and further into the "profession" of the domain of mathematics (or related fields).

Future Research Indicated by the Results of This Study

A number of different possible research directions are suggested by the results of this study.

Adding valenced mastery orientation (mastery approach (MAP) and mastery avoidance (MAV)) to the analyses. This study only considered M (operationally the same as MAP). This was primarily due to a focus on performance endorsement (PAV specifically). This followed the suggestions of the published models and extant literature and seemed sensible given the overt performance framing of the study, and frankly the rather ambiguous nature of MAV. The strong results showing increased M endorsement coupled with decreased performance by females in an enhanced stereotype threat state was a surprise. A future study that focuses on the valence of goals rather than on the competence measure aspect (of goals) would be interesting. How do the negatively valenced goals (PAV, MAP and PAP-) change as group (and individually) compared to the positively valenced goals of MAP and PAP+? An attempt was made to do such a group analysis with the data in hand, however no *additional* information was gained (primarily due to the already discussed,

rather overwhelming, increase in positively valenced goal structures (M and PAP+) for the high threat group and as math identification increased). Still, a full-blown study using Elliot's full 2x2 model (plus PAP-) would likely yield interesting results.

A follow up with the group of students who participated in this study to see if they proceeded to, and were successful at, calculus. This would be particularly interesting if it revealed differential patterns of future success (completion of course or sequence) based on either math identification quartile or motivation profiles from the current study.

An attempt to test the cognitive load hypothesis by manipulating the levels of endorsement for various motivational goals through direct methods and situational cues to test for *increases* in performance associated with increased motivation of any type (other than PAV). Such a finding would cast into doubt the ubiquity of cognitive load issues as a result of increases in motivation.

An extension of the present study to simply gather more data (N) to see if the tendency to report higher PAP- scores (among the enhanced threat females) becomes significant.

Conclusion

People are creatures of their environments. We are remarkably adept at picking up the subtle hints in a situation that inform us about how to behave; what the possible

outcomes of that behavior will be; and how (or if) others will judge that behavior. This is a positive thing much of the time. In social interactions of all kinds our ability to read slight changes in facial expression, tone of voice or body language allows us navigate the tricky waters of interpersonal interaction. We chose our responses based on the outcomes that we desire at the moment; moving away (or toward) negative things (or positive things) in the interaction as deemed appropriate at that time. The interpretation of cues and the choice of response is nearly instantaneous in familiar situations. Our individual "happiness" is closely tied to this ability as we often get immediate feedback when we fail (or succeed) to do this well. Most of us learn to do this at least adequately if not quite well. The inability to read these cues is a major detriment; and those who lack this skill are regarded as "oblivious" at best and "obnoxious" (or worse) at worst. The stakes are high and the learning environment, close to ideal.

This skill can betray us however. In cases where we operate under a known stigma our knowledge of the stigma (and the corresponding cues that tell us that the stigma is relevant) can cause us to perform poorly in situations we care about. In these cases being oblivious may have its benefits. This seems to be the case with the phenomena of stereotype threat where the knowledge of a relevant stigma, unless specifically debunked, causes under performance among the afflicted group. In academic situations the people involved are often very adept at reading and interpreting cues about their performance and its meaning, and they are aware of how they should act and/or think in order to obviate the stigma. They would (as often as not) prefer not to acknowledge the stereotype, they do not believe it and certainly do

not believe it applies to them personally. Yet they are aware of the "threat in the air" and devote cognitive resources in determining how to deal with that aspect of the situation. They do it automatically, as they would with any other aspect of the situation that might be important.

This study illustrates the conundrum faced by women in mathematics (and others operating under stigmata). Those that are in an overtly threatening situation react in a rather positive way (in this study at least); reporting higher levels of desire to show mastery of the material and to approach the performance with the intent of doing well enough to outperform others. They do not, as a group, report increased desire (compared to those in the reduced threat state) to avoid the situation, nor do they report changes in self-efficacy with respect to the material; yet, they are significantly outperformed by those in the reduced threat state. It is tempting to say that those in the reduced threat state are "oblivious" to the threat posed by the stigma. But that is not so. Both groups have been reminded of the stereotype, and have been told that some tests have shown that males perform better than females. The "reduced threat" group has simply been told that this is not so on their particular test. This seems to be enough however, to dampen the need to direct significant cognitive resources to dealing with the additional meaning of the performance (doing as well as the males). This is evident in the muted responses (as evidenced by the scores on the motivation scales) of the reduced threat group as their identification with the domain of mathematics increases. The enhanced threat group's level of PAP+ increases as identification increases, with the most identified women reporting the highest levels of PAP+. They react as if disproving the stigma rests squarely on them. This belief, and

the increase in motivation that accompanies it, does not lead to better performance. Better performance does accompany the reduction of the salience of the threat and the lessening of the need for high motivation of any type.

People are not automatons. They are aware of the stimulus in a situation, and of the range of possible behaviors and consequences implicit in the situation, but they are also aware of how they are supposed to act given who they are and who they want to be. In situations that challenge their ability any reduction in extraneous cognitive load could sensibly be hypothesized to increase performance. Such seems to be the case here if awareness of one's motivation is viewed as cognitive load.

REFERENCES

Aboud, F. (1988). *Children and prejudice*. New York, NY: Blackwell.

Allport, G. (1954). *The Nature of Prejudice.* New York, NY: Addison-Wesley

Aronson, J., Lustina, M. J., Good, C., Keough, K. A., 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.

Bandura, A. (1995). *Guide for constructing self-efficacy scales*. Retrieved from the World Wide Web: http://www.emory.edu/EDUCATION/mfp/banguide.html

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

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology, 133* (4), 584 - 600.

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science,* 12, 225 - 229.

Bridgeman, B. & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement, 31* (1), 37-50.

Brodish, A. B., & Devine, P. G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology, 45*, 180-185.

Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76* (2), 346 -357.

Brown, R. P., & Josephs, R. A. (2000). The importance of importance: The math identification questionnaire. Unpublished manuscript, The University of Texas.

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

Chalabaev, A., Sarrazin, P., Stone, J., & Cury, F. (2008). Do achievement goals mediate stereotype threat?: An investigation on female's soccer performance. *Journal of Sports and Exercise Psychology, 30,* 143-158.

Cook, S. W. (1979). Social science and school desegregation: Did we mislead the Supreme Court? *Personality and Social Psychology Bulletin, 5*(4), 420-437.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 94*(4), 608-630.

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

Dabbs, J. M., Jr. (1998). Testosterone and the concept of dominance. *Behaioral and Brain Sciences, 21*, 370-371.

Davies, P.G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology, 88* (2), 276-287.

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

Dweck, C. S., & Leggett, E. L., (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256-273.

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

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34* (3), 169-189.

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

Elliot, A. J., & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology. 80*: 501–519.

Elliot, A. J., McGregor, H.A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology, 91,* 549–563.

Florida Department of Education, (2007). SAT and ACT concordance tables. Retrieved from the World Wide Web: http://www.fldoe.org/evaluation/xls/actsatct.xls

Gerard, H. (1983). School desegregation: The social science role. *American Psychologist, 38,* 869-878.

Greene, B. A. & Miller, R. B. (1996). Influences on course achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology, 21,* 181-192.

Greene, B. A., Miller, R. B., Crowson, H. M., Duke, B. L. & Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. *Contemporary Educational Psychology 29,* 462–482.

Hall, C. W., Davis, N. B., Bolen, L. M., & Chia, R. (1999). Gender and racial differences in mathematical performance. *The Journal of Social Psychology 139,* 677-689.

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

Inzlicht, M. & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95* (4), 796-805.

Josephs, R. A., Newman, M. G., Brown, R. P.**,** & Beer, J. M. (2003). Stereotypes as status: Explaining the stereotype threat effect with testosterone. *Psychological Science, 14,* 158-163.

Kaiser, H. F. & Coffrey, J. (1965). Alpha factor analysis, *Psychometrika, 30,* 1-14.

Kane, T. J. (1998). Racial and ethnic preferences in college admissions. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 431-456). Washington, D.C.: Brookings Institution Press.

Katz, I., Epps, E. G., & Axelson, L. J. (1964). Effect upon Negro digit symbol performance of comparisons with Whites and other Negroes. *Journal of Abnormal and Social Psychology, 69,* 963-970.

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

Linnenbrink, E. A., & Pintrich, P. R. (2000). Multiple pathways to learning and achievement: The role of goal orientation in fostering adaptive motivation, affect, and cognition. In Sansone, C., and Harackiewicz, J. M. (eds.), *Intrinsic and Extrinsic Motivation: The search for optimal motivation and performance.* San Diego, CA: Academic Press

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: self-evaluation of one's social identity. *Personality and Social Psychology Bulletin, 18* (3) 302 - 318.

Major, B., & Schmader, T. (1998). Coping with stigma through psychological disengagement. In C. Stangor (Ed*.), Prejudice: The Target's Perspective*. New York, NY: Academic Press.

Mazur, A. & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences, 21,* 353-397.

Middleton, M., & Midgley, C. (1997). Avoiding the demonstration of the lack of ability: An underexplored aspect of goal theory. *Journal of Educational Psychology, 89,* 710-718.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A. Kumar, R., Middleton, M., Nelson, J., Roeser, R. & Urdan, T., (2000). *Manual for the Patterns of Adaptive Learning Survey (PALS).* Ann Arbor, MI: University of Michigan.

Miller, R. B., Behrens, J., Greene, B., & Newman, D., (1993). Goals and perceived ability: Impact on student valuing, self-regulation, and persistence. *Contemporary Educational Psychology, 18,* 2-14.

Miller, R. B., & Brickman, S. A. (2004). A model of future oriented motivation and self - regulation. *Educational Psychology Review, 16,* 9–33.

Miller, R.B., Greene, B.A., Montalvo, G.P, Ravidran, B., & Nichols, J.D. (1996) Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology 21,* 388-422.

National Science Foundation (2000). *Women, minorities, and persons with disabilities in science and engineering: 2000*. Arlington, VA: NSF 00-327.

Nicholls, J. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychology Review*, *91*, 328-346.

Osborne, J. W. (1997). Race and academic disidentification. *Journal of Educational Psychology, 89*, 728-735.

Osborne, J. W. (1999). Unraveling underachievement among African American boys from an identification with academics perspective. *Journal of Negro Education*, *68* (4), 555-565.

Osborne J. W. & Simmons C. M. (2002). Girls, math, stereotype threat, and anxiety: Physiological evidence. Poster presented at the national meeting of the American Education Research Association (AERA), New Orleans LA.

Osborne, J. W. & Rausch, J. L. (2001). Identification with academics and academic outcomes in secondary students. Poster presented at the national meeting of the American Education Research Association (AERA), Seattle WA.

Osborne J. W., Walker C., & Rauch, J. L. (2002). The interaction of race and identification with academics in predicting withdrawal from school in secondary students. Paper presented at the national meeting of the American Education Research Association (AERA), New Orleans LA.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38* (1), 1-4.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451- 502). San Diego, CA: Academic Press.

Pintrich, P. R. (2000b). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology*, *25*, 92–104.

Pintrich, P.R., & Garcia, T. (1991). Student goal orientations and self-regulation in the college classroom. In M. L. Maehr, and P. R. Pintrich (Eds.) *Advances in Motivation and Achievement Vol. 7* (pp 371-401). Greenwich CT: JAI Press.

Pintrich, P. R., & Schrunk, D. (1996). *Motivation in Education, theory research and applications,* Upper Saddle River, NJ: Prentice-Hall.

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

Ruble, D., & Martin, C. L. (1998). Gender development. *Handbook of Child Psychology*, *3*, 933–1016.

Ryan K. E., & Ryan R. M. (2005). Psychological processes underlying stereotype threat and standardized math test performance. *Educational Psychologist, 40*(1), 53-63**.**

Scaramella, T., & Brown, W. (1978). Serum testosterone and aggressiveness in hockey players. *Psychosomatic Medicine, 40*, 262-283.

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

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

Schunk, D.,& Pajares, F. (2001) The development of academic self-efficacy. In A.Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 15 - 31). San Diego, CA: Academic Press.

Senko, C., & Miles, M. M., (2008) Pursuing their own learning agenda: How mastery oriented students jeopardize their class performance. *Contemporary Educational Psychology, 33*, 561-583.

Simons, J., DeWitte, S., & Lens, W. (2000). Wanting to have vs. wanting to be: The effect of perceived instrumentality on goal orientation. *British Journal of Psychology*, *91*, 335–351.

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**.,** Morgan, C. L, & White, P. H. (2005). Investigating a measure of computer technology domain identification: A tool for understanding gender differences and stereotypes. *Educational and Psychological Measurement, 65*, 336 - 355 .

Smith, J. L. (2006). The interplay among stereotypes, performance – avoidance goals, and women's math performance expectations. *Sex Roles*, *54* (3/4) 287-296.

Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotype task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology*, *99*(1), 99-114.

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

Steele, C. M. (1992, April). Race and the schooling of Black Americans. *The Atlantic Monthly*, 68–78.

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

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

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*, 323–326.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah NJ: Lawrence Erlbaum Associates Inc.

Wrisberg, C. A., (1994). The arousal-performance relationship. *Quest*, *46* (1). 60-77.

Appendix A – Factor Analyses and Motivation Scales

Table A1

Rotated Factor Matrix(a) for Identity Scales

	Factor	
	Gender	Math Domain
Ident 1		.740
Ident 2	.801	
Ident 3		.362
Ident 4		.797
Ident 5	.913	
Ident 6		.699
Ident 7		.767
Ident 8	.851	
Ident 9		.670
Ident 10	.878	

Method: Alpha Factoring.

Rotation Method: Varimax with Kaiser Normalization.

Table A2

Rotated Factor Matrix for Motivation Scales

	Factor			
	1	2	3	4
G1		.627		
G2				.535
G3			.531	
G4	.687			
G5			.453	
G6			.683	
G7	.541			
G8				.554
G9		.618		
G10			.610	
G11	.774			
G12		.719		
G13	.785		.	
G14				.679
G15		.693		
G16			.563	
G17	.670			
G18				.710

Extraction Method: Alpha Factoring.

Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 15 iterations.

Table A3

Item Statistics for Motivation Sub Scales (Four Factor Model)

Scale (item)	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
M					
G4	10.21	9.611	.662	.443	.818
G11	10.48	8.722	.700	.493	.802
G13	10.60	8.517	.708	.505	.798
G17	10.81	9.057	.680	.466	.810
PAP +					
G1	9.47	10.516	.503	.258	.879
G9	9.98	8.561	.665	.462	.821
G12	10.23	7.515	.803	.712	.758
G15	10.24	7.582	.812	.724	.754
PAV					
G3	9.02	10.004	.498	.253	.774
G6	9.42	9.006	.603	.365	.722
G10	8.75	9.077	.643	.439	.700
G16	9.08	9.881	.618	.418	.717
PAP-					
G2	9.90	8.449	.576	.359	.769
G8	10.01	7.790	.575	.379	.769
G14	10.60	7.559	.618	.482	.748
G18	10.39	7.079	.691	.536	.710

Appendix B– Descriptive Statistics for Math Test Score (for analysis where 24 was not used for missing ACT Math Scores)

Dependent Variable: Math Test Score

Group	Mean	Std. Deviation	N
Reduced threat Females	4.93	2.726	42
Reduced threat males	4.96	2.424	50
Enhanced threat females	3.66	1.783	41
Enhanced threat males	4.29	1.865	42
Total	4.49	2.289	175

Appendix C – Notes on Alternative Regressional Analysis Method

A multiple regression analysis was also attempted in lieu of the MANOVAs discussed in the main body of the dissertation.

IV's – Threat State and Math Identification

Since threat state is categorical some sort of dummy coding had to be used. Additionally, a third variable had to be created to account for the interaction term for threat state and mathematical domain identification. Effect coding was chosen as the methodology. This creates one new variable (since my categorical variable has only two states).

	EFF 1
Enhanced State	-1
Reduced State	+1

This will make the enhanced state the base group.

*Comparisons for sig differences are done with reference to the grand mean.

Comparisons are not reported for the base group.

So the output compared the reduced state mean for the DV to the DV mean for the whole data set.

The interaction term is arrived at by multiplying the two variables so that would be MATH ID * EFF 1 = INT

Regressions were done (for each DV (motivation scale)) independently by entering the variables in "stepwise" (in block one) with the interaction term in block two. At each step (or block) I checked to see if a significant amount of new variance was explained over and above the previous block.

After performing the regression analyses the results were essentially the same as for the MANOVA. The significant result was that for females, the high threat state had a higher mastery motivation score and a lower math test score than the reduced threat state. The regression for M showed a significant amount of variance accounted for by math identification, and further, by threat state, but not the interaction. For PAP+ the only significant variance was accounted for by math identification. This is a bit different than in the MANOVA output where the interaction between math identification and state came up as significant or approaching significance (at least for some scales…which the regression did not identify).

Discussion of regression technique vs. MANOVA:

Regression

a. Disadvantages: More error since an analysis has to be done for each motivation subscale. This technique forces the choice of a "reference state" which is not really appropriate for this material. All participants were manipulated to some extent. Gives math identification preference of place over threat state, which does not seem appropriate.

b. Advantages: More power in that more information is retained on the math identification scales.

Since both approaches give the same basic results and the MANOVA hints at the interaction of math identification and threat state I choose to report the MANOVA results.

Appendix D - Informed Consent Form and Instrumentation

D1 Informed Consent Form

University of Oklahoma
Institutional Review Board
Informed Consent to Participate in a Research Study

Project Title:	**Investigating Why Some Mathematics Tests Show Differences in Scores Based on Sex**
Principal Investigator:	C. Max Simmons
Department:	Educational Psychology

You are being asked to volunteer for this research study. This study is being conducted at Oklahoma City Community College, and the University of Oklahoma. You were selected as a possible participant because of your enrollment in trigonometry or a calculus preparatory algebra course.

Please read this form and ask any questions that you may have before agreeing to take part in this study.

Purpose of the Research Study
In the past some standardized mathematics exams have shown a bias toward males while other tests of the same sort have not. This study looks at items from tests that do and do not show this bias in an effort to determine what accounts for the difference.
Number of Participants
About 600 people will take part in this study.
Procedures
If you agree to be in this study, you will be asked to fill out a brief survey during class and then go to a Testing Center where you will fill out a brief demographic questionnaire; respond to a series of 20 questions taken from standardized mathematics tests such as the ACT, SAT, or GRE; and take two brief surveys.
Length of Participation
The in-class survey should take 15 minutes or less. The session in the Testing Center should take less than 40 minutes. Total participation time is 55 minutes or less.
This study has the following risks:
This study has no risks associated with it, however anyone can end their participation in the study at any time.

Benefits of being in the study are
The benefits to participation are a better general understanding of how to structure and administer exams such as the ones mentioned above. This type of research can also impact classroom testing procedures. These benefits are unlikely to apply directly to you but will help students in the future.
Confidentiality
In published reports, there will be no information included that will make it possible to identify you without your permission. You will be asked for your Institutional ID number (see below) and your name at this time. Your name will be recorded separately so that your instructor can be notified of your participation in the study. Your Institutional ID will be used to obtain your ACT or SAT math score. After those scores are obtained the portion of this form that contains that information will be detached and destroyed. The remaining portion of this form will be filed securely and will not be linked in any way to published data. After agreeing to participate in the study you will be given a randomly assigned study ID that will be used on all test and

survey results (for tracking purposes). Research records will be stored securely and only approved researchers will have access to the records.

There are organizations that may inspect and/or copy your research records for quality assurance and data analysis. These organizations include the OCCC Institutional Review Board and the OU Institutional Review Board.

Compensation

You will be reimbursed for your time and participation in this study through credit assigned to you in your class. The credit varies from section to section but is in the range of 5 to 10 points. Credit is given at the completion of your participation in the study and is not based on your score on the mathematics test or any response to any other question during the study. You must complete both the in-class survey and the portion of the study in theTest Center to receive credit. Your instructor or the investigator will provide you with the specific details for your class before you sign this form. If you have questions please contact me at the email address provided.

Voluntary Nature of the Study

Participation in this study is voluntary. If you withdraw or decline participation, you will not be penalized or lose benefits or services unrelated to the study. If you decide to participate, you may decline to answer any question and may choose to withdraw at any time during either the in class session or the portion of the study that takes place outside of class.

Contacts and Questions

If you have concerns or complaints about the research, the researcher(s) conducting this study can be contacted at

C. Max Simmons: (405) 650-3992 or cmsimmons@ou.edu

Dr. Raymond Miller (405) 325-1501 or rmiller@ou.edu

Contact the researcher(s) if you have questions or if you have experienced a research-related injury.

If you have any questions about your rights as a research participant, concerns, or complaints about the research and wish to talk to someone other than individuals on the research team or if you cannot reach the research team, you may contact the University of Oklahoma – Norman Campus Institutional Review Board (OU-NC IRB) at 405-325-8110 or irb@ou.edu.

You will be given a copy of this information to keep for your records. If you are not given a copy of this consent form, please request one.

Statement of Consent

I have read the above information. I have asked questions and have received satisfactory answers. I consent to participate in the study.

____

Signature **Date**

- -

I give permission for C. Max Simmons or Dr. Raymond Miller to access my ACT or SAT scores as part of the study: **Investigating Why Some Mathematics Tests Show Differences in Scores Based on Sex.**

Name (Print Please)

Student ID

Signature

D2 Identification Survey

(Color not seen on participant version)

About Me

For the questions below, indicate the extent to which you agree or disagree with each statement by writing a number from the following scale in the blank next to the questions: Many items on this survey are similar to others so that we can be sure we understand your perspective.

1----------2----------3----------4----------5----------6----------7
strongly DISagree — strongly agree

*1.___ My math abilities are very important to me.

+2.___ My gender has a lot to do with how I feel about myself.

*3._____ If I took an IQ test of my math abilities and I did poorly on this test, I would be very bothered.

*4.___ Math abilities will probably be very important to me in my future career.

+5.___ An important part of my self-image is my gender.

**6.__ I enjoy math related activities.

**7.__ I will be likely to take a job in a math related field.

+8.___ My gender is an important reflection of who I am.

**9.__ Math ability is important to the sense of who I am.

+10.__ My gender is important to my sense of what kind of person I am.

*From Brown's MIQ (Math identity)

**Adapted From Smith, Morgan and White (2005) (Math Identity)

+From Schmader (2002)…adapted from Luhtanen and Crocker (1992) (Gender Identity) Items 2 and 10 have been altered so that they are "positively worded like all the other items.

D3 Self –Efficacy Instrument

(Adapted from the *Guide for Constructing Self-Efficacy Scales* (Bandura, 1995))
Appraisal Inventory

The following four items are the same type that you will be asked to solve in the upcoming math test In the answer space provided please rate how sure you are that you can solve the problem in one minute or less. Rate your degree of confidence by recording a number from 0 to 100 in each blank using the scale given below. Please be sure to hit "submit" for each answer.

0	10	20	30	40	50	60	70	80	90	100
Cannot do at all				Moderately certain can do						Certain can do

1. $4/(1/2) + 2/(2/3) + 3/(3/8) =$

2. How many positive integers are both multiples of 3 and divisors of 36?

3. For each of *n* people Sam bought a hamburger and a soda at a restaurant. For each of *n* people Laura bought 2 hamburgers and a soda at the same restaurant. If Sam spent a total of $5.40 and Laura spent a total of $12.60 how much did Sam spend for just hamburgers? (Assume that all hamburgers cost the same and all sodas cost the same).

4. What is the least integer value of n such that $(1/2^n)$ is less than 0.001?

D4a Achievement Goal Questionnaire – Reduced Threat Version

Mastery – items 4, 7, 11, 13, 17
PAP+ - items 1, 5, 9, 12, 15
PAP- - items 2, 8, 14, 18
PAV –items 3, 6, 10, 16

My Goals

Standardized tests evaluate you based on how you perform relative to other students who took the same exam. The math test that follows this survey is one such exam that shows women and men doing equally well.

The following are questions about your goals as you take the math test (with items like the ones you just saw in the previous survey).

Many items on this survey are similar to others so that we can be sure we understand your perspective. Please respond on a scale of 1 through 5 where:

1	**2**	**3**	**4**	**5**
NOT AT ALL TRUE		SOMEWHAT TRUE		VERY TRUE

1. It's important to me that my performance is better than that of other students.
2. I am concerned about performing poorly on this math test so I will work really hard at it.
3. One of my goals on this math test is to avoid showing I have trouble doing the work.
4. It's important to me that I learn something about my understanding of mathematics.
5. One of my goals is to perform like this test is easy for me.
6. One of my goals is to keep others from thinking I'm not smart concerning the math on this test.
7. It's important to me that I see evidence of my improving math skills.
8. My fear of not performing well on this test motivates me to try harder.
9. One of my goals is to demonstrate to others that I'm good at math.
10. It's important to me that I don't look stupid on this test.
11. It's important to me that I learn something new about math.
12. One of my goals is to perform well so that I look smart in comparison to other mathematics students.
13. One of my goals on this math test is to learn as much as I can.

14. I will work hard on this math test because I am concerned that I can't do the problems.
15. It's important to me that I do well so that I look smart compared to others in math.
16. It's important to me that people don't think that I know less than others in math.
17. One of my goals is to tackle new challenges on this test.
18. I will try hard on this math test because I am worried that I might not perform well.

D4b Achievement Goal Questionnaire – Enhanced Threat Version

Mastery – items 4, 7, 11, 13, 17
PAP+ - items 1, 5, 9, 12, 15
PAP- - items 2, 8, 14, 18
PAV –items 3, 6, 10, 16

My Goals

Standardized tests evaluate you based on how you perform relative to other students who took the same exam. The math test that follows this survey is one such exam that shows that men perform better than women.

The following are questions about your goals as you take the math test (with items like the ones you just saw in the previous survey).

Many items on this survey are similar to others so that we can be sure we understand your perspective. Please respond on a scale of 1 through 5 where:

1 NOT AT ALL TRUE **2** **3** SOMEWHAT TRUE **4** **5** VERY TRUE

1. It's important to me that my performance is better than that of other students.
2. I am concerned about performing poorly on this math test so I will work really hard at it.
3. One of my goals on this math test is to avoid showing I have trouble doing the work.
4. It's important to me that I learn something about my understanding of mathematics.
5. One of my goals is to perform like this test is easy for me.
6. One of my goals is to keep others from thinking I'm not smart concerning the math on this test.
8. It's important to me that I see evidence of my improving math skills.
8. My fear of not performing well on this test motivates me to try harder.
9. One of my goals is to demonstrate to others that I'm good at math.
10. It's important to me that I don't look stupid on this test.
11. It's important to me that I learn something new about math.
12. One of my goals is to perform well so that I look smart in comparison to other mathematics students.
13. One of my goals on this math test is to learn as much as I can.
14. I will work hard on this math test because I am concerned that I can't do the problems.

15. It’s important to me that I do well so that I look smart compared to others in math.
16. It’s important to me that people don’t think that I know less than others in math.
17. One of my goals is to tackle new challenges on this test.
18. I will try hard on this math test because I am worried that I might not perform well.

D5a

Math Test (from Inzlicht and Ben Zeev, 2003) Reduced Threat Version

The items on this exam have never shown a bias towards males.

Directions: Each of the questions (20) have five answer choices. For each of these questions select the best of the answer choices given. You have 20 minutes to complete this test .

.

---

2. 1/(1/2) + 2/(2/3) + 3/(3/4) =

   a. 1/9 b. 13/12 c. 29/12 d. 8 e. 9

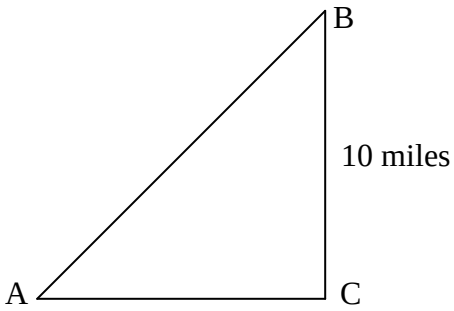
3. If n is not equal to 0 which of the following must be greater than n?

   I. 2n II. n^3 III. 4-n

   a. none of these b. I only c. II only d. I and II e. I and III

4. According to the figure below, traveling directly from point A to point B, rather than form point A to point C and then from point C to point B would save approximately how many miles?

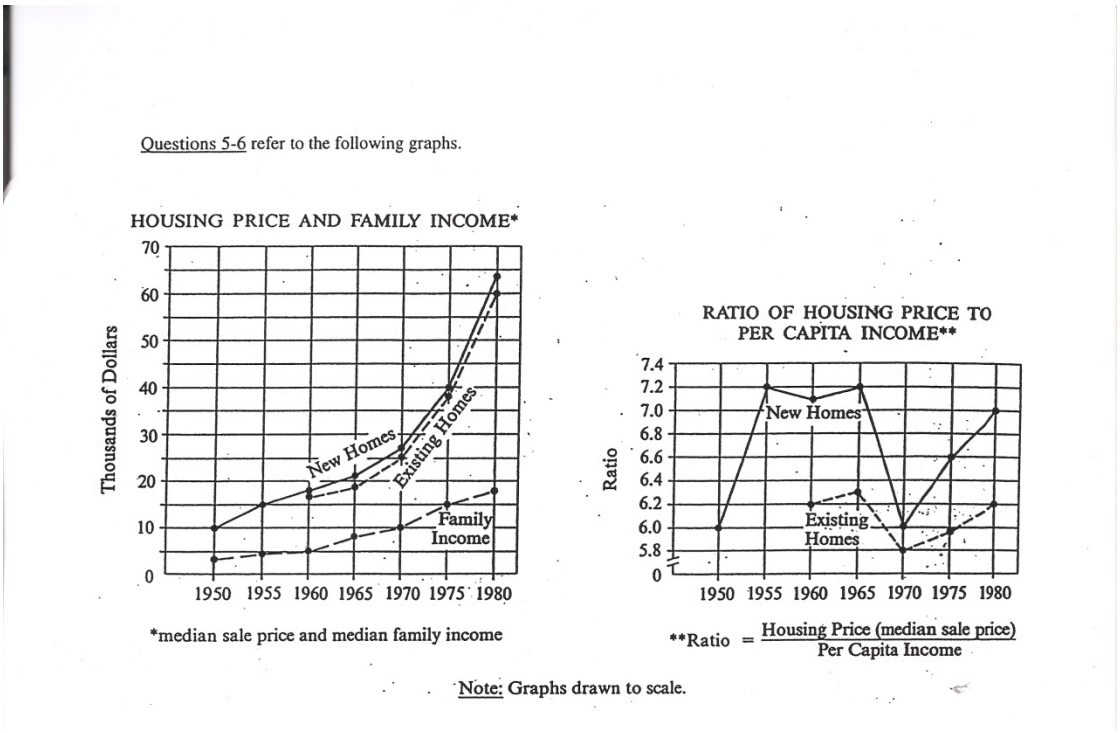
   a. 1 b. 2 c. 2 d. 4 e. 5



5. 0.50% =

a. 1/500 b. 1/200 c. 1/50 d. 1/20 e. 1/2

Questions 5-6 refer to the following graphs



6. If in 1985 the per capita income was $7200 and the ratio of the median sale price of an existing home to per capita income was the same as in 1980, what was the median sale price of and existing home in 1985?

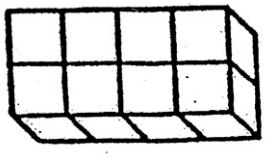
   a. $50,040 b. $44,640 c. $11,600 d. $5,040 e. $1,160

7. By approximately what percent did the median sale price of a new home increase from 1955 to 1975?

   a, 26% b. 37.5 % c. 62.5% d. 167% e. 267%

8. The distance from point X to point Y is 20 miles and the distance from point X to point Z ix 12 miles. If *d* is the distance, in miles, between points Y and Z, then the range of possible values for *d* is indicated by

   a. 8 to 20 b. 8 to 32 c. 12 to 20 d. 12 to 32 c. 20 to 32



9. The rectangular solid above is made up of eight cubes of the same size, each of which has exactly one face painted blue. What is the greatest fraction of the total surface area of the solid that could be blue?

   a. 1/6 b. 3/14 c. 1/4 d. 2/7 e. 1/3

10. What is the least integer value of n such that $(1/2^n)$ is less than 0.01?

    a. 7 b. 11 c. 50 d. 51 e. There is no such least value

11. A distillate flows into an empty 64-gallon drum at spout A and out of the drum at spout B. If the rate of flow through A is 2 gallons per hour, how many gallons per hour must flow out at spout B so that the drum is full in exactly 96 hours?

    a. 3/8 b. 1/2 c. 2/3 d. 4/3 e. 8/3

12. If $a > 0$, $b > 0$, and $c > 0$ then $a + (1/(b + 1/c)) =$

    a. $(a + b)/c$ b. $(ac + bc + 1)/c$ c. $(abc + b + c)/bc$ d. $(a + b + c)/(abc + 1)$ e. $(abc + a + c)/(bc + 1)$

13. If $L = (a-b) - c$ and $R = a- (b-c)$ then $L - R =$

    a. 2b b. 2c c. 0 d. -2b e. -2c

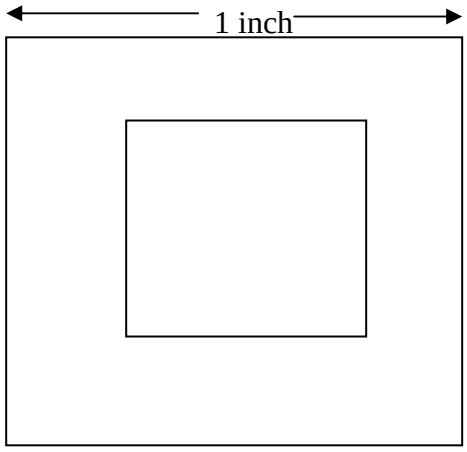
14. For each of *n* people Maggie bought a hamburger and a soda at a restaurant. For each of *n* people Paul bought 3 hamburgers and a soda at the same restaurant. If Maggie spent a total of $5.40 and Paul spent a total of $12.60 how much did Paul spend for just hamburgers? (Assume that all hamburgers cost the same and all sodas cost the same).

a. $10.80 b. $9.60 c. $7.20 d. $3.60 e. $2.40

15. The buyer of a certain mechanical toy must choose 2 of 4 optional motions and 4 of 5 optional accessories. How many different combinations of motions and accessories are available to the buyer?

a. 8 b. 11 c. 15 d. 20 e. 30

16. In the figure below, if the area of the smaller square region is ½ the area of the larger square region, then the diagonal of the larger square is how many inches longer than the diagonal of the smaller square?



a. sqrt2 -1 b. ½ c. sqrt2/2 d. (sqrt2 +1)/2 e. sqrt2

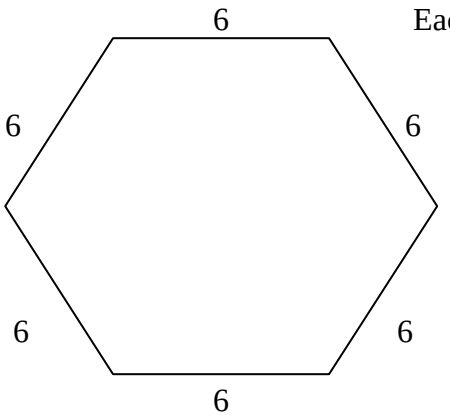
16. A certain recipe makes enough butter for exactly 8 circular pancakes that are 10 inches in diameter. How many circular pancakes, each 5 inches in diameter and of the same thickness as the 10-inch pancakes, should the recipe make?

a. 4 b. 16 c. 24 d. 32 e. 40

17. What is the area of the hexagonal region shown in the figure below?

a. 54 (sqrt 3) b. 108 c. 108 (sqrt3) d. 216

e. It cannot be determined from the information given



18. A farmer has two rectangular fields. The larger field has twice the length and 4 times the width of the smaller field. If the smaller field has area K, then the area of the larger field is greater than the area of the smaller field by what amount?

a.. 2K b. 6K c. 7K d. 8K e. 12K

19. A board of length L is cut into two pieces such that the length of one piece is 1 foot more than twice the length of the other piece. Which of the following is the length, in feet, of the longer piece?

a. (L+2)/2 b. 2L + ½ c. L-1/3 d. 2L + 2/3 e. (2L + 1)/3

20. How many positive integers are both multiples of 4 and divisors of 64?

a. two b. three c. four d. five e. six

D5b

Math Test (from Inzlicht and Ben Zeev, 2003) Enhanced Threat Version

In the past, males have been shown to perform better on this test than females.

Directions: Each of the questions (20) have five answer choices. For each of these questions select the best of the answer choices given. You have 20 minutes to complete this test .

1. 1/(1/2) + 2/(2/3) + 3/(3/4) =

   a. 1/9 b. 13/12 c. 29/12 d. 8 e. 9

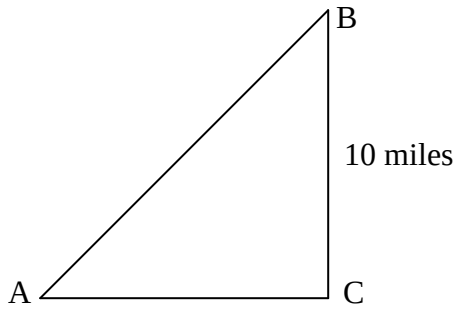
2. If n is not equal to 0 which of the following must be greater than n?

   I. 2n II. n^3 III. 4-n

   a. none of these b. I only c. II only d. I and II e. I and III

3. According to the figure below, traveling directly from point A to point B, rather than form point A to point C and then from point C to point B would save approximately how many miles?

   a. 1 b. 2 c. 2 d. 4 e. 5

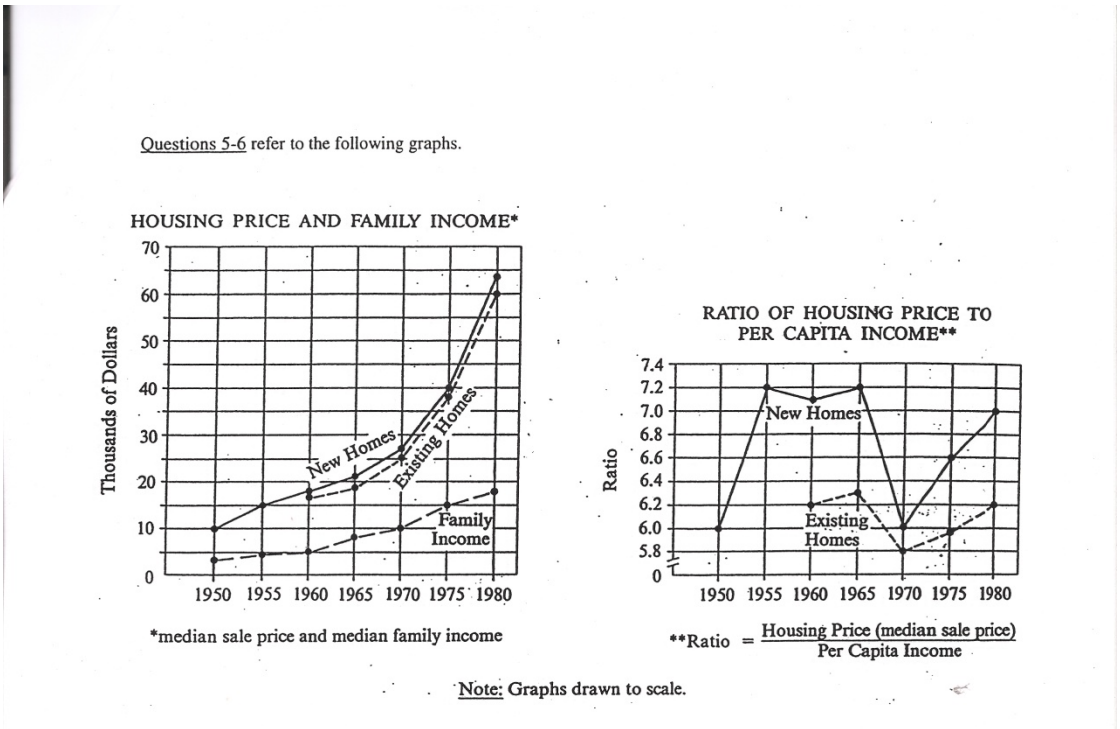


8 miles

4. 0.50% =

   a. 1/500 b. 1/200 c. 1/50 d. 1/20 e. ½

Questions 5-6 refer to the following graphs



5. If in 1985 the per capita income was $7200 and the ratio of the median sale price of an existing home to per capita income was the same as in 1980, what was the median sale price of and existing home in 1985?

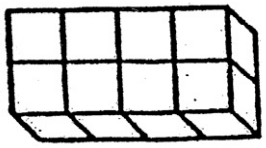
   a. $50,040 b. $44,640 c. $11,600 d. $5,040 e. $1,160

6. By approximately what percent did the median sale price of a new home increase from 1955 to 1975?

   a, 26% b. 37.5 % c. 62.5% d. 167% e. 267%

7. The distance from point X to point Y is 20 miles and the distance from point X to point Z ix 12 miles. If *d* is the distance, in miles, between points Y and Z, then the range of possible values for *d* is indicated by

   a. 8 to 20 b. 8 to 32 c. 12 to 20 d. 12 to 32 c. 20 to 32



8. The rectangular solid above is made up of eight cubes of the same size, each of which has exactly one face painted blue. What is the greatest fraction of the total surface area of the solid that could be blue?

   a. 1/6 b. 3/14 c. 1/4 d. 2/7 e. 1/3

9. What is the least integer value of n such that $(1/2^n)$ is less than 0.01?

   a. 7 b. 11 c. 50 d. 51 e. There is no such least value

10. A distillate flows into an empty 64-gallon drum at spout A and out of the drum at spout B. If the rate of flow through A is 2 gallons per hour, how many gallons per hour must flow out at spout B so that the drum is full in exactly 96 hours?

    a. 3/8 b. 1/2 c. 2/3 d. 4/3 e. 8/3

11. If $a > 0$, $b > 0$, and $c > 0$ then $a + (1/(b + 1/c)) =$

    a. $(a + b)/c$ b. $(ac + bc + 1)/c$ c. $(abc + b + c)/bc$ d. $(a + b + c)/(abc + 1)$ e. $(abc + a + c)/(bc + 1)$

12. If $L = (a-b) - c$ and $R = a- (b-c)$ then $L - R =$

    a. 2b b. 2c c. 0 d. -2b e. -2c

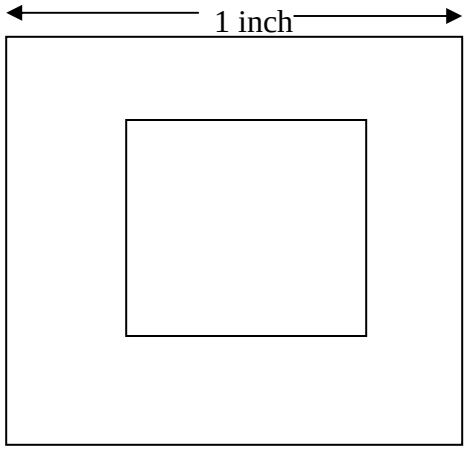
13. For each of *n* people Maggie bought a hamburger and a soda at a restaurant. For each of *n* people Paul bought 3 hamburgers and a soda at the same restaurant. If Maggie spent a total of $5.40 and Paul spent a total of $12.60 how much did Paul spend for just hamburgers? (Assume that all hamburgers cost the same and all sodas cost the same).

a. $10.80 b. $9.60 c. $7.20 d. $3.60 e. $2.40

14. The buyer of a certain mechanical toy must choose 2 of 4 optional motions and 4 of 5 optional accessories. How many different combinations of motions and accessories are available to the buyer?

a. 8 b. 11 c. 15 d. 20 e. 30

15. In the figure below, if the area of the smaller square region is ½ the area of the larger square region, then the diagonal of the larger square is how many inches longer than the diagonal of the smaller square?



a. sqrt2 -1 b. ½ c. sqrt2/2 d. (sqrt2 +1)/2 e. sqrt2

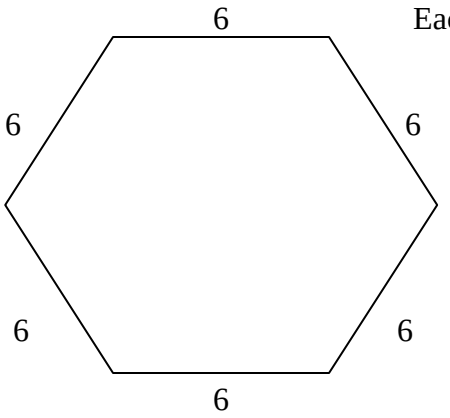
16. A certain recipe makes enough butter for exactly 8 circular pancakes that are 10 inches in diameter. How many circular pancakes, each 5 inches in diameter and of the same thickness as the 10-inch pancakes, should the recipe make?

a. 4 b. 16 c. 24 d. 32 e. 40

17. What is the area of the hexagonal region shown in the figure below?

a. 54 (sqrt 3) b. 108 c. 108 (sqrt3) d. 216

e. It cannot be determined from the information given



Each interior angle is x^o

18. A farmer has two rectangular fields. The larger field has twice the length and 4 times the width of the smaller field. If the smaller field has area K, then the area of the larger field is greater than the area of the smaller field by what amount?

a.. 2K b. 6K c. 7K d. 8K e. 12K

19. A board of length L is cut into two pieces such that the length of one piece is 1 foot more than twice the length of the other piece. Which of the following is the length, in feet, of the longer piece?

a. (L+2)/2 b. 2L + ½ c. L-1/3 d. 2L + 2/3 e. (2L + 1)/3

20. How many positive integers are both multiples of 4 and divisors of 64?

a. two b. three c. four d. five e. six

D6 Demographics Questionnaire

Investigating Why Some Mathematics Tests Show Differences in Scores Based on Sex

Demographics Questionarre

Study ID (on sheet provideed) _____

Sex _____

Age (circle the appropriate range) a) under 18 b) 18 to 22 c) 22 to 30 d) 30 to 40 e) 40 to 50 f) over 50

Race a) Black b) Asian c) Native American d) White e) other _____

Major Field of Study _____

Number of Immediate Family Members (parents, step-parents, siblings) With a College Degree _____

High School Graduated From (name, city and state) _____
_____

Last Math Class Taken in High School _____

Appendix E – Analysis of Variance Results for Males

As expected, due to the very different experiences males and females have with negative stigmatization in mathematics the manner in which threat state and mathematical domain identification affected motivation differed between males and females. As stated in the main body of this work I do not believe comparisons between the two sexes yield much information about the effect of negative stereotypes on females since males do not have the same experience with negative stigma associated with math as do females. However, the following is a brief overview of the results for males (who fully participated in all aspects of the study).

Threat State

The only motivation measure where threat state produced a significant difference in score was for PAP+. For this scale males under the gender neutral condition scored higher than the males in the gender different ($F(1,110) = 5.281$ with $p < .024$). So the males who were told that females did as well on the test as males did were more eager to approach the test with the intent of showing that they could do better than others.

In contrast it will be recalled that the only motivation measure where differences in threat state produced a significant result for females was the mastery measure. In that case, the gender different females reported higher levels of mastery motivation than the gender neutral females.

Although it is, in my opinion, dangerous to compare the results between genders (as stated above) it may be noteworthy that for threat state the male group that

reported higher motivation (PAP+) actually scored highest on the math test while the female group that reported higher motivation (M) scored the lowest on the math test. For both sexes it was the group that would be considered "threatened" by the manipulation (i.e. gender different for females and gender neutral for males) that reported higher levels of some sort of motivation, although the type varied.

Table E1

Descriptive Statistics for Males Only for PAP+

Dependent Variable: PAP+

THREATSTATE	mathidquartile	Mean	Std. Deviation	N
Gender Neutral	1.00	13.4545	3.93354	11
	2.00	11.8750	3.53789	16
	3.00	15.0000	3.07060	15
	4.00	16.6000	3.75690	15
	Total	14.2456	3.92010	57
Gender Dfferent	1.00	11.3636	4.05642	11
	2.00	11.7059	4.14977	17
	3.00	14.2308	2.48843	13
	4.00	13.0000	4.63191	12
	Total	12.5472	3.96410	53
Total	1.00	12.4091	4.04333	22
	2.00	11.7879	3.80590	33
	3.00	14.6429	2.79171	28
	4.00	15.0000	4.47214	27
	Total	13.4273	4.01479	110

Mathematical Domain Identification

Math identification differences resulted in significant differences for three of the four motivation types measured (M, PAV and PAP+).

For M, F (3,110) = 8.351; p< .000. The pattern for males is the same as that for females, namely the higher the math identification quartile the stronger the reported mastery goal score. This seems reasonable and is almost an internal check of the instruments.

Table E2

Descriptive Statistics for Males Only for Mastery

Dependent Variable: mastery

THREATSTATE	mathidquartile	Mean	Std. Deviation	N
Reduced	1.00	12.2727	3.10132	11
	2.00	13.1250	2.80179	16
	3.00	15.2000	2.65115	15
	4.00	15.9333	4.58984	15
	Total	14.2456	3.60199	57
Enhanced	1.00	10.6364	3.41388	11
	2.00	13.8235	4.00367	17
	3.00	15.2308	3.72276	13
	4.00	16.0833	2.50303	12
	Total	14.0189	3.93451	53
Total	1.00	11.4545	3.29107	22
	2.00	13.4848	3.43803	33
	3.00	15.2143	3.13117	28
	4.00	16.0000	3.74166	27
	Total	14.1364	3.75017	110

For PAV (F (3,110) = 2.871; p < .04) the significant result seems to be due to a large jump between quartiles 2 and 3 for the gender different threat state. See Table E3 below. This seems anomalous. There was no significant result for females on this scale.

Table E3

Descriptive Statistics for Males only for PAV

Dependent Variable: PAV

THREATSTATE	mathidquartile	Mean	Std. Deviation	N
Reduced	1.00	13.82	2.401	11
	2.00	11.94	3.872	16
	3.00	13.93	2.939	15
	4.00	12.93	4.803	15
	Total	13.09	3.695	57
Enhanced	1.00	10.64	4.884	11
	2.00	10.35	4.046	17
	3.00	14.00	3.055	13
	4.00	12.58	3.397	12
	Total	11.81	4.067	53
Total	1.00	12.23	4.093	22
	2.00	11.12	3.982	33
	3.00	13.96	2.937	28
	4.00	12.78	4.163	27
	Total	12.47	3.914	110

For PAP+ ($F(3,110) = 4.788$; $p < .004$) the gender neutral males showed more variability (increase between quartiles) and higher scores overall (see Table E1). This is like the gender different condition females' reaction (which makes sense since reduced threat for females is enhanced threat for males) except that the "threatened" males did well on the math test and the "threatened" females did not. For males motivation is motivation, for females it (motivation) is cognitive load? Perhaps the different outcomes are based on the novelty of the thing (for males), versus the reinforcement of old ideas for the females. In fact, it is the "novel" group that does best on the math test for both genders (the female reduced threat group) in this sample.