IMPLICIT MEASUREMENT OF CHALLENGE AND THREAT AS MOTIVATIONAL RESPONSES TO STEREOTYPE THREAT

by

JAMIE O'MALLY

JAMIE DECOSTER, COMMITTEE CHAIR
SHEILA BLACK, CO-CHAIR
MARK DAVIS
DAVID R. EWOLDSEN
JAMES C. HAMILTON

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2011

ABSTRACT

Stereotype threat occurs when people identify with a stigmatized group and experience depressed performance on a task because they become anxious in their desire to disprove a negative stereotype (Steele, 1997). Recent research using a Biopsychosocial (BPS) model has shown that physiological responses to challenge and threat may be an important aspect to consider in understanding the underlying motivational states that influence performance (Vick, Seery, Blascovich, & Weisbuch, 2008). However, physiological data are expensive and time-consuming to collect. The primary goal of this dissertation was to determine whether an implicit cognitive measure could be used in place of physiological equipment to effectively examine motivational responses to challenge and threat. In Study 1, four modified implicit measures were explored to determine which measure would be most effective in examining underlying motivations for challenge and threat. The modified Stroop task demonstrated patterns consistent with stereotype threat effects in Study 1 and was selected for use in Study 2. In Study 2, the relation between performance on the Stroop task and physiological measures of challenge and threat was examined. Additional measures of interest in this study included: working memory, state anxiety, math abilities, and perceptions of task performance. Results indicated that there were no significant effects of stereotype threat conditions on performance on the Stroop or physiological measures. However, stereotype threat significantly influenced state anxiety and perceptions of performance.

.

LIST OF ABBREVIATIONS AND SYMBOLS

ACT American College Testing: Standardized college entrance exam

bpm Measurement of heart rate in beats per minute

BPS Biopsychosocial model: Model used to explain physiological responses to challenge and threat that motivate performance

CO Cardiac Output: Amount of blood in liters that is ejected by the left ventricle of the heart each minute

d Cohen's d: An effect size used to indicate the strength of a relationship between two variables

F Fisher's *F* ratio: A ratio of two variances

GLM General Linear Model

GPA Grade Point Average

GRE Graduate Record Examination: Standardized graduate school entrance exam

HR Heart Rate: Number of heart beats occurring per minute

IAT Implicit Associations Test: Computerized cognitive measure used to assess the strength of associations between items based on reaction time responses

ICG Impedance Cardiography: Technique used to examine cardiovascular function

LVET Left Ventricular Ejection Time: Amount of blood leaving the left ventricle of the heart

M Mean: The sum of a set of measurements divided by the number of measurements in the set

N	Sample size
p	Probability associated with the occurrence under the null hypothesis of a value as extreme as or more extreme than the observed value
r	Pearson product-moment correlation
SAT	Scholastic Aptitude Test: Standardized college entrance exam
SD	Standard deviation: The square root of the variance
STAI	State Trait Anxiety Inventory scale: Self report scale of state anxiety
SV	Stroke Volume: Amount of blood pumped by the heart for a given heartbeat
t	Computed value of t test
TPR	Total Peripheral Resistance: Amount of blood passing through a blood vessel at any moment that is determined by blood pressure and resistance to flow
VC	Ventricular Contractility: Measurement of the amount of blood entering and leaving a ventricle in the heart
$<$	Less than
$=$	Equal to

ACKNOWLEDGMENTS

I am pleased to express my sincere gratitude to the faculty, friends, and colleagues who have made this dissertation possible. I truly appreciate the endless support, patience, guidance, and statistical knowledge provided by my advisor, Jamie DeCoster. I would also like to thank my committee members Sheila Black, Mark Davis, Dave Ewoldsen, and Jim Hamilton for assisting in the development of this project and for their support of my research endeavors. I am truly indebted to my colleagues at the Institute for Social Science Research at the University of Alabama, especially Joan Barth and Debra McCallum whose support and guidance made this project possible. I appreciate the help of various consultants throughout the research process including Brooke Vick and Phil Cusimano. In addition, I am thankful for the friendship and work of an amazing team of enthusiastic undergraduate research assistants who worked vigorously to collect the data for this dissertation: Melissa Bruce, Lauren Griffin, Melissa Hines, JJ Molloy, FabryeAnne Robinson, and Dustin Swartz. I am immeasurably grateful for the continual support of my closest friends and family. I would especially like to thank my fiancé, Mark Martinez, who tenderly and heroically dealt with the brunt of my tears, frustration, and stress by continually reminding me that I could really do this. Finally, this research would not have been possible without the gracious financial support provided by the Institute for Social Science Research and the Psychology Department at the University of Alabama.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Implicit Measurement of Threat and Challenge as Motivational Responses to Stereotype Threat

Initial research on stereotyping focused on the cognitive processes involved in making judgments and decisions as well as the social contexts in which stereotyping is most likely to occur (e.g. Chen, Duckworth & Chaiken, 2000; Sinclair & Kunda, 1999). More recently, researchers have explored the experiences of those who are targets of stereotyping (Aronson et al., 1999; Croizet & Claire, 1998; ; Inzlicht & Ben-Zeev, 2000; Keller & Dauenheimer, 2003; Spencer, Steele & Quinn, 1999; Steele & Aronson, 1995; Vick, Seery, Blascovich, & Weisbuch, 2008). People identified with a stigmatized group may become so anxious in their desire to disprove a negative stereotype that they hinder their own performance, a phenomenon known as stereotype threat (Steele, 1997). Stereotype threat has been proposed as one explanation for the achievement gap between minorities and Whites, in that minorities may experience anxiety when taking tests because of their fears of confirming a negative stereotype against their group (Steele, 1997). In addition to examining depressed performance in the face of a negative stereotype, previous research has also shown that participants can exhibit motivational processes that reflect a state of challenge rather than threat (Chalabaev, Major, Cury, & Sarrazin, 2009; Scheepers, 2009; Vick et al., 2008).

In most research regarding the effects of threat, the target's experience is typically measured explicitly using self-report measures or measures of task performance. Recently researchers have examined underlying motivational processes that occur during situations of stereotype threat. Previous studies have demonstrated that physiological responses to threat may be an important aspect to consider in understanding the underlying motivational states that influence performance under stereotype threat (Vick et al., 2008). According to the Biopsychosocial (BPS) model, when people face a situation in which they feel they have the resources either necessary or exceeding those needed to succeed at the task, they will experience a sense of challenge resulting in improved performance. On the other hand, when people feel that a task demands resources beyond their abilities, they experience a sense of threat resulting in decreased performance (Vick et al., 2008). Combining physiological measures with standard social psychology methodologies has come to the forefront of recent research. However, these methods are extremely costly to employ in terms of training, equipment, and administration time. The main goal of this dissertation was to examine whether researchers can utilize implicit cognitive measures in place of physiological equipment to examine the impact of stereotype threat. By expanding methods of examining the effects of challenge and threat, we can assess underlying motivational states that cannot be understood using explicit measures alone in situations where cumbersome physiological equipment cannot be easily used. This will allow us to move towards identifying new ways to reduce the negative effects of stereotype threat.

Factors Influencing Responses to Stereotype Threat

Although most previous research has focused predominantly on depressed performance under stereotype threat, a few studies have been conducted to examine when a challenge response is likely to be invoked (Alter, Aronson, Darley, Rodriguez, & Ruble, 2010; Keller, 2007; Scheepers, 2009; Vick et al., 2008). Threat responses are considered maladaptive and likely to lead to impaired performance, while challenge responses are considered positively adaptive and likely to lead to facilitated performance (Scheepers, 2009). Threat and challenge motivational processes can occur based on a variety of situational contexts. For example, motivational processes for high and low status groups can change based on the perceived stability of status differences. In research in which status stability was manipulated, low status groups in an unstable status hierarchy condition not only displayed a challenge response physiologically but also outperformed the high status group on an actual performance task (Scheepers, 2009). Challenge responses to threatening situations have been shown to occur for low status groups when status hierarchy is unstable, making the possibility of status improvement seem more possible. In conditions in which status hierarchy was unstable, high status group members experienced motivational processes that reflect threat because their superior position was less secure (Scheepers, 2009).

Common paradigms in stereotype threat research use a manipulation in which instructions before a test either emphasize or reduce a negative stereotype. Several factors have been suggested to play a role in how vulnerable a target will be to the negative effects of stereotype threat. The main influences that have been examined throughout stereotype threat literature are group identification, stigma awareness, domain identification, diagnosticity in testing contexts, and task difficulty (Wheeler & Petty, 2001). Each of these factors is described below to provide a context for understanding how stereotype threat was manipulated in the current study.

Group identification. In order to experience stereotype threat, individuals must first identify themselves as belonging to a stigmatized group. Those who highly identify with a particular stigmatized group and feel that their group identification is an important part of their self-concept are at a higher risk of experiencing the negative effects of stereotype threat than people who are not highly identified with a stigmatized group. Although stereotype threat was initially proposed to explain racial differences in performance (Steele & Aronson, 1995), stereotype threat effects have also been found for other groups, including those defined by gender, age, and socioeconomic status (Croizet & Claire, 1998, Spencer et al., 1999; Hess, Auman, Colcombe, & Rahhal, 2003). Women who are highly identified with their gender have been shown to perform worse than women who are less identified with their gender when performing a gender-stereotyped task (Schmader, 2002).

Stigma awareness. Targets must be aware of the negative stereotype against their group to be subject to stereotype challenge or threat. Most studies ensure salience of the negative stereotype by manipulating instructions prior to the start of a task (e.g. Spencer et al., 1999). However, some negative stereotypes are so prevalent that they become automatically salient when targets are asked to simply report their demographic information prior to taking a test (Croizet & Claire, 1998; Steele & Aronson, 1995). Research has also shown that for participants who are members of both a negatively stereotyped group and a positively stereotyped group, performance is affected by the group identity that is more salient to them at the time. For example, there is a common stereotype that Asians excel at math, while it is also commonly accepted that women are not as good as men at math. How do Asian women respond when taking math tests? Research has shown that when gender identity is activated, performance can be depressed, while activating the ethnic identity can actually improve performance (Shih, Pittinsky, & Ambady, 1999).

Domain identification. When people identify themselves as members of stereotyped groups, their vulnerability to the negative effects of stereotype threat is increased when the task is important to them. Targets that identify highly with a particular domain are more likely to underperform due to stereotype threat effects than participants who are not highly identified with the domain (Aronson et al., 1999). This is intuitive in that those with high domain identification would find it more important to be successful in that particular task than those for whom the domain is less important. With greater importance placed on task success, anxiety concerning ability to overcome the salient stereotype increases. For example, women who are pursuing professions in which excelling in math is critical will be more likely to experience negative effects of threat than women who do not view math as an integral part of their self-concept (Spencer et al., 1999).

Diagnosticity in testing contexts. Not only is it important that individuals identify with the task domain, it is also important that they believe that the task is capable of measuring their abilities in the domain for them to be subject to stereotype challenge and threat. Previous research has shown that the performance of participants who are told that a test is diagnostic of their abilities suffers greater than those in a condition in which the test is portrayed as non-diagnostic. Frustration and anxiety increase for tasks described as diagnostic because the pressure to succeed is greater in circumstances in which participants believe that their performance is directly reflective of their abilities (Croizet & Claire, 1998; Steele & Aronson, 1995). Recent research has further demonstrated that negative effects of stereotype threat can be reduced by reframing a task as challenging rather than threatening. In one study, participants were told that a problem solving task was being used either as a measure of ability or as a tool to help students sharpen their problem solving strategies. Participants in the challenge-framing condition performed better than those in the threat-framing condition. The authors noted that this preliminary work supports the notion that challenge-framing decreases poor performance among stigmatized groups (Alter et al., 2010). Another study showed that when participants were encouraged to view the testing context with a performance-based focus (i.e. assessing strengths), their physiological patterns and performance showed increased performance and physiological markers of challenge; compared to participants who were encouraged to view a testing context with a prevention-based focus (i.e. assessing weaknesses) who showed decreased performance and physiological markers of threat (Chalabaev et al., 2009).

Task difficulty. Because a major component of stereotype threat is based on anxiety experienced during the task, the test should be difficult enough to elicit frustration, but not so difficult that the target would be incapable of succeeding in non-threatening conditions. Spencer and colleagues (1999) administered two sets of tests, one easy and the other difficult. When gender-bias was made salient, women experienced stereotype threat effects on the difficult math test but performed equally as well as men on the easier math test, suggesting that task difficulty is an important factor in vulnerability to stereotype threat (Spencer et al., 1999).

A brief review of various methodologies will shed light on how each of these criteria is incorporated in studying the effects of stereotype threat. In one of the first stereotype threat studies (Steele & Aronson, 1995), Black and White college students were given items from the verbal GRE (Graduate Record Examination). The test was described as either diagnostic of cognitive abilities or as a problem-solving task that was not diagnostic of intellectual abilities. When the test was described as diagnostic it made a negative racial stereotype of the intellectual abilities of Blacks more salient, and caused such anxiety that the performance of Blacks was actually hindered (Steele & Aronson, 1995). Results also showed that Blacks reported exerting the same amount of effort on the exam regardless of condition. In addition, prior SAT scores were used to control for natural, existing abilities. In summary, when threat is removed by describing the test as exploratory rather than diagnostic of ability, Blacks performed as well as Whites after controlling for existing ability. These results support the notion that the achievement gap between Blacks and Whites can be attributed to situational contexts in which tests are taken. Several later studies replicated the effects of stereotype threat on the performance of Blacks taking aptitude tests (Aronson, Fried, & Good, 2002; Blascovich, Spencer, Quinn & Steele, 2001). In addition, stereotype threat effects have been demonstrated among other ethnicities like Latinos (Gonzales, Blanton, & Williams, 2002).

Similar to the effects of racial stereotypes, researchers found that women experienced stereotype threat when gender was made salient for tasks that measured mathematical abilities (Brown & Josephs, 1999; Oswald & Harvey, 2000; Schmader, 2002; Spencer et al., 1999). In an initial study of gender stereotype threat, women with superior math skills were randomly assigned to one of two groups before taking a difficult math test (Spencer et al., 1999). Before beginning the test, one group was told that the test was biased against women, whereas the other group was told the test was fair. Women who were led to believe that the test was fair performed as well as men, however those who believed the test was biased performed worse than men. In addition, the results also showed that women's expectations of their performance and the effort they made were not influenced by condition. This supports the theory that depressed performance was a result of stereotype threat effects rather than decreased effort or self-efficacy (Spencer et al., 1999).

Performance and Self-Report Measures of Challenge and Threat

The present research examines ways in which stereotype threat can be measured and interpreted. By improving methods of examining the effects of threat, we can move towards identifying new ways to reduce threat. Researchers have typically relied on task performance and self-report to gain an understanding of the effects of stereotype threat. The most common methodologies typically measure performance on a test by examining the number of questions correct, the number of questions attempted, and the amount of time spent on each question

(Croizet & Claire, 1998; Spencer et al., 1999; Steele & Aronson, 1995). Self-report measures include participant ratings of their anxiety levels, self-efficacy, and experience of threat (Steele & Aronson, 1995; Stone, Lynch, Sjomeling, & Darley, 1999; von Hippel et al. 2005). For example, women who were asked to rate their anxiety before taking a test reported feeling more anxious when they were told the test was gender-biased than women who were told the test was gender-fair (Spencer et al., 1999). Similarly, Blacks who believed a test was diagnostic of academic ability reported experiencing greater anxiety than Blacks in the non-diagnostic testing condition (Osborne, 2001).

While performance measures and self-report are currently the most widely used method of examining the effects of stereotype threat, they may not be the most accurate. When participants are asked to report their feelings of anxiety, frustration, and self-efficacy using explicit measures, there is a possibility that participants may not be aware of the level of emotions they are truly experiencing or that they may not be willing to respond honestly. In addition, by asking participants to self-report their experiences, we force them to actively reflect on their experiences, which could be disrupting the emotional processes involved in stereotype threat (Vick et al., 2008). Performance measures and self-reports can only give us part of the picture. There are motivational processes that lead to internal responses to challenge and threat that cannot be identified sufficiently using explicit measures.

Physiological Measurement in Psychological Research

For some psychological states, people are unable to accurately introspect their true experiences. In addition, the explicit consideration of one's psychological state can sometimes alter that state. Therefore, it is often necessary to supplement performance measures and self-reports with direct physiological measurement.

Various methodologies have been used to examine physiological responses that underlie psychological phenomena. For example, galvanic skin response, which measures electrical skin conductivity and moisture levels, has been used to explore physiological and psychological arousal in responses to advertising stimuli (Ohme, Reykowska, Wiener, & Choromanska, 2009), situational stress (Al-Fudail & Mellar, 2008), and emotion processing (Ramachandran, Depalma & Lisiewski, 2009). Blood pressure has frequently been used in psychological research to examine psychological stress and pain sensitivity (Frew & Drummond, 2009), stereotype threat and anxiety (Osborne, 2007), and anger and forgiveness (Lawler-Row, Karremans, Scott, Edlis-Matityahou, & Edwards, 2008).

Impedance cardiography (ICG) is a noninvasive method that is used to explore cardiac function. ICG measures thoracic impedance using four electrodes applied to the neck and lower back. A low voltage, high frequency electrical current constantly flows between the electrodes and fluctuations in this flow indicate changes in blood volume that are occurring during the cardiac cycle (Berntson, Quigley, & Lozano, 2007). These changes can be identified in calculations of stroke volume and cardiac output which are derived from heart rate and left ventricular ejection time (LVET) (Kubicek, Karnegix, Patterson, Witsoe, & Mattson, 1966). Cardiac output, which measures the amount of blood pumped by the heart over a given period of time, has been used in psychological research to examine abnormal brain aging (Jefferson, 2010), stress, (Williams, O'Carroll, & O'Connor, 2009) fear and sadness (Kriebig, Wilhelm, Roth, & Gross, 2007), and motivational responses to challenge and threat (Blascovich, Seery, Mugridge, Norris, & Weisbuch, 2004). Cardiac output is calculated by multiplying heart rate by stroke volume. Stroke volume represents the amount of blood pumped from the left ventricle with each heartbeat. Cardiac function is influenced by a variety of factors including changes in activity in the sympathetic nervous system and the parasympathetic nervous system. Increases in cardiac output are marked by an increase in blood flow to skeletal muscles, via dilated arteries, allowing oxygenated blood to move to the periphery more quickly (Blascovich & Mendes, 2000). In this dissertation, cardiac output is measured to explore physiological responses to challenge and threat.

Physiological Measurement of Challenge and Threat

Previous research has used continuous on-line measurement of cardiovascular output, blood pressure, and heart rate variability to examine physiological responses to stereotype threat (Blascovich et al., 2001; Croizet et al., 2004; Vick et al., 2008;). Cardiovascular measures can show differences in motivational processes in which participants either respond to stereotype threat with a physiological response that is representative of a challenged motivational state or a physiological response that is representative of a threatened motivational state (Vick et al., 2008). The bipolar dimensions of threat and challenge are typically interpreted in terms of relativity. The expectation is that those experiencing stronger stereotype threat, (i.e. they believe that the situational demands exceed their resources and abilities) will exhibit physiological responses more indicative of a threatened motivational state, while those for whom stereotype threat is alleviated (i.e. they believe that their abilities and resources meet or exceed situational demands) will exhibit physiological responses indicative of a challenged motivational state.

Research examining physiological interactions with environmental pressures also supports the notion that there may be physiological differences in how people respond to threatening situations. Using testosterone levels as an indicator of internal responses, researchers found that higher testosterone levels influence behavior in situations involving stereotype threat and the challenge response may occur but only for individuals of high status, concluding that testosterone levels in combination with a social situation can influence performance on a cognitive task, such that performance can increase when a challenge response is invoked (Newman, Sellers, & Josephs, 2005). This notion was applied this notion to stereotype threat, finding that higher levels of testosterone led to a more reactive response for participants in high status positions compared to low status positions, and that those in high status positions were better able to use their cognitive resources to complete a task (Newman et al., 2005).

Four cardiovascular measures have been validated as indices of threat and challenge motivational states (Blascovich & Tomaka, 1996). These include heart rate (HR), ventricular contractility (VC), cardiac output (CO), and total peripheral resistance (TPR) (Blascovich, et al., 2004; Townsend, Major, Sawyer, & Mendes, 2010; Scheepers, 2009, Vick et al., 2008). TPR is a measure of the amount of blood passing through a blood vessel at any given moment. This is determined by blood pressure and resistance to blood flow which is largely a result of changes in vessel diameter or structure (Berntson, Quigley, & Lozano, 2007). During motivational states of challenge, TPR decreases as a result of increased HR and VC, leading CO to increase. Although HR and VC also increase during a motivational state of threat, TPR actually increases and little or no change occurs in CO, as a result of the inhibition of epinephrine in response to threat (Blascovich & Tomaka, 1996; Townsend, et al., 2010; Vick et al., 2008). In this dissertation, we used HR, CO, and SV to examine physiological changes in response to challenge and threat. Measurement of TPR was excluded from the current study due to financial limitations. TPR can be used in calculating CO by dividing mean arterial pressure by TPR (Berntson, Quigley, & Lozano, 2007). Differences in challenge and threat responses can be interpreted from relative changes in cardiac output alone because TPR is a function of the same underlying activations (Blascovich, Mendez, Tomaka, Salomon, & Seery, 2003; Vick et al., 2008). The ability to examine physiological responses to stereotype threat is invaluable because with an understanding of the processes that lead to depressed performance under threat, we can begin to explore new ways to alleviate negative effects of stereotype threat as well as ways in which we can work towards increasing motivational states that reflect a challenge response.

Challenge responses to threatening situations occur when there is a physical arousal response to threat that involves increased heart rate, but a decrease in blood pressure, allowing one to maintain focus on the task at hand (Blascovich & Tomaka, 1996). Men and women can show different motivational responses to the same threat based on how the threat applies to the group to which they belong. Vick and colleagues (2008) showed that men and women responded differently to the presence of a stereotype threat for gender. Physiological measures demonstrated that women experienced a threatened motivational state during a math test in which they were told that men typically outperformed women and a challenged state when told the test was gender-fair. Specifically, the physiological measures indicated an increase in cardiac output for men who were given the same instructions regarding a negative bias against women, marking a challenge response. The reverse was true when participants were told the test was gender-fair, such that women exhibited a challenge response and increased cardiac output, while men experienced threat and little or no change in cardiac output (Vick et al., 2008).

There are many benefits to using physiological measures to examine the motivational states to which performance under threat can be attributed. By using continuous measurement, physiological techniques not only alleviate the need to rely on participants' compliance and capability to accurately report their experiences, but it also reduces any interference that could occur when participants are asked to actively reflect on their experiences (Blascovich & Seery, 2007).

It is clear that the use of physiological measures of challenge and threat motivational states has shown promise in examining mechanisms of stereotype threat. The benefit of using measures of internal responses has wide support not only in stereotype threat literature but in other domains of social psychology as well. Unfortunately the expenses involved in collecting physiological data currently restrict many researchers from examining stereotype threat in this way. The equipment needed to conduct a study like those conducted by Vick and colleagues (2008) requires equipment costing upwards of $15,000. It also requires disposable materials such as sensors that cost approximately $12.00 for each participant. Researchers should receive formal training on how to use equipment to collect data which typically costs about $3000 for a group training session. Not only is the use of physiological measures financially taxing, it is also time consuming and methodologically restrictive. Most laboratories will only have a single set of physiological equipment, so participants must be run individually. This not only makes conducting physiological studies time consuming, but also limits our ability to manipulate stereotype threat in larger group settings.

Implicit Measurement of Challenge and Threat

The purpose of the present research is to expand measurement of challenge and threat beyond self-report, performance, and physiological measures. We developed 4 implicit cognitive measures to determine which of these may be used to examine motivational states that occur under threat and challenge. By examining the potential for computerized implicit tasks to identify underlying motivational responses, we hoped to develop a more efficient and accessible technique to identify implicit responses to challenge and threat. There are many existing implicit cognitive paradigms that have been used throughout psychological research but none have been used to examine implicit responses to threat and challenge. We modified 4 common implicit measures in an effort to determine whether a new technique could be developed and integrated with self-report, performance, and physiological measures of challenge and threat effects. We anticipated that this would lead to a more comprehensive understanding of the motivational response that occurs under conditions of challenge and threat. Additionally, we hoped the development of these techniques would increase efficiency of measurement and broaden the accessibility to study this phenomenon, leading to new avenues of stereotype threat research.

We conducted two studies to identify and test the ability of a cognitive implicit task to measure challenge and threat. The first study used four widely-used implicit measures to examine responses to a stereotype threat manipulation. We modified each of these existing measures to use them to detect underlying mechanisms of challenge and threat. The second study selected the implicit measure that was most effective in Study 1 and investigated its relation with a physiological measure of challenge and threat. Our main goal was to find an implicit measure that is comparable to physiological equipment in its ability to examine underlying processes of challenge and threat.

Each of the four implicit measures that we used in Study 1 uses a different method to assess underlying attitudes and beliefs that people may be either unwilling or unable to explicitly express. These implicit measures were modified versions of the Attitude Accessibility Task (Shen, Monahan, Rhodes, & Roskos-Ewoldsen, 2009), Implicit Associations Task (IAT) (Greenwald, McGhee, & Schwartz, 1998), the Stroop Color-Word Interference Task (McNally, Hornig, Hoffman, & Han, 1999), and the Word-Completion Task (McCabe, 1999). Each of the four implicit tasks was designed to measure the salience of challenge and threat for participants who are provided with information that either confirms or disconfirms a common gender stereotype.

We predicted outcomes similar to those in previous stereotype threat literature. For math performance, we predicted that women in a biased condition would experience depressed performance compared to women in a fair condition. Conversely, men in a fair condition would experience depressed performance compared to men in a biased condition. For each of the implicit tasks, we predicted that when a task was described as gender-biased, women would demonstrate stronger activation for threat words, while men would demonstrate a stronger activation for challenge words. When a task was described as gender-fair, we predicted that women would demonstrate a stronger activation for challenge words, while men would demonstrate a stronger activation for threat words.

STUDY 1

Method

Participants

College students (N= 236; 167 females, 69males) enrolled in Introductory to Psychology from the Psychology Subject Pool at the University of Alabama participated in this study. Data were excluded for 17 participants who guessed the hypothesis and 8 participants for whom equipment malfunctioned. Participants (N = 211; 149 females, 62 males) consisted of 77.3% White students, 20.4% Black students and 2.3% who reported another race or ethnicity. Participants had a mean age of 19.3, a mean ACT score of 25.33, and a mean GPA of 3.38. Participants earned 1.5 research credits in exchange for their participation.

Design

A 2 (gender) x 2 (stereotype threat: gender-fair test, gender-biased test) between-subjects design was used in this study. The dependent variables included number of questions answered correctly on a math test and performance on one of four implicit measures (Attitude Accessibility Task, IAT, Stroop Task, and Word Completion Task).

Materials

Math Performance Test. The materials for this measure are presented in Appendix A. The math performance test consisted of 30 problem-solving items taken directly from an ACT study guide (Dulan, 2007). Each multiple-choice question was presented individually on the computer screen. Participants were given 20 seconds to read and answer each question, with a countdown timer appearing below each question. Participants were asked to perform their calculations mentally, without paper or a calculator, and to respond by pressing a key on the keyboard. The test lasted approximately 10 minutes.

Implicit Measures

Participants were randomly assigned to complete one of 4 implicit computerized tasks. Each task took approximately 10 minutes to complete and was administered on a desktop computer using E-prime experimental presentation software. Details for each task are described below.

All implicit tasks used the same challenge and threat word lists. In an online pilot study, participants (N = 67) rated a total of 173 words on whether each word related more to challenge or threat. Participants were given the following instructions:

> "On the next 4 pages, you will be asked to rate words based on whether they relate to challenge or threat. As you are rating the words, please keep in mind the following definitions of challenge and threat. **Challenge**- People experience a feeling of challenge when they are faced with a situation or task that they believe they have the resources or skills necessary to equal or exceed the demands of the situation or task. **Threat**- People experience a feeling of threat when they are faced with a situation or task that they believe outweighs the availability of resources or skills necessary to cope with demands of the situation or task."

Participants then rated each word on a 7-point scale on which the lower end of the scale was labeled "challenge" and the higher end of the scale was labeled "threat." From these ratings, a preliminary word list was created that consisted of 50 words with the lowest mean ratings (challenge words) and 50 words with the highest mean ratings (threat words). From these lists, we then randomly generated 10 lists containing 25 challenge words and 10 lists containing 25 threat words. Means for word length and frequency of use were calculated for each of these lists. Frequency of use was determined using an interactive website that includes 86,800 of the most commonly used English words (Harris, 2003). We then chose the pair of lists that had the lowest mean difference for both word length and frequency of use. These lists are presented in Appendix B.

Attitude Accessibility. According to Fazio (1995), the faster the speed at which an individual evaluates an attitude or belief, the more salient or accessible the attitude is for that person. Attitude accessibility tasks have been used to examine the salience of objects and beliefs in a wide variety of topics like persuasive messages (Shen et al., 2009), religious beliefs (Cohen, Shariff, & Hill, 2008), and self-esteem (DeMarree, Petty, & Strunk, 2010). We based our version of this task on established attitude accessibility procedures (Roskos-Ewoldsen & Fazio, 1992; Shen et al., 2009). We asked participants to indicate whether they liked or disliked a number of different attitude objects. Instructions directed participants to consider how they are feeling right now and to categorize the words on the screen by pressing the "m" key if they liked the object and the "c" key if they disliked the object. The first three blocks consisted of 24 trials using practice words to familiarize the participants with the task. The fourth block consisted of the critical list of randomly ordered words containing 25 challenge, 25 threat, and 25 filler words. We predicted that threat words would be highly accessible for women in gender-biased conditions and that challenge words would be highly accessible for women in gender-fair conditions. We predicted the opposite would be true for men in gender-biased and gender-fair conditions.

Implicit Association Test (IAT). The IAT (Greenwald, McGhee, & Schwartz, 1998) is one of the most well-known and widely used measures in social cognition research. It has been used to examine implicit associations in a number of domains including implicit attitudes towards race, gender, and age (Nosek, Banaji, & Greenwald, 2002). Participants categorize words presented by a computer and reaction times are used to examine implicit associations. We are interested in whether there are stronger associations between math and threat words or between math and challenge words. This version of the IAT is administered in 6 blocks of randomly presented words. The task begins by introducing the keys and concepts to be used in the experiment.

Block 1 is a practice block to familiarize the participants with the keys and concepts of the task. In this block, participants categorize 25 trials of words as relating to either flowers or insects by pressing corresponding keys. In Block 2, participants are presented with 50 trials of words and are asked to press a left-hand key when the word is related to math and to press the right-hand key when the word is not related to math. After participants become familiar with this, the categorization labels reverse and the rules change for the following 4 blocks. In Block 3, participants categorize 50 trials of words by pressing the left-hand key whenever they see a word that they thought related to challenge and pressing the right-hand key whenever they see a word that related to threat. Block 4 includes 100 trials in which participants categorize all stimuli previously seen. Participants are asked to press the left-hand key whenever they see either a math word OR a challenge word, and to press the right-hand key whenever they see a non-math word OR a threat word. In Block 5, the categorizations are reversed so that participants respond to 50 trials by pressing a left-hand key to categorize threat words and a right-hand key to categorize challenge words. In Block 6, participants are presented with 100 trials and asked to press the left-hand key whenever they see either a math word OR threat word, and to press the right-hand key whenever they see a non-math word OR a challenge word. The order of word categories was counterbalanced across participants. Theoretically, the more difficult participants find the final Block, the more strongly they associate challenge words with math, indicating that they are experiencing a motivational state of challenge rather than threat.

Stroop Color-Word Interference Task. This task typically measures selective attention abilities and the amount of cognitive interference in naming a color that is incompatible with a word displayed (Stroop, 1935). This task has been used frequently throughout cognitive literature and experiments using implicit measures (for an extensive review see MacLeod, 1991). In the traditional procedure for this task, a series of names of colors are displayed one at a time. Participants are expected to name the ink colors of incongruent color names. Participants experience interference when they are unable to ignore the printed word in favor of the ink color (McNally et al., 1999). We modified this task by presenting a list of challenge and threat words in 4 colors. Participants were asked to determine the color of each of the words and to ignore the actual word. We predicted that when a word is more salient to a participant, they will respond more slowly to categorize the color of the word because they experience interference by the actual word. We predicted that participants who experienced more interference for threat words than for challenge words would identify the colors of threat words more slowly than challenge words.

Word Completion Task. The materials for this task are presented in Appendix C. This task has been used in studies to examine implicit memory (McCabe, 1999). Words become implicitly salient to someone because of priming, experience, situational context, or memory. We modified this task by providing participants with word puzzles that could be completed most commonly with letters that make up neutral words, but could also be completed with rarer words related to challenge or threat. For this task, participants viewed a series of incomplete words on the computer screen. Participants were shown 25 word puzzles that could be completed that are related to threat, 25 word stems to be completed that can be related to challenge, as well as 15 filler word puzzles. The word puzzles were missing an average of 3 letters. Participants were asked to complete the puzzles as quickly as possible with the first word that came to mind. All word puzzles were presented randomly for 10 minutes to avoid order effects. We predicted that for women in gender-biased conditions, threat-related words would be salient; while those in the gender-fair condition should more frequently call to mind challenge-related words. In contrast, we predicted for men in a gender-biased condition, challenge-related words would be most likely to come to mind, while threat-related words would be most salient for men in the gender-fair condition.

Additional measures:

Demographic questionnaire. This pencil and paper survey, presented in Appendix D, consisted of 8 items that asked participants to report personal information such as age, race, gender, major, standardized test scores, GPA, importance of math to the participant, and math courses taken. The demographics questionnaire took about 10 minutes to complete.

Short Form State-Trait Anxiety Inventory (SF STAI). This is a 6-item version of the Spielberger State-Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). This brief version assesses only state anxiety. The first STAI was completed at the start of the computerized portion of the study. Participants were presented with 6 computerized analogue mood scales. Participants were asked to indicate the extent to which they currently felt calm, tense, upset, relaxed, content, and worried. Each of the words was presented individually. Below each word was a scale consisting of a 15-cm horizontal line. The left side of the scale is labeled “not at all” and the right side of the scale is labeled “very.” To complete these scales, participants used the mouse to move their cursor along the line to a point corresponding to their current mood state and press the mouse button to register their response. This yielded a score between 1 and 30 for each item, depending on which area of the line was selected. Scores for calm, relaxed, and content were reversed. Higher scores indicated higher state anxiety. This measure took less than 5 minutes to complete and was administered at 3 different points in the study.

Math Performance Questionnaire. This 12-item questionnaire, presented in Appendix E, was administered on the computer and asked participants to rate their performance on the math test used in the study, their perceptions of test difficulty, and their agreement with gender stereotypes. Some questions were adapted from a survey used in prior research (McIntyre, Paulson, & Lord, 2003).

Working Memory Task. In this computerized working memory task (Schmader & Johns, 2003) that lasted about 15 minutes, participants were asked to count the number of vowels that occurred in individually presented sentences (Appendix F) while memorizing separate recall words (Appendix G) that flashed onto the screen after each sentence. Each sentence contained approximately 7-12 words and an average of approximately 10 vowels. Participants were told that counting vowels and recalling words were equally important. Participants were presented with 60 sentences over multiple blocks. Each block contained between 4-6 individual sentences that were randomly presented with a new recall word appearing after each sentence. Following each sentence presentation, participants were asked to report the number of vowels they counted in the sentence. After participants reported the number of vowels, a recall word flashed on the screen. At the end of each block, participants were asked to recall the list of 4-6 recall words that were presented in the previous block.

Post-Study Surveys. Two post-study surveys, presented in Appendix H, were administered using pencil and paper. These brief surveys contained a total of 6 items used to assess hypothesis-guessing, problems that may have arisen during the experiment, and a manipulation check. The post-study surveys took less than 10 minutes to complete.

Procedure

Participants were recruited through the psychology subject pool website to participate in the study in groups of up to four. Upon arrival to the lab, participants were given a copy of the study information sheet that they were invited to keep for their own records. Participants were told that the study would last approximately 1 hour. Participants were seated at a Dell desktop computer with a keyboard and mouse in an individual cubicle for the remainder of the study. Participants were first given the demographics questionnaire that they filled out and returned to the experimenter. The experimenter then prepared the computer to administer the remaining instructions and tasks for the study. Participants used the keyboard to submit their responses to each of the measures and the computer recorded all responses. An experimenter was present throughout the study. Participants were told to alert the experimenter if they were confused about instructions, had questions, or decided to discontinue their participation at any time during the experiment.

After completing the demographics questionnaire, participants were presented with instructions on the computer that outlined the general procedures for the study. At 3 different points in the study, participants completed the STAI, a brief measure of state anxiety. Following the first STAI, participants were given the following general instructions:

> "We are interested in examining the performance of college students on multiple measures of mathematical ability. In this study, you will first complete 2 tasks that have been used to identify people who are good at math. Then you will complete a task that measures your general cognitive ability. In order to assess your math ability, the following tasks require you to follow instructions, pay attention to stimuli, and solve problems as accurately and quickly as possible."

Participants were randomly assigned to receive either gender-fair or gender-biased instructions. Before beginning the implicit task, participants were provided instructions that led them to believe that the implicit task was math-related in order to maximize the fair or biased instructions that were provided later. The instructions created to justify the use of each implicit measure in determining mathematical ability are provided below.

"The first task you will complete does not contain math problems, however, it requires you to…

- make quick decisions to categorize concepts to solve a problem. (Attitude Accessibility)
- simultaneously coordinate multiple rules to solve a problem. (IAT)
- focus your attention only on information that is necessary to solve a problem (Stroop)
- intuitively determine the best approach to solve a problem (Word Completion)

…This task has frequently been used to accurately identify people who are good at math."

In order to manipulate stereotype threat, participants in the gender-fair condition were then led to believe that women perform equally as well as men on the tasks, while participants in the gender-biased condition were led to believe that men perform better than women on the tasks. For the gender-fair condition, the following instructions (based on McIntyre, Paulson, & Lord, 2003) were provided prior to beginning the implicit task:

> "Before you begin the following test, we should note that some people believe that women are inferior to men at math related tasks, but empirical evidence has shown that women and men perform equally on this test. Please try your best."

Participants in the gender-biased condition were given the same instructions as those above with the conclusion altered:

> "Before you begin the following tasks, we should note that some people believe that women are inferior to men at math related tasks, but empirical evidence is mixed. Please try your best."

Immediately following the manipulated instructions, participants were randomly assigned to complete one of the 4 possible implicit computerized tasks (Attitude Accessibility, IAT, Stroop, or Word Completion Task).

Following the implicit task, the participants' state anxiety was measured for a second time using the STAI. Participants were asked to respond to the STAI based on how they felt while they were completing the implicit task.

In order to re-emphasize the stereotype threat manipulation presented prior to the implicit task, a similar set of instructions occurred directly before participants began the math performance test. For the gender-fair condition, the following instructions were provided prior to completing the math test:

> "Before you begin the following math test, we should note that some people believe that women are inferior to men at math, but empirical evidence has shown that women and men perform equally on this test. Please try your best."

Participants in the gender-biased condition were given the same instructions as those above with the conclusion altered:

> "Before you begin the following math test we should note that some people believe that women are inferior to men at math, but empirical evidence is mixed. Please try your best."

Following these instructions, participants completed the math performance test. The STAI was administered a third time after the math performance test and asked participants to rate how they felt during the math test. Participants were then asked to complete the Math Performance Questionnaire that asks about their perceptions of their performance on the math test and their endorsement of gender stereotypes. The experimenter then saved the participant's data on the computer and administered two paper and pencil post-study surveys consecutively.

Participants received a debriefing form explaining the purpose of the experiment and any questions were answered by the experimenter.

Results

The main purpose of this study was to determine which of the proposed implicit measures would best capture motivational responses to challenge and threat for use in Study 2. With 4 separate implicit measures being examined, this study was powered for the purpose of exploring possible patterns in the data rather than for drawing inferential statistical conclusions. As we suspected, based on lack of power for determining inferential differences, there were no significant main effects or interactions for any of the implicit measures[1]. Therefore, for the implicit measures in this study, we focused on examining mean differences rather than statistical tests of significance. The scoring procedures for each measure are described below, followed by a table that includes means and standard deviations by condition for each measure. In each case, the effect size was defined such that positive values indicated that more challenge was observed in the condition where participants were expected to feel challenge and more threat was observed in the condition where participants were expected to feel threat.

Attitude Accessibility. This measure was scored using procedures to determine a valenced measure of attitude accessibility described by Rhodes and Roskos-Ewoldsen (2009). First, to address the common issue that reaction times are typically highly skewed, all reaction times were subjected to an inverse transformation and multiplied by 100 before analyses. The transformed reaction times were then multiplied by -1 for "dislike" responses and 1 for "like" responses. We summed these scores for each category of threat, challenge, and neutral words. Lower reaction times indicate higher accessibility. A difference score for this measure was then created by

[1] Statistical tests examining the interaction of gender and bias condition for each measure indicated that there were no significant differences for performance on Attitude Accessibility [$F(1,42)$ 2.51, $p = .12$], Stroop [$F(1,41)$ 2.31, $p = .136$], IAT [$F(1,49)$ .159, $p = .69$], or on the Word Completion Task [$F(1,37)$.28, $p = .6$].

subtracting reaction times for challenge responses from reaction times for threat responses. Higher numbers for the difference score indicate stronger activation for challenge words than for threat words. The means and standard deviations for this measure are presented by condition in Table 1. These results indicate that participants had a higher activation for challenge words than threat words in all conditions. An effect size of -.82 indicated that these results were not consistent with the predictions.

Table 1. Means and standard deviations by condition for the Attitude Accessibility task.

	Fair	Bias
Men	N = 9 *M* = -.18, *SD* = .033	N = 2 *M* = -.25, *SD* = .01
Women	N = 17 *M* = -.19, *SD* = .04	N = 19 *M* = -.20, *SD* = .05

Implicit Associations Test. This measure was scored using revised scoring procedures for the IAT detailed by Greenwald, Nosek, and Banaji (2003). First, data was restricted to include only test trials and not practice trials. All trials with reaction times greater than 10,000 ms were excluded. Participants who had reaction times lower than 300 ms for more than 10% of their trials were also excluded. Means of reaction times for correct trials were computed for each block. Incorrect trials were set to have missing reaction times. A pooled *SD* was created for the block that included paired categories of Math/Challenge with Not Math/Threat and for the block that included paired categories of Not Math/Challenge with Math/Threat. Error latencies were replaced with the mean of correct latencies for each block, plus 600 ms. New means were then created using the revised means from the error trials. The revised raw mean differences were then divided by the pooled *SD* to create a final score for IAT performance in which positive values indicated a stronger association between math and challenge. This score was reversed in later analyses so that lower values for the difference score indicate a stronger association between math and challenge than math and threat. The means and standard deviations for this measure are presented by condition in Table 2. These results indicate that participants had a higher activation for challenge words than threat words in all conditions. An effect size of -.10 indicated that these results were not consistent with the predictions. The results for this measure show that participants had a higher activation for challenge words than threat words in all conditions.

Table 2. Means and standard deviations by condition for the Implicit Associations Test.

	Fair	Bias
Men	N = 7 *M* = -.26, *SD* = .35	N = 12 *M* = -.18, *SD* = .28
Women	N = 20 *M* = -.25, *SD* = .27	N = 15 *M* = -.2, *SD* = .36

Stroop Task. This measure was scored using procedures similar to those used by Mogg, Bradley, Williams, and Mathews (1993). Reaction times were excluded that were less than 100 ms, greater than 4000 ms, or greater than the mean plus 3 *SD*. Means of reaction times were calculated for each condition: challenge, threat, and neutral. Longer reaction times indicated greater interference. Difference scores were then created by subtracting reaction times for challenge from reaction times for threat. Lower values for the difference score indicate a stronger association between math and challenge than between math and threat. The means and standard deviations for this measure are presented by condition in Table 3. The results from the current study indicate that women in the fair condition and men in the bias condition had higher activation for challenge words compared to threat words than men in the fair condition and women in the bias condition. A medium effect size (d = .49) indicated that these results are consistent with our predictions for the effect of stereotype threat.

Table 3. Means and standard deviations by condition for the Stroop task.

	Fair	Bias
Men	N = 7 *M* = 28.03, *SD* = 78.47	N = 8 *M* = -8.25, *SD* = 27.13
Women	N = 16 *M* = -6.62, *SD* = 50.54	N = 15 *M* = 9.95, *SD* = 53.21

Word Completion Task. For this measure, participants responded by generating words that were later compared to the list of challenge and threat words that were used in the other implicit measures in this study. A score for challenge words was created by summing the number of words that the participant generated that matched words on the challenge list, and a score for threat words was created by summing the number of words that the participant generated that matched words on the threat list. Higher sums indicate a greater activation for that particular type of word. Difference scores were then created by subtracting the score for challenge words created from the score for threat words created. Lower values for the difference score indicate a stronger activation for challenge words than for threat words. The means and standard deviations for this measure are presented by condition in Table 4. These results indicate that participants had a higher activation for challenge words than threat words in all conditions. An effect size of -.18 indicated that these results were not consistent with the predictions. The results from the current study indicate that participants had a higher activation for challenge words than threat words in all conditions.

Table 4. Means and standard deviations by condition for the Word Completion task.

	Fair	Bias
Men	N = 7 *M* = -2.43, *SD* = .35	N = 6 *M* = -1.5, *SD* = 1.87
Women	N = 14 *M* = -.5, *SD* = 2.9	N = 15 *M* = -.6, *SD* = 2.06

These 4 implicit measures were examined for patterns that would be consistent with stereotype threat predictions. The Stroop task was the only measure in which means of difference scores indicated a pattern of results that is consistent with stereotype threat literature such that challenge would be activated for women in a fair condition and men in a bias condition, while threat will be activated for women in a bias condition and men in a fair condition. The other 3 implicit measures indicated a pattern of results that is inconsistent with a stereotype threat prediction. The Stroop task produced a medium-sized effect, which was deemed acceptably strong. Based on these results, the Stroop task was selected for use in Study 2.

Working Memory Task. Scores for this measure were created by totaling accuracy for reporting the number of vowels in the sentences and for totaling the number of words correctly recalled (Schmader & Johns, 2003). The general linear model (GLM) was used to examine main effects and interactions of gender and bias condition on the number of words accurately recalled and correctly reporting the number of vowels for each sentence. The results indicated that participants were able to recall an average of 37 words and were able to accurately report the number of vowels for an average of 38 sentences. There were no main effects of gender [$F(1, 185)=.623$, $p = .431$] or bias condition [$F(1,185)=.297$, $p = .586$] on number of words recalled. The bias condition by gender interaction was also not significant for number of words recalled [$F(1,185)=.00$, $p = .986$]. For the number of vowels accurately counted, results indicated that there were no main effects of gender [$F(1,185)=.551$, $p = .459$] or bias condition [$F(1,185)=.585$, $p = .445$] and the bias condition by gender interaction was not significant [$F(1,185)=.017$, $p = .898$]. These results indicate that working memory was not significantly effected by stereotype threat in this study.

Math Performance Test. GLM was used to examine main effects and interactions of gender and bias condition on performance on the math test. On average, participants answered 14.76 math problems correctly. Results indicated a significant main effect for ACT on the math performance test such that participants with higher ACT scores performed better on the math performance test [$F(1,184)= 95.88$, $p <.001$]. A main effect of gender for reaction times to respond to questions on the math performance test indicated that women responded more quickly than men [$F(1,184)= 8.36$, $p =.004$]. There were no significant main effects for gender [$F(1,184)= 1.81$, $p =.180$] or bias condition [$F(1,184)= .92$, $p =.34$] and no significant interaction for gender by bias condition [$F(1,184)= .59$, $p =.443$] on the number of math questions that were answered correctly. Overall, the results suggest that bias condition and gender did not have a significant influence on performance on the math test in this study.

Math Performance Questionnaire. The GLM was used to examine main effects and interactions of bias condition and gender on various questions regarding participants' perceptions of their own performance on the math test. ACT scores were included in the model as a covariate for all analyses on this measure. When participants were asked to compare their performance on the math test to the average performance of other students of various education levels, there were main effects and interactions for specific education level comparisons. There were significant main effects of gender such that men, compared to women, reported that they felt that they performed better on the math test than the average freshman [$F(1,184)= 7.29$, $p = .008$] and the average sophomore [$F(1,184)= 5.00$, $p = .027$], but believed they performed worse than the average junior [$F(1,184)= 4.45$, $p = .036$]. Other significant main effects of gender showed that women's ratings of their performance on the test was significantly lower than men's ratings [$F(1,184)= 5.00$, $p = .027$], and women were less likely than men to attribute their performance on the test to gender influences [$F(1,184)= 6.55$, $p = .011$]. There was also a significant bias condition by gender interaction such that men in the bias condition estimated that their performance would be worse than college freshmen compared to men in the fair condition [$F(1,184)= 4.58$, $p = .034$].

Before the bias manipulation, participants were also asked to rate how important it is to them to do well in math. No main effects or interactions were found for gender or bias condition. The mean ratings for math importance (2.05 on a scale of 1 to 5) indicated that the majority of participants in the study rated their success in math as very important to them. Additionally, there were no significant differences for mean ratings of test difficulty. The mean rating of 2.76 on a scale of 1 to 5 indicated that the majority of participants rated the test as somewhat difficult.

STAI. A within-subjects repeated-measures analysis of the effects of time, gender, and bias condition was used to predict self-reported state anxiety during the study. A significant main effect was found for time, such that self-reported anxiety increased significantly each time the STAI was administered throughout the study [$F(2,412)= 139.02$, $p <.001$]. As illustrated in Figure 1, there was a significant interaction of time by bias condition by gender that suggests that men in the fair condition reported feeling significantly more anxiety during time two than men in the bias condition. Figure 2 illustrates that women in the fair condition report significantly greater anxiety at time 3 than women in the biased condition [$F(2,412)= 5.21$, $p =.006$].

Figure 1. Estimated marginal means for men's responses to the STAI at each time point.

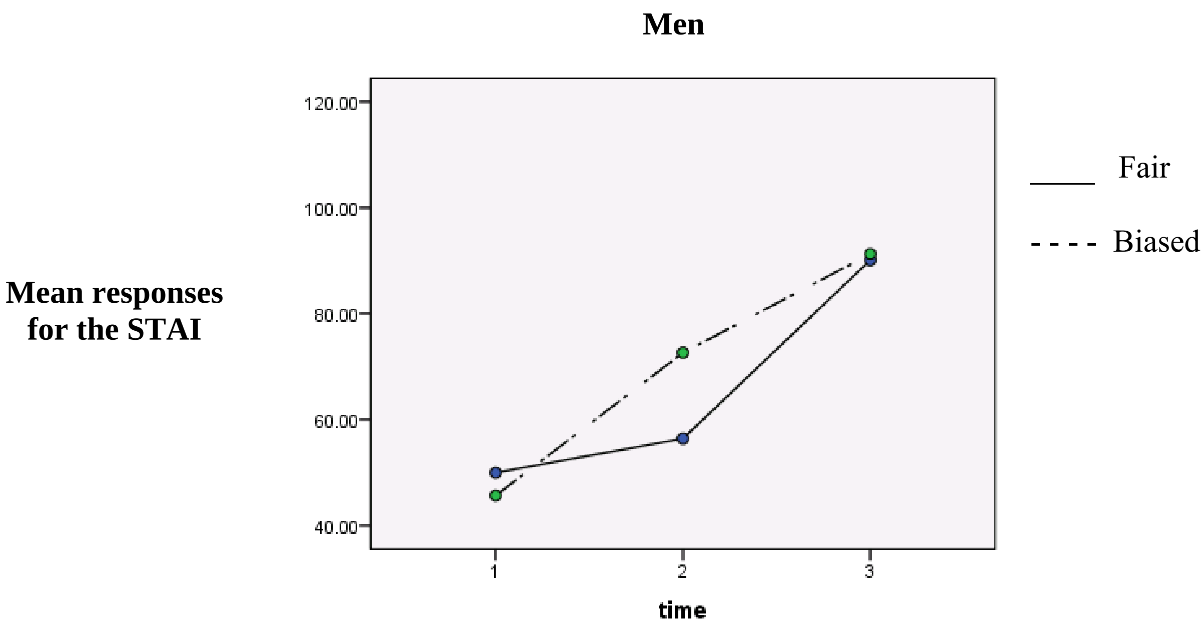
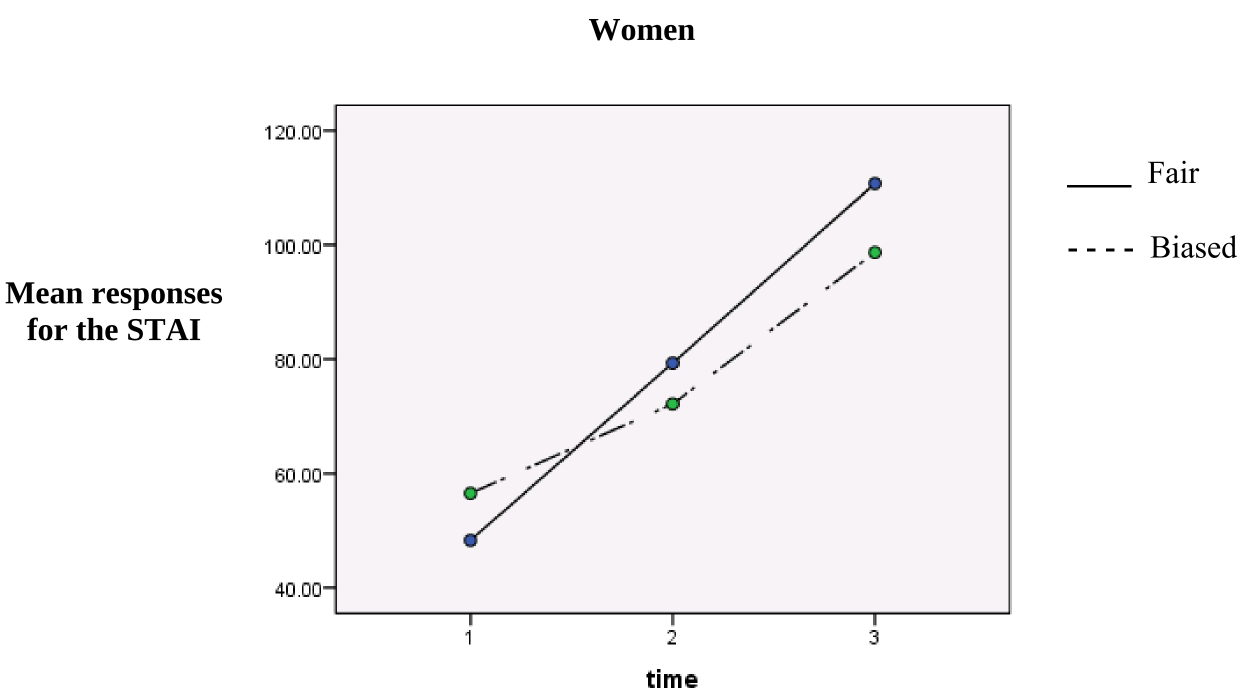


Figure 2. Estimated marginal means for women's responses to the STAI at each time point.



Discussion

The Stroop task had a medium effect size and demonstrated patterns were consistent with predictions for stereotype threat in all conditions. The three other implicit measures examined in Study 1 resulted in patterns that were inconsistent with stereotype threat predictions. In all three of these implicit measures, participants had stronger associations with challenge than with threat regardless of bias condition or gender. As a result of the findings of Study 1, the Stroop was selected for use in Study 2. We predicted that stereotype threat would impair working memory; however, there were no significant differences among groups for performance on the working memory task used in this study.

Performance on the math test was normally distributed, with ACT scores significantly predicting performance. Although women tended to respond to questions significantly faster than men on the math test, there were no gender differences for accuracy. Additionally, bias condition did not have an effect on math performance in this study. The majority of participants reported that being successful in math was very important to them. Performance on other measures was not influenced by ratings of math importance due to limited variability for this variable. On average, participants rated the math test as somewhat difficult. The math test proved to be an appropriate measure for this study because although participants viewed it as challenging, it was not so difficult that participants could not perform well, resulting in a normal distribution of scores. Men, compared to women, rated their own performance on the test higher and generally felt that they performed better than other students in math. These results support the idea that men are more confident and have higher esteem than women for math-related tasks.

State anxiety significantly increased throughout the study for participants regardless of gender and bias condition. Additionally, an interaction among time, bias condition, and gender revealed that men who were told that the test is gender fair increased their state anxiety levels immediately after the implicit measure more than men in the bias condition. This is consistent with the prediction that men in the fair condition would experience greater threat or anxiety on the measures than men in the bias condition. Additionally, women in the fair condition reported significantly greater anxiety after the math test than women in the biased condition. This is contrary to the prediction that women would experience less threat and anxiety following instructions indicating that the test is fair.

STUDY 2

Based on the results of Study 1, we selected the Stroop task to be used in Study 2 to test the ability of this measure to examine implicit motivations for challenge and threat. In Study 2, a larger sample size was used to examine these effects using statistical tests. In addition, Study 2 paired the Stroop task with physiological measures to expand the BPS model to include physiological responses to challenge and threat.

Method

Participants

College students (N= 181; 116 females, 65 males) enrolled in Introductory to Psychology from the Psychology Subject Pool at the University of Alabama participated in this study. Data were excluded for 12 participants who guessed the hypothesis and 7 participants for whom equipment malfunctioned. Participants (N = 162 ; 103 females, 59 males) consisted of 85.2% White students, 9.9% Black students and 4.9% who reported another race or ethnicity. Participants had a mean age of 18.75, a mean ACT score of 24.55, and a mean GPA of 3.46. Additional participants were excluded in analyses for the physiological component of this study due to faulty electrodes and malfunctioning equipment (N = 116). Participants earned 1.5 research credits in exchange for their participation.

Design

A 2 (gender) x 2 (stereotype threat: gender-fair test, gender-biased test) between-subjects design was used in this study. The dependent variables include number of questions answered correctly on a math test, performance on the Stroop task, and cardiac output.

Materials and Training

The materials and procedures in Study 2 generally replicate those in Study 1. The exceptions were that all participants completed the Stroop task, the working memory task was excluded, and physiological measurement was added. All physiological equipment used in this study was purchased from Biopac Systems, Inc. All administrators of the experiment attended a formal training session in which a representative from Biopac provided demonstrations and instructions on using the equipment and software, including equipment setup, participant preparation, proper application and removal of electrodes, recording procedures, and data analysis techniques.

AcqKnowledge Software. This software is used to view, measure, analyze and transform data that is collected by Biopac data acquisition systems. Version 3.9.1 was used to record the physiological data in this study and version 4.1.1, which has more efficient data analysis capabilities, was used to prepare and analyze the data in this study.

MP100 Data Acquisition System. This is used with Biopac amplifiers and accessories to allow various modules to interface with one another in the data recording process.

STP100C. This Biopac module was required to allow the physiological equipment to interface with E-Prime, the computer program used to administer the experiment.

NICO100C. This Biopac module connected to leads (Biopac Lead130) that were clipped to 4 pre-gelled, disposable band electrodes (Biopac EL506) worn by the participant to measure impedance cardiography (ICG). Two band electrodes were positioned on the back the participant's neck the other 2 band electrodes were positioned on the participant's lower back. The electrodes alternate emitting a low level (400 μA) current allowing impedance to be measured by changes that occur in the thorax volume. By recording the changes in impedance that occur over time, the NICO100C allows us to measure cardiac heart rate (HR) and (LVET) using noninvasive methods. These measurements allowed us to calculate stroke volume (SV) which represents the amount of blood pumped by the heart for a given heartbeat and cardiac output (CO) which is the amount of blood in liters that is ejected by the left ventricle of the heart each minute. We then multiplied HR by SV to derive a measure of CO from our recordings.

Procedure

Participants for Study 2 were recruited through the psychology subject pool website and were tested individually in the lab. Upon arriving at the lab, participants were given the information sheet for Study 2. Participants were told that we are using physiological equipment to examine responses that occur when people perform cognitive performance tasks. They were told that we are measuring their heart rate and cardiac output while they complete the study and that the measures were non-invasive and not harmful.

Standard procedures were followed to collect ICG data (Berntson, Quigley, & Lozano, 2007; Blascovich, Mendes, Hunter, & Salomon, 1999; Brownley, Hurwitz, & Schneiderman, 2000; Scheepers, 2009; Sherwood, et al., 1990; Townsend, et al., 2010). A trained experimenter applied 2 electrodes to the participant's neck and 2 electrodes to the lower back of the participant and measured the distance between the two innermost bands. The length of the area between the bands was used in the calculation for SV. Using this technique, changes are generated in the output voltage between the electrodes that indicate changes in SV (Brownley, Hurwitz, &

Schneiderman, 2000). The participant was seated at the computer and was given time to relax and fill out a demographics form. While the participants filled out the demographics form, the experimenter began recording data. After the participant completed the demographic form, the experimenter administered the computerized portion of the experiment. First, participants were provided with a welcome screen that provided a brief explanation of the study and emphasized the importance of minimizing movement while wearing the electrodes to avoid introducing artifact to the physiological recordings. Participants were then told that a blank screen would appear for 5 minutes in which they were asked to relax their bodies to allow the equipment to calibrate. This 5 minute resting period would later be used as a baseline measurement of HR and CO (Blascovich et al., 2004; Mendes, Major, McCoy, & Blascovich, 2008; Scheepers, 2009; Townsend, et al., 2010). After 5 minutes, the computer automatically progressed to the next portion of the study. The participant then completed the first STAI, as in Study 1. Next, the computer presented the following instructions:

> "The following 2 tasks have been used to identify people who are good at math. In order to assess your math ability, these tasks require that you pay attention to stimuli and solve problems as accurately and quickly as possible."

Participants were then provided with the rationale used in Study 1 which justified the use of the Stroop task as a measure of math-related ability. Participants were randomly assigned to fair or bias test conditions and all participants then completed the Stroop task. The remaining procedures replicated those used in Study 1. Physiological data was recorded throughout the duration of Study 2, however, only 1-minute time slices of data were used in the analyses in accordance with standard procedures used in previous research (Blascovich, Mendes, Hunter, Lickel, & Kowai-Bell, 2001; Blascovich & Seery, 2007; Blascovich et al., 2004; Hoyt & Blascovich, 2007; Hoyt & Blascovich, 2010; Mendes, Blascovich, Hunter, Lickel, & Jost, 2007;

Mendes et al., 2008; Scheepers, 2009; Sherwood, 1990; Tomaka, Blascovich, Kelsey, & Leitten, 1993; Tomaka, Blascovich, Kibler, & Ernst,1997; Vick et al., 2008). At the conclusion of the computerized portion of the experiment, the experimenter removed the electrodes from the participant, administered the post study surveys used in Study 1, provided a debriefing form, and answered any questions raised by the participant.

Consistent with previous literature, we predicted that HR would increase during both challenge and threat states, however during threatened states, little or no change would occur in CO as a result of inhibition of epinephrine (Vick et al., 2008). These cardiac measures have been validated in several studies as appropriate methods for examining cardiovascular indexes of challenge and threat (Blascovich et al., 2004; Blascovich & Seery, 2007; Tomaka, Blascovich, Kelsey, & Leitten, 1993; Tomaka et al., 1997).

Results

The main purpose of Study 2 was to determine whether the Stroop task would be a valid measure for examining motivational responses to challenge and threat by relating it to physiological responses. First, the GLM was used to test for significant main effects and interactions for gender and bias condition on participants' Stroop performance when controlling for the influence of ACT scores. Then the relation between performance on the Stroop task and physiological measures was also examined.

Stroop Task. Procedures used to score the Stroop task were the same as those described in Study 1. Lower values for the difference score indicate a stronger association between math and challenge than between math and threat. The means and standard deviations for this measure are presented by condition in Table 5. An effect size of -.21 indicated that these results were not consistent with the predictions.

Table 5. Means and standard deviations by condition for the Stroop task in Study 2.

	Fair	Bias
Men	N = 30 *M* = -8.18, *SD* = 49.09	N = 29 *M* = .88, *SD* = 51.80
Women	N = 49 *M* = 3.63, *SD* = 38.04	N = 50 *M* = -7.52, *SD* = 52.84

There were no significant main effects for gender [F(1,137)= .327, p =.568] or bias condition [F(1,137)= .031, p =.861] and no significant interaction for gender by bias condition [F(1,137)= .886, p =.348] on Stroop performance. According to the means for Stroop performance, men demonstrated a stronger activation for challenge when they were in the fair condition than men in a bias condition and women in the bias condition demonstrated greater challenge than women in the fair condition. These results are contrary to our predictions for stereotype threat effects on the Stroop task, however these differences are not statistically significant. Overall, the results suggest that gender and bias condition did not have a significant influence on performance on the Stroop task used in this study.

Cardiovascular Measures. The last minute of baseline was compared to HR, CO, and SV during 3 different time segments of the experiment (Vick et al., 2008). These time segments included the first minute of the practice portion of the Stroop task, the first minute of the critical portion of the Stroop task, and the first minute of the math performance test. First, the raw means were Winsorized to adjust for outliers that were greater or less than 3 standard deviations (Townsend et al., 2010). The mean heart rate at baseline was 79.4 bpm. Second, the last minute of HR at baseline was tested to confirm that there were no significant baseline differences for the fair and bias groups [t(112) = .074, p = .09].

Difference scores were then created for each time segment by subtracting the last minute of baseline for HR, CO, and SV from the first minute of each task segment. These scores were then standardized and the GLM was used to test for significant main effects and interactions of gender and bias condition on HR and CO for each time segment. There were no significant main effects or interactions on any of these time segments for HR, CO, or SV. For example, there were no main effects of gender [$F(1,110)= .008$, $p =.928$] or bias condition [$F(1,110)= .027$, $p =.869$] or a bias condition by gender interaction [$F(1,110)= .956$, $p =.330$] for HR on the first minute of the math test, when accounting for differences at baseline. Overall, these results suggest that bias condition and gender did not have a significant influence on changes in HR, CO, or SV during this study.

Relating Stroop Performance to Cardiovascular Measures. Performance on the Stroop task was correlated with physiological measures to determine whether a significant relationship existed between the two measures. We predicted that participants in a state of challenge would show greater interference for challenge words than threat words on the Stroop and increased CO from baseline. Lower values for the difference score indicate a stronger association between math and challenge than between math and threat. The correlation, although not significant, shows a pattern consistent with our predictions such that decreases in threat compared to challenge detected by the Stroop were correlated with increased cardiac output [$r(100)= -.116$, $p = .247$). Heart rate is expected to increase from baseline for both challenge and threat states. Correlations with the Stroop task were not significant for challenge responses [$r(100)= .029$, $p = .772$] and threat responses [$r(100) = .036$, $p = .721$].

Math Performance Test. The GLM was used to examine main effects and interactions of gender and bias condition on performance on the math test. On average, participants answered 14.57 math problems correctly. Results indicated a significant main effect for ACT on the math performance test, such that participants with higher ACT scores performed better on the math performance test [F(1,137)= 37.895, p <.001]. There were significant main effects for gender on number correct and reaction times for responses, such that men answered significantly more math questions correctly than women [F(1,137)= 8.56, p =.004] and women responded significantly more quickly than men [F(1,137)= 7.65, p =.006]. The interaction for bias condition by gender was not significant for accuracy [F(1,137)= .067, p =.796] or response time [F(1,137)= .911, p =.341]. These results indicate that gender had a significant effect on the number of math questions answered correctly, but the stereotype threat manipulation did not significantly influence math performance in this study.

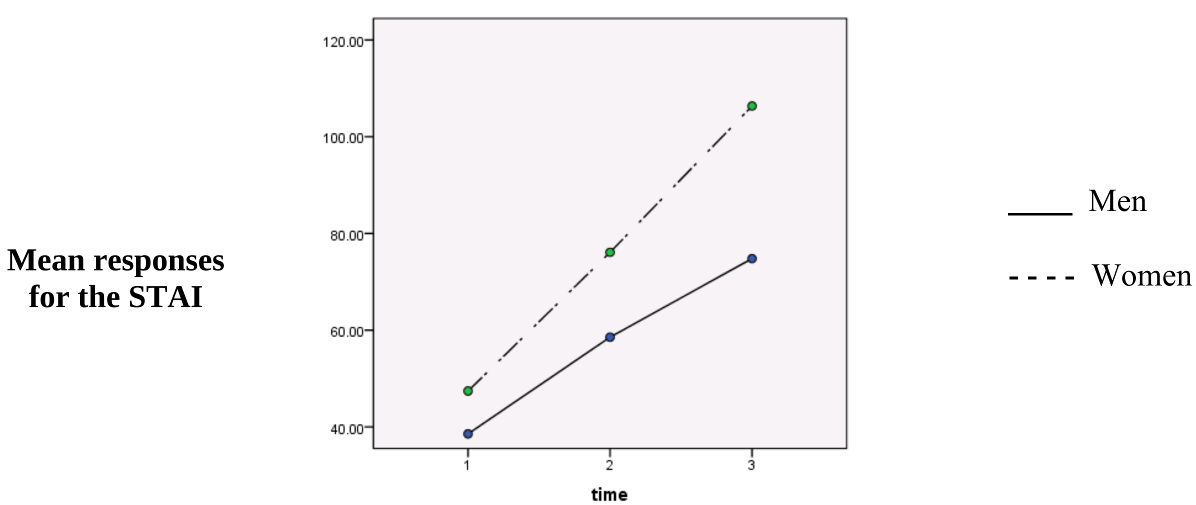
Math Performance Questionnaire. The GLM was used to examine main effects and interactions of bias condition and gender on various questions regarding participants' perceptions of their own performance on the math test. ACT scores were included in the model as a covariate for all analyses on this measure. There was a significant main effect of bias condition such that people in the biased condition, compared to the fair condition, rated their performance on the math test as significantly worse than an average college male student regardless of their gender [F(1,137)= 6.185, p =.014].

The interaction of gender by bias condition was not significant, however, there were several significant main effects for gender. Women were more likely than men to rate themselves as worse at math in general [F(1,137)= 8.06, p =.005] and on the math test used in the study [F(1,137)= 25.52, p <.001]. Additionally, women in the study rated the math test as more difficult than men did [F(1,137)= 4.61, p =.034]. When participants were asked to compare their performance on the math test to the average performance of other students of various education levels, women, consistently rated their performance as worse than every comparison group. Specifically, women were significantly more likely than men to rate their own performance as worse than the average freshman [$F(1,137)= 14.85$, $p <.001$], sophomore [$F(1,137)= 17.85$, $p <.001$], junior [$F(1,137)= 28.98$, $p <.001$], senior [$F(1,137)= 9.34$, $p =.003$], math major [$F(1,137)= 10.27$, $p =.002$], man [$F(1,137)= 13.15$, $p <.001$], and even compared to other women [$F(1,137)= 9.42$, $p =.003$].

Before the stereotype threat manipulation, participants were also asked to rate how important it is to them to do well in math. No main effects or interactions were found for gender or bias condition on this measure. The mean ratings for math importance (2.04 on a scale of 1 to 5) indicate that the majority of participants in the study rated their success in math as very important to them.

STAI. A within-subjects repeated measures analysis was used to examine the effects of time, gender, and bias condition on self-reported state anxiety during the study. A significant main effect was found for time, such that self-reported anxiety increased significantly each time the STAI was administered throughout the study [$F(2,306)= 169.58$, $p <.001$]. As illustrated in Figure 3, there was a significant time by gender interaction that suggests that the advantage of men over women increased over time [$F(2,306)= 9.81$, $p <.001$]. These results indicate that the stereotype threat manipulation did not have a significant effect on state anxiety.

Figure 3. Estimated marginal means for the STAI at each time point.



Discussion

The Stroop task was used in Study 2 to examine implicit motivational responses to challenge and threat. The results indicated that there were no significant differences based on the influence of gender or bias condition. In fact, men in a fair condition demonstrated a stronger activation for challenge than men in a bias condition and women in a bias condition demonstrated stronger challenge than women in a fair condition. Although these differences are not significant, they directly oppose our predictions for the effects of stereotype threat on this task. Results for the cardiovascular measures indicated that gender and bias condition did not influence HR, SV, and CO during the study. No changes occurred throughout the study based on these conditions that were significantly different when controlling for differences during baseline.

Although men answered significantly more questions correctly than women on the math test, bias condition did not play a role in performance on the math test. Likewise, there were no significant interactions between bias condition and gender on participants' perceptions of their performance on the math test. Women, regardless of bias condition, were significantly more likely than men to rate their performance as worse than other students, rate their test performance lower, and rate the math test as more difficult. Consistent with results from Study 1, state anxiety increased regardless of bias condition over time during the study, with state anxiety for women increasing more over time than for men. In light of these results, it is possible that stereotype threat was not successfully activated by the manipulation.

GENERAL DISCUSSION

The Stroop task was selected as a result of the findings of Study 1 to examine the potential effectiveness to detect differences in challenge and threat responses in Study 2. Physiological measures were recorded while participants completed the Stroop task and a math performance test in Study 2. Although Study 1 indicated that the Stroop would likely be an effective tool to examine challenge and threat, the results of Study 2 indicated that the Stroop measure was unable to detect changes consistent with predictions for stereotype threat. Overall, the manipulation was unsuccessful in activating stereotype threat for all measures which diminished the possibility of detecting a relation between performance on the Stroop and physiological measures of challenge and threat.

Limitations for these studies include methodological constraints and our inability to activate stereotype threat. In Study 1, we modified four widely used implicit measures. The modification of these measures included developing lists of challenge and threat words. Although these words were pretested for relating to challenge or threat, frequency of use, and length, there may be other factors to be considered in the development of these lists for use in developing implicit tasks. The four implicit measures used in this study were simply chosen for their popular use in cognitive research. However, we do not claim that these four implicit measures are the best possible measures available for examining underlying motivations. Although we believed that a medium effect size for the Stroop task was adequate for pursuing the use of this task in Study 2, a larger sample size would have allowed for statistical testing of the bias condition by gender interaction for each implicit measure. The limited sample size restricted our ability to conduct more predictive statistical tests and our ability to explore other implicit measures that might have shown stronger effects. Our sample size was further decreased due to equipment errors. Standard procedures were followed to maintain a more complete data set for analyses of cardiovascular measures by excluding participants due to equipment malfunction (Townsend et al., 2010). In this study, 28.4% of participants were excluded due to errors noise in the recordings, signal loss, and faulty electrodes. Despite efforts to correct these problems, it is not uncommon to exclude 10-23% of participants based on errors in recording physiological data (Blascovich et al., 2004; Seery, Weisbuch, & Blascovich, 2009; Townsend et al., 2010).

The second limitation, failure to activate stereotype threat, is more detrimental to the goals of this study. Although some of the measures showed gender differences, these differences should have resulted in significant interactions with bias conditions. Given that the manipulation was unsuccessful, variability decreases for challenge and threat responses reducing our ability to detect the effect of challenge and threat on the implicit and physiological measures.

We predicted that performance on a working memory task would be impaired as a result of stereotype threat; however, our results indicated no differences in performance on this measure. It is unclear whether the lack of differences was due to an ineffective manipulation, or to procedural differences in the administration of the working memory task. Working memory tasks used in stereotype threat research are typically administered immediately following stereotype threat activation (Schmader & Johns, 2003). However, the working memory task used in this study was included at the end of the study following the tasks of primary interest. Due to programming space limitations, the working memory task was the only measure that was included in a separate program presentation from the other measures. After all of the computerized measures were completed in using one program, the participants notified the experimenter who then administered the working memory task from a separate program. It is possible that because the working memory task followed a break in the procedure and was presented at the end of the study within a separate program, participants may have inadvertently been led to believe that working memory was unrelated to the previous tasks causing the stereotype threat manipulation to become less salient.

Performance on the math test did not appear to be influenced by bias condition in either study. Although performance tasks such as this are typically influenced by stereotype threat manipulations, it is not necessary to demonstrate differences in underlying motivations when examined using the BPS model (e.g. Blascovich, Mendes, Hunter, & Salomon, 1999; Mendes et al., 2008; Scheepers, 2009). Therefore, although the math test used in this study did not show significant differences as a result of the threat manipulation, it may still be considered an effective tool in exploring physiological and implicit differences.

Several methodological considerations should be made in future research. In order to improve chances of successfully activating stereotype threat, other methods of threat activation should be explored. Stereotype threat is commonly invoked by subtly manipulating the diagnostic ability of the task (Croizet & Claire, 1998), theories of the malleability of intelligence (Aronson et al., 2002), or by explicitly emphasizing potential group differences on a task (Spencer et al., 1999). Some studies activate stereotype threat through very subtle methods including asking a participant to report their gender at the start of a study (e.g. Steele & Aronson, 1995). Other researchers activate threat by taking a more direct approach in which participants are explicitly reminded of the threat conditions (e.g. Spencer et al., 1999). In this study, we chose a moderately direct approach, in which participants read instructions that told them that the test was either gender fair or biased. Future research should explore whether a stronger manipulation is necessary to demonstrate challenge and threat responses for implicit measures.

Methodological practices can be improved by considering particular factors that might influence cardiovascular performance on physiological measures. Noise in the data might be reduced by using a shorter task to minimize participant movement throughout the study. Although the Stroop patterns were consistent with stereotype threat predictions in Study 1, these results were not replicated in Study 2. The only methodological difference between these two studies that might have influenced these results is that in Study 2, physiological responses were recorded during the Stroop. It is possible that attaching leads and telling participants that we were recording their physiological responses may have increased their awareness or discomfort which may have influenced response patterns on the physiological and implicit measures. To examine this possibility, the Stroop should be examined alone, using a sample size that would be adequate to test for statistical differences among groups. The main difference between Study 2 and previous research that examined physiological responses to challenge and threat is that the study took longer to complete in order to include the Stroop task. The time frame may be an important difference and future research would benefit by examining these measures separately. Additional implicit measures and word lists for challenge and threat should also be examined in future research. Newly developed measures take many repetitions to validate and these measures might benefit from further exploration including applying them to different domains related to stereotype threat.

Efforts were made in this study to consider the influence of individual differences including academic ability and domain importance. Participants in this study were selected from a college sample and their average GPAs and ACT scores indicate that they are successful students. Perhaps being successful in school in general weakened the negative effects that we predicted. A neutral condition should also be utilized in future research to examine responses for those in a stereotype-free situation compared to those in a bias or fair condition. Additional influences should be examined including strength of gender identification and tendencies to adhere to gender norms.

The results of these studies have several implications. First, these results demonstrate the complexity of successfully activating stereotype threat. Second, the studies introduce an implicit measure that may be refined in future research to potentially expand the BPS model to include cognitive implicit measures. Third, it would be useful to determine whether implicit motivations can be manipulated to induce particular responses to stereotype threat when stereotype threat is successfully activated.

References

Al-Fudail, M. & Mellar, H. (2008). Investigating teacher stress when using technology. *Computers & Education, 51,* 1103-1110.

Alter, A.L., Aronson, J. Darley, J.M., Rodriguez, C., & Ruble, D.N. (2010). Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. *Journal of Experimental and Social Psychology, 46*, 166-171.

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

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

Berntson, G.G., Quigley, K.S., & Lozano, D. (2007). Cardiovascular psychophysiology. In J.T.Cacioppo, L.G., Tassinary, & Berntson, G.G. (Eds.), *Handbook of Psychophysiology* (3rd ed.), 182-210. Cambridge: Cambridge University Press.

Blascovich, J. & Mendes, W.B. (2000). Challenge and threat appraisals: The role of affective cues. In J.P. Forgas (Ed.), Feeling and thinking: *The role of affect in social cognition* (pp. 59-82). New York, NY: Cambridge University Press.

Blascovich, J., Mendes, W.B., Hunter, S.B., Lickel, B. & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social*

Psychology, 80, 253–267.

Blascovich, J., Mendes, W.B., Hunter, S.B., & Salomon, K. (1999). Social "facilitation" as challenge and threat. *Journal of Personality and Social Psychology, 31,* 422-429.

Blascovich, J. Mendes, W.B., Tomaka, J., Salomon, K., & Seery, M. (2003). The robust nature of the Biopsychosocial model: A reply to Wright and Kirby. *Journal of Personality and Social Psychology, 7,* 234-243.

Blascovich, J. & Seery, M.D. (2007). Visceral and somatic indexes of social psychological constructs: History, principles, propositions, and case studies. In A. Kruglanski & E.T. Higgins (Eds.) *Social psychology: Handbook of basic principles.* New York: Guilford.

Blascovich, J., Seery, M.D., Mugridge, C.A., Norris, R.K. & Weisbuch, M. (2004). Predicting athletic performance from cardiovascular indexes of challenge and threat, *Journal of Experimental Social Psychology*, *40*, 683–688.

Blascovich, J., Spencer, S.J., Quinn, D.M., & Steele, C.M. (2001). Stereotype threat and the cardiovascular reactivity of African-Americans. *Psychological Science, 12*, 225-229.

Blascovich, J. & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. *Advances in Experimental Social Psychology, 28,* 1-51.

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

Brownley, K.A., Hurwitz, B.E., & Schneiderman, N. (2000). Cardiovascular psychophysiology. In J.T.Cacioppo, L.G., Tassinary, & Berntson, G.G. (Eds.), *Handbook of Psychophysiology* (2^{nd} ed.), 224-264. Cambridge: Cambridge University Press.

Chalabaev, A., Major, B., Cury, F., & Sarrazin, P. (2009). Physiological markers of challenge and threat mediate the effects of performance-based goals on performance. *Journal of Experimental Social Psychology, 45* (4), 991-994.

Chen, S., Duckworth, K., & Chaiken, S. (2000). Motivated heuristic and systematic processing. *Psychological Inquiry, 10* (1), 44-49.

Cohen, A.B., Shariff, A.F., & Hill, P.C. (2008). The accessibility of religious beliefs. *Journal of Research in Personality, 42,* 1408-1417.

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

Croizet, J., Despres, G., Gauzins, M., Huguet, P., Leyens, J., & Meot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30*, 721-731.

DeMarree, K.G., Petty, R.E., & Strunk, D. R. (2010). Self-esteem accessibility as attitude strength: On the durability and impactfulness of accessible self-esteem. *Personality and Social Psychology Bulletin, 36,* 628-641.

Dulan, S.W. (2007). McGraw-Hill's ACT 2008 Edition. McGraw-Hill Companies, Inc.

Fazio, R. H. (1995). Attitude as object-evaluation associations: Determinants, consequences and correlates of attitude accessibility. In R. E. Petty & J. A. Kronsnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247-282). Mahwah, NJ: Lawrence Erlbaum.

Frew, A.K. & Drummond, P.D. (2009). Opposite effects of opioid blockade on the blood pressure-pain relationship in depressed and non-depressed patients. *Pain, 142*, 58-78.

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

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464-1480.

Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Harris, J. (2003). http://www.wordcount.org/index2.html. Retrieved 06/18/2009.

Hess, T.M., Auman, C., Colcombe, S.J., & Rahhal T.A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58*, 3-11.

Hoyt, C. & Blascovich, J. (2007). Leadership efficacy and women leaders' responses to stereotype activation. *Group Processes and Intergroup Relations*, *10*, 595–616.

Hoyt, C.L., & Blascovich, J. (2010). The role of leadership self-efficacy and stereotype activation on cardiovascular, behavioral, and self-report responses in leadership domain. *The Leadership Quarterly, 21*(1), 89-103.

Inzlicht, M. & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365-371.

Jefferson, A.L. (2010). Cardiac output as a potential risk factor for abnormal brain aging. *Journal of Alzheimer's Disease, 20,* 813-821.

Keller, J. (2007). When negative stereotypic expectances turn into challenge or threat: The moderating role of regulatory focus. *Swiss Journal of Psychology, 66* (3), 163-168.

Keller, J. & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting effect on women's math performance. *Personality and Social Psychology Bulletin, 29,* 371-381.

Kriebig, S.D., Wilhelm, F.H, Roth, W.T. & Gross, J.J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear and sadness-inducing films. *Psychophysiology, 44,* 787-806.

Kubicek, W.G., Karnegis, J.N., Patterson, R.P., Witsoe, D.A., & Mattson, R.H. (1966). Development of an impedance cardiac output system. *Aerospace Medicine, 37,* 1208-1212.

Lawler-Row, K.A., Karremans, J.C., Scott, C., Edlis-Matityahou, M., & Edwards, L. (2008). Forgiveness, physiological reactivity and health: The role of anger. *International Journal of Psychobiology, 68,* 51-58.

MacLeod, C.M. (1999). Half a century of research on the Stroop Effect: An integrative review. *Psychological Bulletin, 109 (2),* 163-203.

McCabe, R.E. (1999). Implicit and explicit memory for threat words in high- and low-anxiety sensitive participants. *Cognitive Therapy and Research, 23 (1),* 21-38.

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

McNally, R.J., Hornig, C.D., Hoffman, E.C., & Han, E.M. (1999). Anxiety sensitivity and cognitive biases for threat. *Behavior Therapy, 30,* 51-61.

Mendes, W.B., Blascovich, J., Hunter, S.B., Lickel, B., & Jost, J.T. (2007). Threatened by the unexpected: Physiological responses during social interactions with expectancy violating partners. *Journal of Personality and Social Psychology*, *92*, 698–716.

Mendes, W.B., Major, B., McCoy, S., & Blascovich, J. (2008). How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance, *Journal of Personality and Social Psychology*, *94*, 278–291.

Mogg, K., Bradley, B.P., Williams, R., & Mathews, A. (1993). Subliminal processing of emotional information in anxiety and depression. *Journal of Abnormal Psychology, 102,* 304-311.

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics: Theory, Research, and Practice, 6* (1), 101-115.

Newman, M.L., Guinn Sellers, J., & Josephs, R.A. (2005). Testosterone, cognition, and social status. *Hormones and Behavior, 47*, 205-211.

Ohme, R., Reykowska, D., Wiener, D., & Choromanska, A. (2009). Analysis of neurophysiological reactions to advertising by means of EEG and galvanic skin response measures. *Journal of Neuroscience, Psychology, and Economics, 2,* 21-31.

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

Osborne, J.W. (2007). Linking stereotype threat and anxiety. *Educational Psychology, 27,* 135 154.

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

Ramachandran, V., Depalma, N, & Lisiewski, S. (2009). The role of mirror neurons in processing vocal emotions: Evidence from psychophysiological data. *International Journal of Neuroscience, 119,* 681-690.

Rhodes, N. & Ewoldsen, D.R. (2009). Attitude and norm accessibility and cigarette smoking. *Journal of Applied Social Psychology, 39,* 2355-2372.

Roskos-Ewoldsen, D. R., & Fazio, R. H. (1992). On the orienting value of attitudes: Attitude accessibility as a determinant of an object's attraction of visual attention. *Journal of Personality and Social Psychology, 63,* 198-211.

Scheepers, D. (2009). Turning social identity threat into challenge: Status stability and cardiovascular reactivity during inter-group competition. *Journal of Experimental Social Psychology, 45* (1), 228-233.

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,* 440-452.

Seery, M.D., Weisbuch, M., & Blascovich, J. (2009). Something to gain, something to lose: The cardiovascular consequences of outcome framing. *International Journal of Psychophysiology, 73,* 308-312.

Shen, L., Monahan, J.L., Rhodes, N., Roskos-Ewoldsen, D.R. (2009). The impact of attitude accessibility and decision style on adolescents' biased processing of health-related public service announcements. *Communication Research, 36,* 104-129.

Sherwood, A., Allen, M.T., Fahrenberg, J. , Kelsey, R.M., Lovallo, W.R. & van Doornen, L.J., (1990) Methodological guidelines for impedance cardiography. *Psychophysiology*, *27*, 1-23.

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

Sinclair, L. & Kunda, Z. (1999). Reactions to a black professional: motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77* (5), 885-904.

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

Spielberger, C.D., Gorsuch, R.L., Lushene, P.R., Vagg, P.R., Jacobs, A.G. (1983). *Manual for the State Trait Anxiety Inventory (Form Y).* Consulting Psychologists Press, Inc.: Palo Alto.

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

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

Stone, J., Lynch, C.I., Sjomeling, M., & Darley, J.M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology, 77* (6), 1213-1227.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Tomaka, J. Blascovich, J., Kelsey, R.M., & Leitten, C.L. (1993) Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology*, *65*, 248–260.

Tomaka, J. Blascovich, J., Kibler, J., & Ernst, J.M. (1997). Cognitive and physiological antecedents of threat and challenge appraisals. *Journal of Personality and Social Psychology*, *73* (1), 63–72.

Townsend, S.S., Major, B., Sawyer, P.J., & Mendes, W.B. (2010). Can the absence of prejudice be more threatening than its presence? It depends on one's worldview. *Journal of Personality and Social Psychology, 99,* 933-947.

Vick, S.B., Seery, M.D., Blascovich, J., & Weisbuch, M. (2008). The effect of gender stereotype activation on challenge and threat motivational states. *Journal of Experimental Social Psychology, 44,* 624-630.

Von Hippel, W., Von Hippel, C., Conway, L., Preacher, K.J., Schooler, J.W., Radvansky, G.A. (2005). Coping with stereotype threat: Denial as an impression management strategy. *Journal of Personality and Social Psychology, 89* (1), 22-35.

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

Williams, L., O'Carroll, R.E., & O'Connor, R.C. (2009). Type D personality and cardiac output in response to stress. *Psychology & Health, 24,* 489-500.

Appendix A

Math Performance Test

Please press the letter key that corresponds to the best possible answer.

1. 4x-9 = 11, then x =
    A. 5
    B. 6
    C. 6.5
    D. 9
    E. 16

Correct: A

2. If P = 5a and Q = 3b-2a, then what is the value of P-Q?
    A. 7a + 3b
    B. 3a + 3b
    C. 7a – 3b
    D. 3a – 3b
    E. 5a – 3b

Correct: C

3. Amanda ate lunch at a restaurant, where her bill was $27.60. She tipped 15%. What was the amount of Amanda's tip?
    A. $4.14
    B. $12.60
    C. $15.00
    D. $23.46
    E. $31.74

Correct: A

4. The average of 8 numbers is 6.5. If each of the numbers is decreased by 3, what is the average of the 8 new numbers?
    A. 0.0
    B. 3.5
    C. 4.0
    D. 7.5
    E. 9.5

Correct: B

5. The expression 5a + 5b is equivalent to which of the following?
    A. 5(a-b)

B. 10(a+b)
C. 5ab
D. 5(a+b)
E. 10ab

Correct: D

6. A classroom has (r+s) rows of seats and t seats in each row. Which of the following is an expression for the number of seats in the entire classroom?
   A. rst
   B. (rs) + (rt)
   C. t + (rs)
   D. r + s + t
   E. (rt) + (st)

Correct: E

7. If 20% of x equals 16, then x = ?
   A. 2
   B. 3.2
   C. 32
   D. 80
   E. 800

Correct: D

8. You have been asked to make punch for your friend's birthday party. The punch recipe calls for 9 quarts of fruit juices to 4 quarts of soda. To make 52 quarts of this punch, how many quarts of soda should you use?
   A. 4
   B. 9
   C. 13
   D. 16
   E. 36

Correct: D

9. If 5x + 4 = 7(x-2), then x = ?
   A. 4
   B. 5
   C. 7
   D. 9
   E. 18

Correct: C

10. If x is a positive integer that divides evenly into both 64 and 96 but divides evenly into neither 16 nor 20, what should you get when you add the digits in x?
    A. 3
    B. 5
    C. 7
    D. 8
    E. 10

Correct: B

11. Which of the following lines has the greatest slope?
    A. $y = 2x + 5$
    B. $y = 5x - 4$
    C. $y = 3x + 8$
    D. $3y = 9x + 6$
    E. $4y = 4x - 8$

Correct: B

12. A certain rectangle is 4 times as long as it is wide. Suppose the length and width are tripled. The area of the second rectangle is how many times as large as the area of the first?
    A. 3
    B. 4
    C. 9
    D. 12
    E. 16

Correct: C

13. What is the smallest possible value for the product of 2 real numbers that differ by 12?
    A. –36
    B. –27
    C. –11
    D. 0
    E. 13

Correct: A

14. If $2x + 5 = 17$, then x =?
    A. 3
    B. 6
    C. 10
    D. 11
    E. 24

Correct: B

15. A carton of 12 cans of soda is priced at $6.60 now. If the soda goes on sale for 20% off the current price, what will be the price of the carton?

   A. $0.55
   B. $1.32
   C. $5.28
   D. $6.36
   E. $6.40

Correct: C

16. If x= -4, then 21 – 3(x-2) =?

   A. 3
   B. 11
   C. 15
   D. 27
   E. 39

Correct: E

17. The expression 8x – 8y is equivalent to which of the following?

   A. 8(x – y)
   B. 8(x + y)
   C. 8xy
   D. –8(x + y)
   E. –8xy

Correct: A

18. Which of the following equations has both x = -3 and x = 6 as solutions?

   A. (x – 6)(x + 3) = 0
   B. (x + 6)(x + 3) = 0
   C. (x + 6)(x – 3) = 0
   D. (x – 6)(x – 3) = 0
   E. x – 6 = x + 3

Correct: A

19. For all x, 7 – 2(x – 10) = ?

   A. 5x + 27
   B. 5x – 27
   C. –2x +27
   D. –2x – 13
   E. –2x +3

Correct: C

20. If 60% of x equals 90, then x =?
   A. 5.4
   B. 15
   C. 54
   D. 150
   E. 1500

Correct: D

21. The price of 1 box of popcorn and 1 drink together is $5.10. The price of 2 boxes of popcorn and 1 drink together is $8.35. What is the cost of 1 drink?
   A. $0.75
   B. $1.85
   C. $2.15
   D. $2.55
   E. $3.25

Correct: B

22. An oil tank contains 4,800 gallons of oil. Each gallon of oil weighs approximately 6 pounds. About how many pounds does the oil in the tank weigh?
   A. 800
   B. 4,806
   C. 6,000
   D. 28,800
   E. 46,800

Correct: D

23. How many prime numbers are there between 36 and 54?
   A. 4
   B. 5
   C. 6
   D. 7
   E. 8

Correct: A

24. What is the smallest possible value for the product of 2 real numbers that differ by 8?
   A. 8
   B. 6
   C. –2
   D. –4
   E. –16

Correct: E

25. The expression a(b – 2c) is equivalent to:
    A. ab – 2a – 2c
    B. ab – 2ac
    C. ab – 2bc
    D. ab – b – 2c
    E. ab – 2b – c

Correct: B

26. The expression 4c – 2d is equivalent to which of the following:
    A. 4(c – 2d)
    B. 2cd
    C. 2(c – d)
    D. 4(c – d)
    E. 2(2c – d)

Correct: E

27. Which of the following is a simplified form of 4x + 2x + y – x?
    A. 3x + y
    B. 5x + y
    C. 2(x + 2)(x + y)
    D. 6x – y
    E. x(6 + y)

Correct: B

28. If 75% of x equals 180, then x =
    A. 24
    B. 135
    C. 240
    D. 450
    E. 2,400

Correct: C

29. A cd that normally sells for $14.50 is on sale for 30% off. What is the cost of the cd during the sale to the nearest cent?
    A. $4.49
    B. $10.47
    C. $11.21
    D. $11.96
    E. $44.85

Correct: B

30. The area of a rectangular kitchen is 180 square feet. If the length of the floor is 2 feet less than twice the width, what is the width of the floor in feet?
    A. 5
    B. 10
    C. 12
    D. 16
    E. 24

Correct: B

Appendix B

Challenge and Threat Word Lists for Implicit Measures

CHALLENGE	THREAT
brave	aggressive
conquer	harmful
challenge	imposing
achieve	peril
courage	assault
adventure	unable
rewarding	intimidate
ability	futile
win	frightful
proud	defensive
capable	threat
endure	fight
dare	cower
success	attack
trump	menace
fearless	damaging
determined	helpless
victory	scary
resolve	tense
prevail	alarm
goal	defeated
ace	looming
hardy	failure
strong	bully

Appendix C

Word Completion Task

achieve	_ _ _ i e v e	alarm	a l _ _ _
ability	a _ _ _ i t y	attack	a t t _ _ _
capable	c a p _ _ _ _	bully	b _ _ _ y
challenge	c h a _ _ _ _ _ e	danger	d a _ _ _ r
courage	c o _ _ _ _ e	defeated	d e f _ _ _ e d
conquer	c o _ _ _ e r	frightful	f _ _ _ _ _ f u l
endure	e_ _ _ r e	failure	f a _ _ _ r _
fortitude	f o r t _ _ _ _ _	helpless	h _ _ _ l e s s
goal	g o _ _	harmful	h a r _ _ _ _
proud	p _ _ _ d	menace	m e n _ _ _
prevail	p r _ _ _ _ l	threat	t h r _ _ _
resolve	r e s _ _ _ e	assault	_ _ _ _ u l t
rewarding	r e w _ _ _ _ n g	cower	_ o w e r
success	s _ _ _ e s s	damaging	_ _ _ _ g i n g
victory	v i _ _ _ _ y	defensive	_ _ _ _ n s i v e
ace	_ c e	fight	_ i g h t
adventure	_ _ _ _ _ t u r e	futile	_ _ t i _ e
brave	_ _ a v e	imposing	i m p o _ _ _ _
dare	d a _ _	intimidate	i n t i m _ _ _ _ _
determined	_ e t e _ _ _ _ e d	looming	l _ _ _ i n g
fearless	_ _ _ r l e s s	peril	p e r _ _
hardy	h a _ _ y	scary	s _ _ r y
strong	_ _ _ o n g	tense	t e n _ _
trump	t r u _ _	unable	u n _ _ _ e
win	_ i n	aggressive	_ _ _ _ e s s i v e

Appendix D

Participant # _____

Demographics Questionnaire

The tasks in this study will assess your cognitive abilities. First we would like to know some information about you and about your academic performance. Please answer the questions below honestly and to the best of your ability. If you can not recall the appropriate response for a particular question, please write "DK- do not know" as your response for that item. If you have any questions, please alert the experimenter.

1. Age: ___________

2. Gender- Circle one: Male Female

3. Race/Ethnicity- Circle one.
   - a.) White
   - b.) Black
   - c.) Hispanic
   - d.) Asian
   - e.) Other not listed ____________________________________

5. What is your current major or major of interest? ______________________________

4. Current GPA*: ________ (*If you are a first semester freshman, please indicate your high school GPA).

5. SAT score: ____________ Math ___________ Verbal ___________total

6. ACT score: ____________

7. Please list the college math courses you have taken so far.

____
____
____
____
____
____

8. How important do you believe it is for you to do well in math?
   - a.) Extremely important
   - b.) Very important
   - c.) Somewhat important
   - d.) Slightly important
   - e.) Not at all important

Appendix E

Participant # _____

Math Performance Questionnaire

Please rate your performance on the math test that you took during this experiment.

1.) In general would you consider yourself ________________ in math?

a.) Extremely above average
b.) Above average
c.) Average
d.) Below average
e.) Extremely below average

2.) How difficult was the math test?

a.) Extremely difficult
b.) Very difficult
c.) Somewhat difficult
d.) Slightly difficult
e.) Not at all difficult

3.) How well do you think you did on the math test?

a.) Extremely well.
b.) Very well.
c.) Somewhat well.
d.) Not very well.
e.) Extremely poorly.

4.) Please rate your performance on the math test compared to the average college <u>freshman</u>.

a.) Extremely above average
b.) Above average
c.) Average
d.) Below average
e.) Extremely below average

5.) Please rate your performance on the math test compared to the average college <u>sophomore</u>.

a.) Extremely above average
b.) Above average
c.) Average
d.) Below average
e.) Extremely below average

6.) Please rate your performance on the math test compared to the average college <u>junior</u>.

a.) Extremely above average
b.) Above average
c.) Average
d.) Below average

e.) Extremely below average

7.) Please rate your performance on the math test compared to the average college senior.
   a.) Extremely above average
   b.) Above average
   c.) Average
   d.) Below average
   e.) Extremely below average

8.) Please rate your performance on the math test compared to the average math major.
   a.) Extremely above average
   b.) Above average
   c.) Average
   d.) Below average
   e.) Extremely below average

9.) Please rate your performance on the math test compared to an average male college student.
   a.) Extremely above average
   b.) Above average
   c.) Average
   d.) Below average
   e.) Extremely below average

10.) Please rate your performance on the math test compared to an average female college student.
   a.) Extremely above average
   b.) Above average
   c.) Average
   d.) Below average
   e.) Extremely below average

11.) Please rate how much you agree with the following statement: I believe that my performance on this math test was influenced by my gender.
   a.) Strongly agree
   b.) Agree
   c.) Somewhat agree
   d.) Disagree
   e.) Strongly disagree

12.) Please rate how much you agree with the following statement: Men are better at math than women.
   a.) Strongly agree
   b.) Agree
   c.) Somewhat agree
   d.) Disagree
   e.) Strongly disagree

Appendix F

Working Memory Task Sentences

(Schmader & Johns, 2003)

1. They celebrate by dancing wildly.
2. The meeting was delayed again.
3. Don't give the fish too much food.
4. She likes to sing in the shower.
5. We need a pump to inflate the ball.
6. The audience was totally confused.
7. The power went out during the storm.
8. She speaks to people often.
9. He borrowed my bike to go to the store.
10. They were married in the spring
11. He kept his shoes freshly polished.
12. The bus was constantly late.
13. They barely won the last game.
14. The windows were covered with mud.
15. I'm going to be there next year.
16. The building was closed for repair.
17. The flowers bloomed after the rain.
18. They stopped to ask for directions.
19. The music was playing too loud.
20. She painted her room last weekend.
21. How many days are in March?
22. I didn't have time to do laundry.

23. She wasn't friendly to her guests.

24. We couldn't leave our bags in the trunk.

25. Please put your shoes in the closet.

26. I like to read the paper sometimes.

27. The phone woke me up this morning

28. Our cat got stuck in the tree.

29. There is a rabbit in the backyard.

30. She read a book on the train.

31. New products can create demand.

32. How many times can you apply?

33. A frog can't fly a plane.

34. He is always lying about something.

35. The quilt is warm on cold nights.

36. Try to keep your room clean.

37. She wore a new dress to work.

38. The stars were shining brightly.

39. I will go home next week.

40. We left the sprinklers running all night.

41. The people there were very friendly.

42. They didn't have enough cold drinks.

43. She travels a lot during the summer

44. The sun was behind the clouds.

45. She delivered an opening statement.

46. I watched a boring movie last night.

47. The paper was cut into small squares.

48. He always arrives too late.

49. He fell and sprained his wrist.

50. Some people don't believe in honor.

51. They drink fresh juice everyday.

52. The box was filled with magazines.

53. The wind blew the door open.

54. I was happy to hear from them.

55. I think they are a perfect match.

56. He won't be hungry after dessert.

57. It's hard to be nice to rude people.

58. He prepared a meal for his friends.

59. My dog likes to sleep on the couch.

60. The bottle was cracked on the side.

Appendix G

Working Memory Task Recall List

(Schmader & Johns, 2003)

1. green
2. fact
3. dress
4. heat
5. sum
6. knife
7. help
8. talk
9. need
10. hard
11. score
12. blue
13. lock
14. dust
15. beach
16. bad
17. forth
18. head
19. out
20. send
21. fight
22. gas
23. brain
24. theme

25. trade
26. file
27. moon
28. ear
29. east
30. jump
31. farm
32. bay
33. week
34. shall
35. aid
36. town
37. stay
38. camp
39. league
40. mouth
41. care
42. back
43. guy
44. list
45. guest
46. cause
47. spot
48. wife
49. rain
50. close
51. wait

52. set

53. beat

54. rock

55. tree

56. fit

57. bible

58. truck

59. nose

60. miss

Appendix H

Participant # _____

Post-Study Survey Page 1

What do you think this study was about?

Was there any time during the study where you weren’t sure what you were supposed to be doing?

Did anything unusual happen during the course of the study?

Other than the announcement on the research pool website, had you heard anything about this study before you came in today? If so, what did you hear?

Participant # _____

Post-Study Survey Page 2

Please circle which of the following instructions you were given before beginning the math test.

A. Some people believe that women are inferior to men at math, but empirical evidence has shown that women and men perform equally on this test.

B. Some people believe that women are inferior to men at math, but empirical evidence is mixed.

C. I do not remember being given either of these sets of instructions before beginning the math test.

[*The question below was only included in Study 2*].

So that we can improve accuracy in interpreting changes in your heart rate throughout the study, please provide the following information (estimate if necessary!):

Height (in feet and inches): __________

Weight (in pounds): __________

Appendix H

IRB Approval Form

Office for Research

Institutional Review Board for the Protection of Human Subjects

THE UNIVERSITY OF ALABAMA
RESEARCH

June 15, 2010

Jamie O'Mally
Department of Psychology
College of Arts & Sciences
Box 870216

Re: IRB#: 10-OR-199 –ME "Implicit Measurement of Threat and Challenge as Motivational Responses to Stereotype Threat"

Dear Ms. O'Mally:

The University of Alabama Medical Institutional Review Board has granted approval for your proposed research.

Your application has been given expedited approval according to 45 CFR part 46. Approval has been given under expedited review category 4 as outlined below:

(4) Collection of data through noninvasive procedures (not involving general anesthesia or sedation) routinely employed in clinical practice, excluding procedures involving x-rays or microwaves. Where medical devices are employed, they must be cleared/approved for marketing.

Your application will expire on June 14, 2011. If your research will continue beyond this date, complete the relevant portions of Continuing Review and Closure Form. If you wish to modify the application, complete the Modification of an Approved Protocol. Changes in this study cannot be initiated without IRB approval, except when necessary to eliminate apparent immediate hazards to participants. When the study closes, complete the appropriate portions of the Continuing Review and Closure Form.

Please use reproductions of the IRB approved stamped information sheets to obtain consent from your participants.

Should you need to submit any further correspondence regarding this proposal, please include the above application number.

Good luck with your research.

Sincerely,

Carpantato T. Myles, MSM, CIM
Director & Research Compliance Officer
Office of Research Compliance
The University of Alabama

152 Rose Administration Building
Box 870117
Tuscaloosa, Alabama 35487-0117
(205) 348-5152
FAX (205) 348-8882