Sticks, stones, & stigmas: A social-cognitive account of stereotype threat mechanisms

Michael C. Philipp

University of Northern Iowa

Copyright ©2005 Michael C. Philipp

Follow this and additional works at: https://scholarworks.uni.edu/etd

Part of the Social Psychology Commons

Let us know how access to this document benefits you

Recommended Citation

Philipp, Michael C., "Sticks, stones, & stigmas: A social-cognitive account of stereotype threat mechanisms" (2005). Dissertations and Theses @ UNI. 585.
https://scholarworks.uni.edu/etd/585

This Open Access Thesis is brought to you for free and open access by the Student Work at UNI ScholarWorks. It has been accepted for inclusion in Dissertations and Theses @ UNI by an authorized administrator of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.
STICKS, STONES, & STIGMAS: A SOCIAL-COGNITIVE ACCOUNT OF STEREOTYPE THREAT MECHANISMS

An Abstract of a Thesis
Submitted
In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Michael C. Philipp
University of Northern Iowa
July 2005
ABSTRACT

Stereotype threat effects have been a popular domain of much psychological inquiry over the past decade. A number of psychological dispositions (e.g., high levels of stigma consciousness, high levels of social dominance orientation, and high levels of domain identification), situational factors (e.g., out-group presence and task difficulty), and physiological characteristics (e.g., levels of circulating testosterone) have been identified as factors that determine one's susceptibility to the performance debilitating effects of negative stereotype activation. Although each of these variables has been found to be important in eliciting underperformance under threat, no attempts have yet investigated the relationships between these variables. This study explains the theoretical mechanisms proposed by previous investigations, proposes a framework in which these many mechanisms might be related, and then tests a portion of the framework to examine the potential connections between these factors. One hundred twenty-five female participants gave pre-manipulation data regarding levels of stigma consciousness, social dominance orientation, math identification, personality dominance, circulating testosterone, and math ability. Following these measures, participants were presented with one of three stereotype activation manipulations consisting of a high threat condition (consisting of a relevant, negative stereotype concerning women and math ability), a control condition (that mentions no stereotype at all), or a low threat condition (consisting of a statement meant to debunk a relevant, negative stereotype concerning women and math ability). Activation of a stigmatizing stereotype slowed response times to incorrectly answered items, and high levels of stigma consciousness enhanced
performance following a message debunking stigmatizing stereotypes. Furthermore, stigma consciousness enhanced performance scores among participants in the low threat condition. No effects of dominance measures (i.e., personality dominance, social dominance orientation, or testosterone) were found. The findings provide some support for the theory that cognitive resources are diminished following stigmatizing stereotypes resulting in a slower performance on stigma-related tasks. The moderating role of stigma consciousness, however, seems to differently affect performance by affecting task ability directly. Implications of these findings are discussed in terms of a social-cognitive theory of stereotype threat phenomena.
STICKS, STONES, & STIGMAS: A SOCIAL-COGNITIVE ACCOUNT

OF STEREOTYPE THREAT MECHANISMS

A Thesis
Submitted
In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Michael C. Philipp
University of Northern Iowa
July 2005
This Study by: Michael C. Philipp

Entitled: Sticks, Stones, & Stigmas: A Social-Cognitive Account of Stereotype Threat

has been approved as meeting the thesis requirement for the

Degree of Master of Arts.

8-26-05
Date

Dr. Helen C. Hatton, Chair, Thesis Committee

7-26-05
Date

Dr. Mary Lus, Thesis Committee Member

8-26-05
Date

Dr. Cathy DeSoto, Thesis Committee Member

8-31-05
Date

Dr. Susan J. Koch, Dean, Graduate College
ACKNOWLEDGMENTS

First and foremost thanks go to Dr. Helen Harton for all of her guidance and insight for the duration of this project. I am thankful for having had the opportunity to work with Dr. Harton during my stay at UNI; I am a better researcher today because of her.

Sincere appreciation goes to the psychology faculty who assisted me during this project. Special thanks go to Dr. Mary Losch for encouraging me to explore new facets of psychological inquiry and serving invaluabley on my thesis committee. I could not have conceived of this project without Dr. Losch's advise and counsel. Thanks also go to Dr. Cathy DeSoto for her insight regarding the theory and pragmatics of psychoendocrinology and her service as a dedicated thesis committee member. Finally, thank you to Dr. Mel Gonnerman whose sage words and congenial conversations made much of the thesis process more bearable.

Huge thanks are also extended to faculty in Chemistry and Biology for their invaluable guidance and assistance with the testosterone assay process. Thank you to Dr. Ira Simet for demystifying the assay process and contributing many resources and much time to my project. Thanks are also extended to Dr. Jim Jurgenson, Dr. Kavita Dhanwada, and Dr. Lisa Belts for sharing the time, lab space, and equipment to make this project possible.

Also invaluable to this project were those who programmed the computer-administered instruments used in the study. A great thanks goes to Dustin Daugherty for helping me derive the program design and for programming the initial version of the
program. Humongous thanks go to Tom Turner who stepped in and fixed the program when it broke, improved the program when it got clunky, and made the program work seamlessly during the data collection process. This project could not have happened without Tom's assistance and suggestions.

Finally, thanks go to all of those who have given selflessly to help out during the thesis project. Thank you to Natalie Jensen who volunteered countless hours assisting with data collection and data entry. Thanks also go to Brad Okdie for both his willingness to talk shop almost any time of day and his equal propensity for constructive procrastination. And, most of all, a huge thanks is extended to Erin Wolfe for her proof reading, vigilance in eliminating unwarranted commas and her attention to my incessant rants and (fewer) raves about the thesis writing process.
TABLE OF CONTENTS

LIST OF TABLES ....................................................................................................... viii

CHAPTER 1. REVIEW OF LITERATURE .................................................................... 1

   Stereotype Threat Phenomena .............................................................................. 1
   Defining Stereotype Threat Phenomena .................................................... 2
   Mechanisms of Stereotype Threat ............................................................ 4
   An Examination of Stereotype Threat Moderators ................................................ 5
   Perception Factors .................................................................................... 7
   Dispositional Factors .............................................................................. 11
   Task-Performance Factors ...................................................................... 15

   Study Scope ....................................................................................................... 18

   Hypotheses ........................................................................................................ 20

CHAPTER 2. METHODOLOGY ................................................................................. 22

   Participants ........................................................................................................ 22
   Procedure ........................................................................................................... 23

   Participant Intake .......................................................................................... 24
   First Saliva Sample ................................................................................ 25
   Modular Arithmetic Tutorial .................................................................. 27
   Computer Administration ....................................................................... 28
   Second Saliva Sample ............................................................................ 32
   Debriefing ......................................................................................................... 32
<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Mean Pre-test Scale Scores</td>
<td>40</td>
</tr>
<tr>
<td>2 Manipulation Checks</td>
<td>43</td>
</tr>
<tr>
<td>3 Post-Manipulation Measures</td>
<td>44</td>
</tr>
<tr>
<td>4 Post-Manipulation Performance Indicator Estimated Means by Treatment (Controlling for Pre-Test Indicator and Math ACT Score)</td>
<td>46</td>
</tr>
<tr>
<td>5 Summary of Regression Analysis for Assessment One Corrected Scores, ACT Math Scores, and Stigma Consciousness Predicting Assessment Two Corrected Scores</td>
<td>48</td>
</tr>
<tr>
<td>6 Summary of Regression Analysis for ACT Math Scores and Stigma Consciousness Predicting Assessment One Performance Indicators</td>
<td>50</td>
</tr>
<tr>
<td>7 Summary of Regression Analysis for ACT Math Scores and Stigma Consciousness Predicting Assessment One Difficult-Item Performance Indicators</td>
<td>50</td>
</tr>
<tr>
<td>8 Post-Manipulation Stigma Consciousness Means by Treatment (Controlling for Pre-Manipulation Stigma Consciousness)</td>
<td>51</td>
</tr>
<tr>
<td>9 Post-Manipulation Means of Social Dominance Orientation and Testosterone Level (Controlling for Pre-Manipulation Levels of Each)</td>
<td>53</td>
</tr>
<tr>
<td>10 Average Within-Cell Correlations Between Stigma Consciousness and Dominance Measures</td>
<td>54</td>
</tr>
<tr>
<td>11 Correlations Between Study Variables</td>
<td>101</td>
</tr>
</tbody>
</table>
CHAPTER 1

REVIEW OF LITERATURE

Stereotype Threat Phenomena

Over the past decade of psychological inquiry, much attention has been paid to performance disparities between different social groups. Issues of assessment validity, racial bias, and gender discrimination have been hot topics among students, educators, and social scientists alike (e.g., Young, 2003). Many tests, such as the Scholastic Aptitude Test (SAT), have reconceptualized test items in an effort to reduce racial and gender disparities (Cloud, 2003). Despite such efforts, however, some people insist that performance disparities will likely persist across social groups (e.g., Zwick, 2002). One popular psychological explanation for the inhibited performance of stigmatized groups is the stereotype threat phenomenon.

Stereotype threat phenomena occur when the activation of stereotypes and stigmas related to a person's group memberships cause the person to perform a task at a diminished level. For example, women, but not men, perform more poorly on a math test when reminded of the stereotype that women are not good at math (e.g., Brown & Josephs, 1999). Similarly, when performing in mixed-ethnicity groups, African American students may under-perform on a task purported to measure intellectual ability, and yet European American students will under-perform on the same task when it is purported to measure athletic ability (Stone, Lynch, Sjomeling, & Darley, 1999). In each of these

...
examples, stereotype threat effects inhibit the performance of individuals whose stigmatized group memberships are made salient.

Although the stereotype threat phenomenon has been documented in settings ranging from the classroom to the workplace, many who purport to investigate it regularly fail to capture the factors that define stereotype threat events. The effect is sometimes over-simplified and misinterpreted as one group being better at particular tasks than another (for an example see Osborne, 2001). Other times research may set out to test stereotype threat effects without manipulating stereotype activation and without measures of math performance (e.g., Ford, Ferguson, Brooks, & Hagadone, 2004). Discussions of stereotype threat effects must recognize the key features of the phenomenon in order to properly attribute findings to a true stereotype threat situation. The three most basic features of stereotype threat are that (a) it only affects the stereotyped group, (b) it leads to deficits in performance, and (c) it is mediated by very active and conscious cognitive processes.

**Defining Stereotype Threat Phenomena**

The first characteristic that distinguishes stereotype threat phenomena from other stereotype activation effects is the notion that stereotype threat phenomena are solely concerned with effects that occur in individuals who are referenced by the stereotype (i.e., self-stereotype activation; Wheeler & Petty, 2001). Stereotype activation may certainly have effects on individuals not referenced by the stereotype (Levy, 1996), but such effects do not have the same motivational features that are attributed to stereotype threat effects. Stone and colleagues (1999) demonstrated this in a study with a mixed-
ethnicity group of male participants at a miniature golf course. Participants were asked to complete the course in the fewest number of shots. European American putters performed worse only when the stereotype of athleticism was made accessible (i.e., participants were told that putting ability was indicative of athletic ability). African American putters, on the other hand, performed worse only when the stereotype of intellectual ability was made accessible (i.e., participants were told that putting ability was indicative of intellectual ability). Stereotype threat phenomena only occur when individuals believe that their group membership is referenced by a stereotype.

The second defining characteristic of stereotype threat effects is that it is solely concerned with explaining why stereotyped group members perform worse on a task following the activation of a negative stereotype relevant to the task (Steele, 1997). Although many studies have revealed that the activation of a positive stereotype may enhance performance, such findings occur under conditions atypical of stereotype threat phenomena and are generally not attributed to stereotype threat mechanisms (e.g., Levy, 1996; Shih, Pittinsky, & Ambady, 1999; Wheeler & Petty, 2001).

Perhaps the key characteristic that differentiates stereotype threat from other stereotype activation effects is that stereotype threat phenomena are strongly mediated by very active motivational mechanisms (i.e., hot motivational mechanisms) that distract affected individuals from the task at hand by affecting their feelings and motivations. Although more automatic mechanisms (i.e., cold motivational mechanisms) may have substantial effects on behavior (Greenwald & Banaji, 1995), stereotype threat theory explicitly states that the negative self-stereotype activation leads to motivational
mechanisms that mediate negative effects on performance among the referenced group members (Wheeler & Petty, 2001).

**Mechanisms of Stereotype Threat**

Initially, Steele and Aronson (1995) proposed that the motivational mechanism that led to stereotype threat effects stemmed from feelings of anxiety over confirming negative expectations. Empirical analyses, however, demonstrate that even among those researchers who most promote the anxiety-moderator hypothesis (e.g., Spencer, Steele, & Quinn, 1999), there is no clear evidence that anxiety moderates reduced performance under stereotype threat conditions (Schmader & Johns, 2003; Spencer et al., 1999). Instead, the cognitive mechanism responsible for eliciting stereotype threat seems to be a reduction of working memory capacity (Quinn & Spencer, 2001; Schmader & Johns, 2003).

Stereotype threat conditions appear to induce threatened individuals to think about performing faster and trying harder at the task (Croizet, Despres, Gauzins, Huguet, Leyens, & Meot, 2004), and this increased cognitive activity may, ironically, deplete the pool of cognitive resources available for completion of the task (Ashcraft, 1998; Ashcraft & Kirk, 2001). Examining performance latencies illustrates the cognitive load effect. An example lies in a study investigating European Americans' racial attitudes through an Implicit Association Test (IAT) that displayed White and Black faces paired with positive and negative words (Frantz, Cuddy, Burnett, Ray, & Harst, 2004). Stereotype threat was activated in half of the participants by reminding them that White people are more racially prejudiced than other ethnicities. Threatened participants were just as successful
in pairing stimuli with the provided exemplars. However, compared to control participants, threatened participants showed longer response latencies when pairing Black faces with positive words—a sign hypothesized to be a significant predictor of racial prejudice. The stereotype activation did not decrease the "correctness" of responses; it simply led participants to take longer to evaluate counter-stereotypic stimuli. These findings demonstrate that very active cognitive mechanisms are triggered by stereotype threat stimuli. These cognitive mechanisms do not directly hinder performance accuracy but instead cause a performance slow-down when confronted with stereotype-relevant tasks.

An Examination of Stereotype Threat Moderators

The present study is most concerned with evaluating the relationships between some of the more popular factors proposed to moderate the effects of negative stereotype activation and proposing a theoretical model that elucidates the mechanism by which these factors may affect psychological dispositions and task performance. With this in mind, an organizational framework is proposed that draws together various stereotype threat findings and explains mechanisms by which many of these disparate findings might be related.

Stereotype threat researchers have discovered numerous variables that moderate threat effects, including gender identification (Ambady, Paik, Steele, Owens-Smith, & Mitchell, 2004, Schmader, 2002; Shih et al., 1999), domain identification (Aronson, Lustina, Good, Keough, Steele, & Brown, 1999), stereotype awareness (Good, Aronson & Harder, 2000), stigma consciousness (Brown & Pinel, 2003), task diagnosticity
(Croizet & Claire, 1998), task difficulty (Spencer et al., 1999), out-group presence (Sloan, Glenn, & Craig, 2004), social dominance (Philipp & Harton, 2005), and baseline testosterone levels (Josephs, Newman, Brown, & Beer, 2003). Those who extensively review the stereotype threat literature (e.g., Smith, 2004; Steele, Spencer, & Aronson, 2002; Wheeler & Petty, 2001) note the many various findings related to stereotype threat phenomena and work toward differentiating stereotype threat effects from other stereotype activation effects. Yet these reviews pose more questions than answers as to the mechanisms of stereotype threat phenomena. Some current manifestations of stereotype threat theory still posit that anxiety is a prime mediator of stereotype threat effects and at the same time acknowledge that "anxiety may mediate stereotype threat effects only under specific conditions" (Steele et al., 2002, p. 400). Such statements illustrate that stereotype threat mechanisms are still not fully understood. However, a close examination of the many factors reviewed in the literature begins to reveal potential common causal factors (e.g., testosterone and social dominance) and hierarchical arrangements (e.g., gender identification, domain identification, and stereotype awareness all contribute to stigma consciousness) that may provide a foundation for a social cognitive theory of stereotype threat phenomena.

A close examination of the many moderators identified in the stereotype threat literature reveals three types of factors important to the manifestation of stereotype threat effects: Perception factors, dispositional factors, and task-performance factors. Together these define the psychologically necessary characteristics that precede stereotype threat phenomena.
Perception Factors

Fundamentally, an individual must be attuned to the stigmas associated with a particular group membership in order to suffer from stereotype threat effects. That is, a person does not need to believe the stereotype is true to be affected: He or she must simply perceive that others use the stereotype when judging his or her performance. When a person perceives that a stigmatized or "spoiled" (Goffman, 1963) identity is being used to judge his or her performance, that person is said to be conscious of the stigmas associated with membership in the group, or simply "stigma conscious" (Pinel, 1999). Individuals higher in stigma consciousness tend to have higher levels of self consciousness, conform more to gender roles, and exhibit less interpersonal trust than those lower in stigma consciousness (Pinel, 1999).

Levels of stigma consciousness vary among individuals, although cues within an environment may heighten the salience of their stigmatized status (Brown & Pinel, 2003; Pinel, 1999). Regardless of whether a person is naturally high in stigma consciousness or a particular situation promotes stigma awareness, the presence or absence of stigma consciousness is determined by the presence of four critical factors: (a) stereotype plausibility, (b) stereotype-implied negativity (stigma), (c) stereotyped-group identification, and (d) stereotyped-domain relevance to the task at hand.

Stereotype plausibility. Stigmatized group members must feel that the stereotype about their group is credible and that it is in some manner socially established in order to stimulate stigma consciousness (Brown & Pinel, 2003). Explicitly activated stereotypes that are unfamiliar or counter-intuitive (e.g., "Men are bad at math") will not heighten
stigma consciousness because the proposed "stigmatized" group will not believe that others will use the stereotype to make attributions about their performances. However, even previously unknown stereotypes may be perceived as plausible to the extent that they reflect culturally implicit social-roles or expectations. For example, a stereotype that "community college students perform worse on standardized tests than university students" seems plausible to most people. Certainly not everyone would believe the stereotype is true, but it seems likely that some people may carry such a stereotype because cultural expectations promote the notion that university students are intellectually "brighter" than community college students. If a group of community college students was informed of this stereotype they too might find it plausible. However, the potential stereotype "university students perform worse on standardized tests than community college students" is far less plausible; there is simply no social expectation or mechanism that can explain how such a stereotype might be true.

The notion of stereotype plausibility is important when considering how to reduce high levels of stigma consciousness. Highly stigma conscious people often self-activate stereotypes that are relevant to a particular situation without any explicit cues (Brown & Pinel, 2003; Pinel, 1999). However, because plausibility is essential in maintaining stigma consciousness, making a chronically-accessible stereotype seem less plausible reduces the negative effects in stigmatized individuals. This plausibility reduction is often accomplished by either explicitly refuting the stereotype's claim (Blascovich, Spencer, Quinn, & Steele, 2001) or by informing students of how stereotype threat effects function (Schmader & Johns, 2003). Either way, by explicitly debunking the stereotype myth,
researchers have found a way to eliminate the stereotype threat effect. As of yet though, no research has carefully examined whether baseline levels of stigma consciousness are affected by such stigma-inoculations.

**Stereotype-implied negativity.** A given stereotype must be perceived to denote an undesirable characteristic of the group to stimulate stigma consciousness. Stigma consciousness rises only to the extent that the group member believes that the stigma soils his or her own identity. It is for this reason that domain identification has been a popularly studied moderator of stereotype threat effects (Aronson et al., 1999). Those naturally identified with the domain of the stereotype threat will perceive a domain-related stigma as an undesirable, negative characteristic. Interestingly, however, stereotype threat effects have also been found to occur among those low in domain identification (Philipp & Harton, 2005). The apparent importance of domain identification in stimulating stereotype threat effects may lie in the procedural knowledge and innate motivation possessed by those highly identified with a particular domain—that is, people who are adept within a domain are those who identify most with that domain. One who does not strongly identify with the domain may still perceive a stereotype as an undesirable stigma so long as he or she is otherwise motivated to perform a domain-related task. The stigmatized domain must simply be a domain in which a person is motivated to participate. With this in mind, there is still likely no better way to control for the perception of stereotype-implied negativity than to ensure that participants identify with the disparaged domain of the stereotype.
Stereotyped-group identification. Before people become stigma conscious, they must perceive that the stereotype references a group to which they belong. Many group memberships are chronically accessible: People's gender memberships, for example, are usually very ingrained into their self-concept. People automatically behave in accordance with the social expectations of gender membership. Attire, for example, is largely determined by which gender a person identifies. On other occasions, however, people may identify more with a particular group because a contrasting group is present (Stets & Burke, 2000). One may highly identify him or herself as "American" when surrounded by foreigners. Yet, many times people identify more with a particular group simply because they are explicitly reminded of their membership in that group (e.g., "As a parent, wouldn't you like to know if a sex-offender lived next door?").

If a person does not identify with a stereotyped group, the stereotype is irrelevant and there is no need to be concerned with a stigma. Certainly, if an individual is not a member of the stigmatized group or, especially, is a member of a contrasting group, that person's stigma consciousness will be unaffected by the stereotype (Steele et al., 2002). Similarly, people may not identify with a group to which they belong because the group is perceived as irrelevant to them or they are unaware of their membership in the group. The group membership must be meaningful and salient to the individual before he or she begins to highly identify with that group. Even in the case of gender, research has demonstrated that men (non-stigmatized outgroup members) must be immediately present for women to underperform in math-related stereotype threat conditions (Inzlicht & Ben-Zeev, 2003), and European-Americans must be present to evoke intelligence-
related stereotype threat effects from African-American students (Sloan et al., 2004). Without an outgroup presence, the stigmatized group membership is less salient and stereotype activation has no effect.

**Stereotyped-domain relevance.** Before a person's stigma consciousness can rise in a particular context, that person must perceive that the disparaged domain mentioned in the stereotype is relevant to the task at hand. The impending task must be perceived to be diagnostic of the domain specified in the stereotype. In many stereotype threat contexts the relevance of the stereotype is plainly obvious to the task (e.g., a math task is relevant to the stereotype that women do more poorly in tests of math than men). Such cases generally require little explanation for how the domain and task are related.

In other cases, the link between the behavioral domain and the stereotype is more ambiguous and requires an explicit connection. In the case of the miniature golf example (Stone et al., 1999), a convincing connection had to be made to explain how golf related to athletic prowess (a stereotype threat to European American putters) and how golf related to intellectual prowess (a stereotype threat to African American putters). If such an explicit connection between domain and task is not apparent, it becomes impossible to know whether or not the individual considers the stereotype relevant to the task at hand.

**Dispositional Factors**

The extent to which a person has a more dominant disposition may also moderate negative effects of stereotype activation. The rationalization generally offered for this moderating effect is that dominant individuals are very concerned about their status in social contexts and when that high status is challenged (e.g., by a relevant stigma) those
highly dominant individuals begin to ruminate about their status and will be unable to perform other cognitively taxing tasks well (Josephs et al., 2003).

Initial findings regarding the role of dominance moderating stereotype threat effects were reported by Josephs et al. (2003). In this study, testosterone samples from saliva were taken from male and female participants prior to administering a stereotype threat prime. Following the saliva sampling, participants, tested alone, completed a questionnaire containing items that either primed stereotype threat (e.g., "I think that some people feel I have less math ability because of my gender") or did not prime any stereotype (e.g., "School can be very rewarding"). Participants were then given twenty minutes to complete written quantitative Graduate Records Exam (GRE) problems. High testosterone (using a median split) women in the stereotype threat condition underperformed on the GRE problems relative to the high testosterone women in the control condition. Additionally, the stereotype threat condition did not seem to at all affect low-testosterone women. Although overall men outperformed women, men's testosterone levels did not affect performance nor were men's scores affected by the stereotype threat prime. These findings were first to suggest that the cognitive processes that lead to stereotype threat phenomena might be moderated by concerns related to status and social hierarchies.

In order to further explore the role of dominance related to stereotype threat effects, Philipp and Harton (2005) examined whether social dominance orientation moderated stereotype threat effects in a manner similar to testosterone. In this study, male and female participants were run in groups of ten to twenty. First participants completed
the social dominance orientation questionnaire. Next, participants were told that the purpose of the study was either to investigate why men outperform women on tests of math ability (the stereotype threat treatment) or to investigate collegiate math assessment techniques (the control treatment). Following the treatment, participants completed math identification items and a math assessment. Hierarchical regression was used to evaluate both the main effect and interaction effects of social dominance orientation on math performance. Similar to testosterone findings, the analyses showed that social dominance orientation negatively predicted performance among threatened female participants and positively predicted performance among control female participants. Social dominance orientation did not differently affect male performance across the treatments. These findings support suggestions that stereotype threat effects are moderated by cognitive activities of more dominant individuals.

Many different methods exist for assessing dominance-related characteristics. Self-report measures of how accepting a person is of social hierarchies and social inequities (i.e., Social Dominance Orientation; Pratto, Sidanius, Stallworth, & Malle, 1994) tap dominance with respect to how accepting a person is of dominant social hierarchies without measuring how dominant the individual is personally. Other self-report measures gauge the extent to which a person feels more or less submissive on a daily basis (i.e., Simple Adjective Test; Grant, 1992; Grant & France, 2001), thus tapping personality dominance and also correlating highly with biological measures of dominance (i.e., testosterone). Some behavioral measures of dominance include peer ratings of toughness and aggression in men (Dabbs, Frady, Carr, & Besch, 1987) and the absence of
smiling and more sexual partners in women (Cashdan, 1995). Although these measures of dominance assess discrete characteristics of the individual, the endurance of these traits and the correlations between these beliefs and behaviors can often be traced to the testosterone levels of the person in question (for examples see Cashdan, 1995; Dabbs et al., 1987; Grant, 1992).

Levels of testosterone have been found to co-vary with differences in personalities and dominant behaviors in both men (Mazur & Booth, 1998) and women (Cashdan, 1995; Grant & France, 2001). Research has demonstrated that the relationship between social behaviors and testosterone is reciprocal. For instance, testosterone levels rise following interactions with the opposite sex (Roney, Mahler, Maestripieri, 2003), successful competitive events (Bernhardt, Dabbs, Fielden, & Lutter, 1998), and changes in social hierarchies (Jeffcoate, Lincoln, Selby, & Herbert, 1986). Although baseline testosterone levels are not predictive of sexual arousal in women (Davis, Davison, Donath, & Bell, 2005), increases in testosterone from baseline levels are positively related to sexual arousal (Dabbs & Mohammed, 1992; Roney, Mahler & Maestripieri, 2003), and baseline levels have been found to co-vary with status-seeking behaviors (Purifoy & Koopmans, 1979) and ratings of toughness in men (Dabbs et al., 1987). Such relationships, however, are attenuated compared to those in non-human species (Brook, Starzyk, & Quinsey, 2000). High levels of testosterone co-occur with dominant behaviors among humans, behaviors oriented toward seeking out social hierarchies, high status within the hierarchy (Jeffcoate et al., 1986), and seeking to identify others' place in that hierarchy (e.g., through aggression or competition).
The behaviors that distinguish dominant individuals may also hint at larger cognitive processes. Higher testosterone people have been found to think more about the world around them and focus on resolving unresolved issues more than their lower testosterone counterparts (Dabbs, Strong & Milun, 1997). Higher testosterone levels may also correspond to greater levels of distraction in tasks that require selective attention (van Honk et al., 1999). These findings suggest that higher testosterone people may differently employ cognitive resources in evaluative settings, resulting in fewer resources to apply to new tasks. This link between testosterone and available cognitive resources may explain why higher testosterone people are more vulnerable to stereotype threat phenomena (Josephs et al., 2003). If high testosterone individuals engage in cognitively demanding tasks (e.g., hierarchy evaluation) during an evaluative task, fewer cognitive resources may remain to devote toward tasks. This may be especially true if the high testosterone individuals perceive status hierarchies within the task setting.

Besides stigma consciousness, testosterone, and social dominance orientation, few other individual difference measures have been found to moderate stereotype threat effects. Some research has examined the role of regulatory focus as a mediator that causes stereotypes to affect performance (Seibt & Forster, 2004), but such findings are unclear as to whether regulatory focus actually moderates stereotype threat performance decrements.

**Task-Performance Factors**

Current syntheses of stereotype threat effects (Steele et al., 2002; Wheeler & Petty, 2001) suggest that threat effects occur most strongly when the task is difficult for
the individual (Spencer et al., 1999) and when the individual is highly identified with the
domain of the stereotype (Aronson et al., 1999). A difficult task is important to eliciting
stereotype threat effects because it demands that the individual commit all available
cognitive resources to the task. The cognitively arresting effects of stereotype threat
phenomena emerge most clearly when people's full cognitive ability is devoted to the
task at hand (Spencer et al., 1999). Although stereotype threat conditions may affect a
person during a simpler task, the effect will likely not noticeably affect quantifiable
measures of performance (e.g., number correct) because the individual's cognitive
capacity is enough to accommodate the task and threat induced decrements. Finer
measures of performance (e.g., response latencies), however, have demonstrated that
negative stereotype activation does affect performance even on very easy tasks (Frantz et
al., 2004).

Task difficulty is often controlled by selecting problems drawn from college-level
standardized tests such as the GRE (e.g., Spencer et al., 1999) or the Graduate
Management Admissions Test (GMAT; e.g., Quinn & Spencer, 2001) and specifically
selecting respondents for whom such tests are difficult (e.g., Quinn & Spencer, 2001). In
order to more fully engage cognitive resources, tasks that require many steps (e.g., math
story-problems) may be used to maximize cognitive processing (Quinn & Spencer,
2001). The maximization of cognitive processing must be balanced with consideration of
the respondent's knowledge-based ability to solve the task. If a task requires knowledge
that is inaccessible to the participant or if the participant is simply unmotivated to apply
his or her effort to a difficult problem, no cognitive resources will be put forth by the
individual and no cognitively-handicapping stereotype threat effects can be observed. Thus, in order to observe stereotype threat effects, highly threat-susceptible individuals must be able and willing to perform the task so that the cognitively taxing effects of stereotype threat can be fully evaluated.

The question of how to control for motivation and ability within a given domain is answered by measuring domain identification. As explained previously, individuals who identify with the domain of the stereotype (which is also the domain of the task) are more likely to dedicate cognitive resources to the task at hand because they likely enjoy the domain and are able to use their expert knowledge to consider methods for approaching and solving the tasks within that domain. Even if motivated to solve a task, a person lacking knowledge of how to approach the task cannot suffer any additional performance deficits since she or he is unable to complete the task in the first place (Canobi, Reeve, & Pattison, 2003). Thus, although it is important for a task to be difficult for stereotype threat effects to emerge, the procedures for solving the task should be accessible for those completing the task. Task knowledge is not the only reason that domain identification is important, though. The second reason that highly domain-identified people dedicate more cognitive resources to a domain-related task is that they are more motivated to engage in domain-related tasks than others for whom the domain is not as important. A person may possess the procedural knowledge to engage in a task, yet that same person may not be motivated to dedicate his or her cognitive resources to the task. For these reasons it is more likely for stereotype threat effects to emerge among those who highly identify with the domain of a negative stereotype.
Study Scope

Many psychological mechanisms have been purported to lead to stereotype threat effects in assessment-related activities. I have organized the various stereotype threat moderators into three fundamental factors that define the necessary psychological conditions for stereotype threat effects to surface. According to my model, a person will suffer stereotype threat effects only when three criteria are met.

First, a person must be high in stigma consciousness (i.e., the person must feel that others use the negative stereotype to make judgments about his or her performance). In the present study, levels of stigma consciousness were measured with Pinel’s (1999) stigma consciousness questionnaire. Additionally, to ensure that the high-threat manipulation was effective in raising levels of stigma consciousness, the stereotype used in this research was a well-known, culturally accepted stereotype (i.e., believable stereotype) that referenced gender membership as the disparaged identity (i.e., relevant stereotype). Alternately, the low-threat manipulation for this study was designed to debunk the “myth” of the stereotype (thereby eliminating believability), consequently reducing stigma consciousness. The control manipulation for this study did not mention the negative stereotype at all; instead it provided a means to evaluate the effects of stigma consciousness on performance. Participants' salience of group membership was enhanced in each condition by recruiting non-stigmatized outgroup participants to be present during the study. The presence of non-stigmatized participants facilitates underperformance among highly stigma conscious participants (Inzlicht & Ben-Zeev, 2003; Sloan et al., 2004)
The second criterion necessary for stereotype threat to affect an individual is that the person has a dominant disposition and is highly attuned to hierarchical and status information in the environment. Because of this chronic attention to status cues, the activation of a negative stereotype should stimulate more cognitive activity in high dominant individuals than in less dominant individuals. In this study, levels of dominance were measured and examined for changes due to the stereotype manipulation. However, there is some question as to what measures can truly distinguish higher and lower dominant individuals.

With this in mind the present study used two paper and pencil measures and one physiological measure of dominance to explore the extent to which these measures are related. Social dominance orientation (Philipp & Harton, 2005) and base-line testosterone levels (Josephs et al., 2003) have both been found to moderate performance-inhibiting effects of negative stereotype activation. Although the moderating effects of personality dominance on stereotype threat has not yet been explicitly researched, personality dominance measures such as the Simple Adjective Test (SAT; Grant, 1994) have been found to correlate highly with testosterone levels. These measures of dominance were examined to evaluate what aspects of dominance trigger stereotype threat effects. Additionally, the study examined the relationship between stigma consciousness and measures of dominance both before and after the experimental manipulation to understand what, if any, changes occured in these measures as a result of stereotype activation.
The third and final criterion that must be met for stereotype activation to inhibit performance is that the performance task must be procedurally easy enough for a person to complete and yet, cognitively taxing to the extent that any other cognitive processes measurably detract from performance on the task. If the task is too difficult, performance may falter due to lack of knowledge—the person will be unable to apply cognitive energies to the task. On the other hand, if a task is not sufficiently taxing on cognitive resources, the inhibitory effects of stereotype threat on cognitive resources will be concealed (as sufficient resources would be available to perform both tasks). This research used a mathematical task (i.e., a modular arithmetic task adapted from that used by McConnell, Beilock, Jellison, Rydell, & Carr, 2004) that was both procedurally straightforward (adding, subtracting, multiplying, and dividing) and still cognitively taxing as the task was both novel (requiring new cognitive strategies for solving the items) and a speed task (eliminating possible ceiling effects). Furthermore, each participant engaged in the task both before and after the experimental manipulation. This pre-test/post-test design allowed stereotype threat effects to be observed within each participant (by comparing performances on the pre-test and post-test) as well as between experimental groups.

Hypotheses

Based on the theoretical model asserted above, I propose the following hypotheses and research questions to be tested:

H1: Stereotype activation and stigma consciousness will moderate the effect of previous task experience on task performance.
H1a: Activation of a negative, relevant stereotype will lead to decreased task performance.

H1b: The debunking of a negative, relevant stereotype will improve task performance but only among more highly stigma conscious individuals.

H1c: Higher levels of stigma consciousness will predict reduced levels of pre-manipulation task performance when controlling for previous domain experience (i.e., ACT scores).

H2: Changes in stereotype awareness (i.e., activation of a negative stereotype or refutation of a negative stereotype) will lead to changes in stigma consciousness.

H3: Dominance (i.e., Testosterone, Social Dominance, and Personality Dominance) will moderate the effects of stereotype threat on performance. Specifically, higher levels of dominance will increase the negative effects of stereotype activation on performance.

RQ1: Do changes in stereotype activation correspond to changes in dominance?

RQ2: Do levels of dominance and stigma consciousness correlate?
CHAPTER 2

METHODOLOGY

Participants

Two hundred ninety-nine female psychology students were recruited to participate in mass testing sessions that included the pre-test for the final study in exchange for course credit. During these sessions participants always first completed the pre-test informed consent (see Appendix A). Following consent, participants completed pre-test measures of stigma consciousness (see Appendix B; Pinel, 1999), social dominance orientation (see Appendix C; Pratto et al., 1994), and math identification (see Appendix D; Smith & White, 2001). After completing the pre-test measures, participants indicated their email address on a separate sheet of paper if they were interested in participating in the follow-up final study. Pre-test participants were invited by email to register for the final study within two weeks of their pre-test participation.

One hundred twenty-five female psychology students (67 freshmen, 32 sophomores, 17 juniors, 8 seniors and 1 graduate student) participated in the final study in exchange for course credit. Participant age was not recorded, although most participants were of traditional college age. Most participants classified themselves as European in descent (115 European-Americans, 4 African-Americans, 2 Asian-Americans, and 4 of unclassified ethnicities). Data from six participants were dropped from further analyses because no pre-test data sets were available to be matched to the final study data. Additionally, two participants were dropped from analyses due to lack of
participation during the study (i.e., reading non-experimental materials and playing with a cell phone during the timed assessments). Unless otherwise noted, analyses were based on a sample of 117 participants. Furthermore, one to two male, college-aged confederates or two to five male, college-aged participants participated in each session of the study to enhance gender salience and reduce suspicion regarding the female-focus of the study. Data of male participants were not retained for analysis.

Procedure

Each session of the final study began at 4:00pm or 4:30pm on a weekday and accommodated up to 22 female participants and up to five male participants or confederates. Sessions were scheduled to begin at this time in order to be consistent with previous research (e.g., Josephs et al., 2003), to control for daily fluctuations in testosterone, and because testosterone levels vary less in the late afternoon than earlier in the day (Dabbs, 1990; Granger, Schwartz, Booth, & Arentz, 1999). Sessions lasted approximately 55 minutes. Sessions took place in a computer lab consisting of 24 computer terminals and an instructor computer connected to a video projector.

Before the sessions began, each computer was loaded with one of three programs. Each program was identical with the exception of one paragraph of instructions that contained both a text and audio operationalization of the stereotype threat treatment (i.e., high threat message, low threat message, or control message). The control condition program was loaded on terminals reserved for male participants and confederates. The remaining 22 computer terminals were randomly assigned to load one of the three programs when logged in. A pair of headphones was connected to each computer, and
next to each monitor were two five-milliliter cryovials labeled with the session number and computer terminal number, two polypropylene funnels, and a paper cup set on top of two napkins.

**Participant Intake**

Participants arriving early to the session remained outside the session room until the room was fully prepared. Male confederates arrived approximately 5 minutes before each session and behaved as participants. Before participants entered the computer lab, a male researcher greeted them and asked them to rinse their mouths before entering the lab; participants were given the option of rinsing at a water fountain or using a paper cup provided by the researcher. Participants were also asked to refrain from eating or drinking until the conclusion of the session.

As participants entered the session room, the researcher attempted to verify each participant's identity and pre-test participation and then assigned the participant a unique identifier number in order to link the participant's pre-test data to her final study data. Unique identifier numbers consisted of a two-digit session number and a two-digit terminal number. Six participants participated in the final study without pre-test data because they insisted that they had participated in the pre-test earlier in the week. No pre-test data were ever located for these six participants, and their final data were not used in analyses.

Each unique identifier number allocated to a participant assigned her to one of the computer terminals in the lab. A female research assistant helped to guide participants to their assigned terminals. As each participant was seated, she was provided with two
copies of the informed consent form and asked to read and sign the form if she consented to participating. The informed consent (see Appendix E) described the study as an evaluation of how daily and seasonal biological changes affect performance on assessments of arithmetic ability. No participants explicitly refused participation in the study.

Following the arrival and consent of all participants, the researcher introduced himself and then introduced the research assistant as a student who was ostensibly performing a pre-test for a different study. The research assistant introduced the Simple Adjective Test (see Appendix F; Grant, 1994) as a survey of personality descriptors. Participants recorded their assigned unique identifier at the top right corner of the survey. When all participants finished, the research assistant retrieved the completed surveys and left the room.

At this point the researcher explained the ostensible nature of the study. Specifically, the researcher explained that daily and seasonal fluctuations in biological chemistry might affect people's abilities to perform well on academic tests, especially tests of arithmetic skill. The researcher further explained that the study was designed to account for possible psychological and biological factors that affect concentration and arithmetic ability.

First Saliva Sample

Following the researcher's explanation of the study, the research assistant returned to the session room. The researcher explained that before beginning the assessment tests, the first saliva sample would need to be collected as an initial biological measure. The
researcher explained that participants should position the narrow end of the provided funnel over the larger opening of the provided vial. Participants were instructed to place the vial on the table in front of them and lean over it placing the side of the funnel between their lower lip and their chin (with their mouth positioned over the hole of the funnel). The researcher demonstrated the position before continuing. The instructions further indicated that participants should not spit into the funnel but only passively drool. Passively drooling simply involves allowing saliva to drip from the bottom lip into the funnel thereby reducing the amount of mucus and other foreign substances in the saliva. Before beginning, participants were told to swallow any saliva currently in their mouths, and the male participants/confederates were asked to accompany the researcher to an adjacent room to reduce any anxiety participants might feel about having the opposite sex present while salivating into vials. After departing the room, the researcher collected the males' vials and explained that male saliva samples were not required. The men were asked to not disclose their lack of salivation to the female participants.

The female research assistant remained in the room to assist participants with any questions or difficulties during the saliva collection stage. Pictures of citrus fruits were displayed on an overhead projector to stimulate salivation. During some sessions the research assistant cut and peeled oranges and lemons to further encourage salivation. Participants' saliva was collected in a polypropylene cryovial. Participants were told to cease their salivation if the saliva level reached the 1ml mark on the vial. The saliva collection lasted approximately four minutes. After four minutes participants capped their cryovials and submitted them to the research assistant. The male participants and the
researcher also returned after four minutes, and the researcher handed the research assistant the male vials. After collecting the vials, the research assistant left the lab and notated each vial as a pre-manipulation sample with a dot from a red permanent marker. Following the session, each sample was stored in deep freeze (-75 degrees Fahrenheit) in order to preserve the saliva sample without affecting the measurable testosterone in the sample (Granger, Shirtcliff, Booth, Kivlinghan, & Schwartz, 2004).

**Modular Arithmetic Tutorial**

After the research assistant left the room, the researcher explained the arithmetic assessments. A brief tutorial of modular arithmetic was given to participants followed by an explanation of how each assessment was structured. The arithmetic problems on each assessment required participants to apply simple arithmetic operations (i.e., addition, subtraction, or multiplication) to successive pairs of numbers. The provided answer derived from each operation was given in terms of a modular notation. Each modular notated answer consisted of a positive integer followed by a modulus (e.g., 3 mod[12]). Assessment items simply asked participants to indicate whether each given equation was correct or incorrect. Although a variety of strategies can be used to assess the correctness of each equation, the easiest method is to calculate the standard answer by apply the given operation to the given pair of numbers, divide the standard answer by the modulus number, and calculate the remainder. If the remainder equals the integer appearing before the modulus, the equation is correct. If the remainder does not equal the integer, the equation is incorrect.

For example, the above strategy can be used to evaluate the equation below.
The sum of 8 and 5 is 13. Dividing the standard answer of 13 by the given modulus of 3 would leave a remainder of 1. Thus, the remainder matches the integer preceding the modulus and the equation is deemed correct.

**Computer Administration**

Following the introduction to modular arithmetic, the researcher explained that the remaining portion of the study would be conducted through computer interface, with the brief exceptions of a final saliva collection stage and a concluding debriefing statement. At this point participants donned the headphones connected to their respective terminals. To ensure that audio portions of the program were audible to participants via the headphones, they were instructed to activate an audio test with the computer mouse. After each participant confirmed that she could hear the audio test, the researcher gave a password that allowed participants to begin the assessment portion of the study.

**Assessment One.** After entering the password, participants were immediately presented with the assessment one instructions in text form on the computer screen and audibly, read by a male voice. The assessment one instructions were the same for each participant and read as follows:

The following assessment was first used by college admissions personnel who were looking for an innovative way to test arithmetic ability. Students' performances on this modular arithmetic task aided academic advisors in knowing how prepared a student was for various math courses. Even though modular arithmetic requires little more than simple math operations, this test was found to be quite predictive of how well a student performed in advanced mathematical
coursework. The questions that follow simply require you to assess whether the equations provided are correct or incorrect. You will be given 5 minutes to complete as many questions as possible. Although the assessment is timed, accuracy is more important than the number of questions you answer. For each equation assessed accurately you will receive 2 points. For each equation assessed incorrectly you will lose 1 point. Please work on these problems to the best of your ability.

When the audible instructions ended, a continue button appeared that, when clicked, began assessment one by presenting the first equation. At the top right corner was a counter that displayed the time remaining for assessment one. Two buttons were available to push depending on whether the equality was correct or incorrect. If the participant deemed the equality correct, she would push the button labeled "Correct." On the other hand, if the equality was deemed incorrect, she would push the "Incorrect" button. Following the activation of either button, the answer and response latency were written to a data file and a new equality was presented to the participant. When the time counter reached zero, the assessment ended and following statement appeared in text only:

Assessment 1 is now complete. When you are ready to proceed to the next assessment, please press the button below.

Threat Treatments. After a few seconds, a button appeared prompting the participant to continue to the assessment two instructions. After pressing the button, assessment two instructions were presented both in text and audibly, read by the same male voice. Three different versions of the instructions were presented depending on the assigned threat treatment. The control treatment instructions are based on those used by Schmader and Johns (2003). In these instructions no mention of gender is made; yet the
purpose of the assessment is vague and the wording does not expressly deny that gender differences may exist. The control condition instructions read as follows:

As you may know, math skills are crucial to performance in many subjects in college, yet surprisingly little is known about how to best assess math ability in a collegiate setting. In part, this second task is designed to help us better understand why some people do better on math assessments than others. The method for answering each question is identical to the previous test—simply indicate whether each equation is correct or incorrect. Your performance on this task will be compared to the scores of students at other institutions, so please complete these items to the very best of your ability.

The instructions depicting the high stereotype threat condition clearly state that males typically out-perform females on the task and that the task is examining why females are inferior to males on tests of arithmetic ability. The high threat condition instructions stated:

As you may know, math skills are crucial to performance in many subjects in college, yet surprisingly little is known about how to best assess math ability in a collegiate setting. As you may also know, a good deal of research indicates that men consistently score higher than women on standardized tests of math, but thus far there is not a good explanation for this difference. Prior use of this next assessment has demonstrated this gender difference—that is, on average men perform better on these problems than women. Thus, this second task is designed to help us better understand the reasons for gender differences on mathematical assessments. The method for answering each question is identical to the previous test; simply indicate whether each equation is correct or incorrect. Your performance on this task will be compared to the scores of students at other institutions, so please complete these items to the very best of your ability.

Finally, the low-threat instructions first dispelled any notions of innate gender differences regarding math ability and secondly expressly stated that the assessment has been found to be gender fair. The low-threat instructions stated:

As you may know, math skills are crucial to performance in many subjects in college, yet surprisingly little is known about how to best assess math ability in a collegiate setting. As you may also know, there has been some controversy about whether there are gender differences in math ability. Previous research in
math assessment has sometimes shown gender differences and sometimes shown no gender difference. Yet, little of this research has been carried out with women and men who are expressly interested in math. You were selected to take this test because of your indicated interests in math. Prior use of this next assessment has shown it to be gender-fair—that is, men and women perform equally well on these problems. In part, this second task is designed to help us better understand why different social groups do better on math assessments than others. The method for answering each question is identical to the previous test; simply indicate whether each equation is correct or incorrect. Your performance on this task will be compared to the scores of students at other institutions, so please complete these items to the very best of your ability.

**Assessment Two.** After the instructions were audibly read to participants, a button appeared that prompted the beginning of assessment two. The equations presented in assessment two were very similar in type and difficulty to those equations presented in assessment one. The only substantive difference between the assessments was that assessment two allowed participants ten minutes to assess as many equations as possible.

**Attitudes Survey.** Following participants’ completion of the second assessment, participants were prompted to continue to a survey of attitudes. The attitudes survey began with items from the Math Identification Questionnaire (Smith & White, 2001), the Social Dominance Orientation questionnaire (Pratto et al., 1994), and the Stigma Consciousness Questionnaire (Pinel, 1999). All items from each scale were presented sequentially, and scales were presented in the order listed above. Each scale's items consisted of a statement to which participants were asked to rate their level of agreement on a scale ranging from *Strongly Disagree* (1) to *Strongly Agree* (9). Following these attitude items, participants responded to questions designed to assess nervousness during the assessments, degree of effort expended on assessment items, comparative difficulty of the assessments, and the extent to which the participant was cognizant of math-related
stereotypes about women. A brief demographics section followed requesting participants to report their gender, ethnicity, year in college, and American College Testing (ACT) scores. Participants were also asked for permission to retrieve their math and composite ACT scores from the University’s Registrar’s Office. Participants gave consent for ACT score retrieval by providing their student ID numbers.

Second Saliva Sample

After all participants completed the computer survey, the female research assistant returned to the lab to administer the second saliva collection. Once again the researcher left the computer lab accompanied by male participants or confederates. The exact procedures used in the first saliva collection were repeated. After collecting the vials, the research assistant left the lab and notated each vial as a post-manipulation sample with a dot from a black permanent marker.

Debriefing

The study concluded with a debriefing statement that the researcher read to the participants. The debriefing read as follows:

Before we complete this study, I would like to provide you with some detailed information about the study you have just completed. In order to provide the most realistic context for studying behavior, it is necessary to set up a scenario that closely matches real-life settings. In the case of this study, the actual interest of this research is to examine the effects of stereotypes on performance. More specifically, this research is examining the factors that cause people to underperform when confronted with a negative stereotype. For this reason some participants were presented with the common stereotype that women are naturally worse at math than men. Other participants in this study were presented with a message explaining that stereotypes about women’s poor math abilities are false and that there is no scientific basis for believing that any sex differences truly exist in regards to math ability. We exposed participants to these different messages to see whether this information would affect math performance.
In reality, it is difficult to determine whether one sex is really naturally better at mathematical operations than the other. You can probably think of instances in which men might perform better on tests of math ability, but there are also instances in which women outperform men. In fact, one major review of female math abilities suggests that on average, females achieve higher grades in math throughout elementary school and middle school—during this time female students exceed males on general computation tasks and on classroom assessments of math ability (Kimball, 1989). As children age, however, cultural pressures gradually encourage women to become less interested in mathematical domains. Also, because of social conditioning, women are more likely to admit a shortcoming regarding math ability. Men, on the other hand, are more reluctant to express shortcomings and instead express their difficulties with math by claiming that could do it they tried harder. By better understanding these tendencies and through understanding the stereotype threat phenomenon, women and other minorities can avoid being negatively affected by its effects.

Before you leave I would like to ask you to be discreet about your participation in this study. The behaviors that we are studying are only displayed in true assessment conditions, so please do not discuss the purpose of this study with any one else. If someone asks you about the study, describe it as a study that is evaluating different methods of mathematical assessment. This study will continue through the Fall and into next Spring; with this in mind, our research team asks that you do not discuss this study with any other students on campus—does that seem reasonable to everyone?

Great -- THANK YOU, we appreciate your participation today. I will gladly answer any questions you have about the study. are now free to go. Thanks, once again, for coming to our study.

Materials

Stigma Consciousness Questionnaire for Women (SCQ-W)

The SCQ for women (Pinel, 1999) is an instrument that measures the extent to which women "interpret their experiences in light of their group membership" (p. 117) (see Appendix B). The instrument consists of ten statements (e.g., "Most men have a problem viewing women as equals") to which participants are asked to rate their reaction on a seven-point scale that ranges from Strongly Disagree (1) to Strongly Agree (7). Seven of the items are reverse-scored, and all are averaged to obtain a stigma
consciousness score. Higher scores represent higher levels of stigma consciousness. The SCQ-W has fair internal reliability ($\alpha = .77$) and an adequate test-retest reliability after one month, $r(42) = .76$ (Pinel, 1999).

In the present study, the pre-test version of the SCQ-W was identical to the original Pinel scale with a similarly adequate reliability ($\alpha = .77$). The final study version of the SCQ-W only differed from the original by changing the rating scale used from a 7-point scale to a 9-point scale. A 9-point scale was used in order to standardize the scales used for measures in the final study and to maximize the potential variance of the scale. Reliability for the final study version of the SCQ-W was, again, adequate ($\alpha = .79$), and it correlated highly with the pre-test measure ($r = .68$).

**Social Dominance Orientation (SDO)**

The SDO (Pratto et al., 1994) is designed to measure a person's preference for social hierarchies and acceptance of social inequalities among social groups (see Appendix D). The basic version of the scale consists of 15 statements (e.g., "Sometimes other groups must be kept in their place") to which participants rate their reaction on a 7-point scale that ranges from *Strongly Disagree* (1) to *Strongly Agree* (7). Half of the items on the scale are reverse scored, and all item scores are averaged to obtain a social dominance score. Higher SDO scores represent higher levels of social dominance orientation. Reliability coefficients of the SDO scale range from $\alpha = .80$ to .89, and tests of validity have distinguished the SDO from other measures of dominance and political orientation (Pratto et al., 1994).
The pre-test used the original version of the SDO and derived adequate reliability ($\alpha = .86$). The final study version of the SDO only slightly changed the rating scale from a 7-point scale to a 9-point scale (see Appendix C). A 9-point scale was used in order to standardize the scales used for measures in the final study and to maximize the potential variance of the scale. Reliability for the final study version was excellent ($\alpha = .93$) and correlated highly with the pre-test scores ($r = .81$).

Math Identification Questionnaire (MIQ)

The MIQ (Smith & White, 2001) is a 20-item scale that measures to what extent a person identifies with the general domain of academics and the specific domains of English and math. Although all questions do not concern math, the measure was originally constructed to assess math identification while maintaining low levels of face validity. The instrument consists of three sections, each composed of a different measurement technique. The first eight items ask participants to rate their level of agreement to domain related questions (e.g., "Mathematics is one of my best subjects") using a 5-point scale that ranges from Strongly Disagree to Strongly Agree. The next ten items ask participants to rate how much each domain-related statement describes them (e.g., "How much do you value being a college student?") using a 5-point scale that ranges from Not At All to Very Much. The final two questions ask participants to rate their level of proficiency in Math and English by comparing themselves to other students (e.g., "Compared to other students, how good are you at math?") using a 5-point scale that ranges from Very Poor to Excellent. Scale scores are calculated by averaging the scores across the relevant scale items. Higher scores indicate higher levels of
identification. Overall the Math scale (α = .93), the English scale (α = .90) and the Academics scale (α = .75) have been demonstrated to be acceptably reliable in addition to demonstrating good convergent validity (Smith & White, 2001).

In the present study, a modified version of the MIQ was used in both the pre-test and the final study to assess math identification (see Appendix D). The first eight items of the MIQ were combined with modified versions of the remaining items and a few newly derived items to create a 22-item scale that asked participants to rate their level of agreement with each statement on a scale from Do Not Agree (1) to Completely Agree (9). This modification of the scale permitted each item to be rated using the same scale. Statements on the adapted version referenced abilities in Math, English, Social Sciences, and Academics in general. The addition of some question regarding social sciences further reduced any suspicion of the study's math-only focus. Only statements referencing math ability were of interest to this study. Similar to the original MIQ, higher scores indicated higher levels of math identification. The math scale demonstrated excellent reliability in both the pre-test (α = .95) and the final study (α = .96). Additionally, participants' scores on the pre-test version and the final study version were highly correlated (r = .91).

Simple Adjective Test (SAT)

The SAT (Grant, 1992) is designed to measure people's feelings of dominance and submissiveness on a personal level. The test consists of 50 words drawn from Russell and Mehrabian's (1977) list of submissive- and dominant-rated words. In this test, participants check the words from the list that affirm the question, "Do you quite often
feel ___?" Each word is assigned a weighting based on the data collected by Russell and Mehrabian. Scores are calculated by summing the weights associated with all words checked by the participant. Higher scores represent higher levels of personality dominance. Although no previous reliability statistics are available for this test, the test has significantly predicted fetus sex in pregnant women (Grant, 1992) and is significantly correlated with serum testosterone levels in women (Grant & France, 2001). For this reason, the test was included as an indicator of dominance.

In the present study, the SAT was employed only as a pre-test measure immediately preceding assessment one in the final study (see Appendix F). The version of the SAT used in this study employed 61 words from Russell and Mehrabian’s (1977) original list and weighted each word according to its calculated dominance weight. According to this system, adjectives rated as connoting more dominance are assigned larger, positive numbers. On the other hand, those adjectives rated as connoting more submissive personalities are assigned larger negative numbers. Dominantly “neutral” words are represented by numbers closer to zero. Reliability for the SAT was adequate (α = .82).

Salivary Assay

The testosterone level of saliva samples was measured using Salimetrics's (State College, Pennsylvania) salivary testosterone enzyme immunoassay kits. Each kit is capable of testing 80 samples of saliva. Only 91 of the 117 final study participants provided enough saliva for testosterone testing. The 80 participants with the highest math
identification scores (MIQ pre-test scores > .9) were tested\(^1\). During testing, one saliva sample became inadequate for testing due to the large amount of particulate matter in the sample. Thus, the pre- and post-manipulation saliva samples of 79 participants were tested in duplicate.

The author conducted each of the saliva assays. The University of Northern Iowa Departments of Chemistry and Biology provided supplementary testing materials and laboratory space for the assay procedures. An enzyme-linked immunosorbent assay (ELISA) was used to measure the testosterone concentrations. Specifically, 50 μL of each saliva sample (and the serially diluted testosterone standards) were measured into wells precoated with testosterone antibodies. For each assay plate, two uncoated non-specific binding (NSB) wells served to measure the small fraction of antigen that nonspecifically binds in the absence of antibodies (Paul, 1999). The NSB wells and two zero standard wells were each filled with 50 μL of assay diluent. Next, 150 μL of an enzyme conjugate was added to each well. The enzyme conjugate molecules (consisting of testosterone linked with horseradish peroxidase) competed for antibody binding sites with testosterone in the standards and unknowns. Following an incubation period, all unbound materials are washed from each well. The concentration of the enzyme conjugate was then measured by adding a solution (tetramethylbenzidine; TMB) that binds to the enzyme conjugate. The TMB, then, produces a blue hue that indicates the amount of enzyme conjugate bound to antibodies in the well. Because the enzyme conjugate directly

\(^1\) Due to budget constraints, only eighty participants' saliva samples could be tested.
competes with testosterone for the antibody binding sites, the amount of testosterone present in the sample is inversely proportional to the density of color in each well (Salimetrics, 2005).

In order to arrive at a standardized concentration score within each plate, the optical densities of duplicate samples were averaged together and the average optical density of the NSB wells were subtracted from each well (to correct for non-specific binding of the enzyme conjugate). Then the percentage of antibodies bound to enzyme conjugate was calculated by dividing the optical density of each standard and sample by the average optical density of the zero standard wells. The resulting score indicated the percentage of antibody binding sites occupied by enzyme conjugate (B/B0). This percentage is inversely proportional to the concentration of testosterone in each sample. These assay procedures were conducted in accordance with the instructions provided by Salimetrics (2005; see Appendix H).
CHAPTER 3

RESULTS

Pre-test

The mean score for all pre-test participants is provided in the last row of Table 1.

Pre-test score comparisons between those who participated only in the pre-test and those who participated in both the pre-test and the final study demonstrate that math identification was higher among final study participants than pre-test only participants, $t(298) = 2.16, p = .03, d = .03$.

Table 1

Mean Pre-test Scale Scores

<table>
<thead>
<tr>
<th></th>
<th>Math Identification$^*$</th>
<th>Social Dominance Orientation*</th>
<th>Stigma Consciousness#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Pretest-Only Participants ($n=183$)</td>
<td>4.36$a$</td>
<td>2.35</td>
<td>2.49$a$</td>
</tr>
<tr>
<td>Final Study Participants ($n=117$)</td>
<td>4.95$b$</td>
<td>2.25</td>
<td>2.34$a$</td>
</tr>
<tr>
<td>All Pretest Participants ($n=300$)</td>
<td>4.59</td>
<td>2.32</td>
<td>2.43</td>
</tr>
</tbody>
</table>

Note. Mean scores between Pretest-Only Participants and Final Study Participants were compared using $t$-test analyses. Pre-test and final study means in the same column that do not share a subscript differ at $p < .05$.

$^*$ Scale ranged from 0 to 9; larger numbers indicate higher levels of Math Identification.

* Scale ranged from 1 to 7; larger numbers indicate higher levels of Social Dominance Orientation.

# Scale ranged from 1 to 7; higher numbers indicate higher levels of Stigma Consciousness.
Stigma consciousness and social dominance orientation did not differ between the two groups. This finding demonstrates that final study participants were similar to the pre-test only participants except on math identification.

Final Study

Data Preparation

Four performance indicators were calculated for each assessment. A corrected score for each assessment was calculated by adding the number of correct items and then subtracting one-half of the incorrect items. To represent the speed of item completion, the number of items answered per minute was also calculated. Finally, latency scores for both correct and incorrect items were recorded and averaged to derive an average latency for correct items and an average latency for incorrect items.

As mentioned previously, testosterone levels were estimated through a salivary testosterone enzyme-linked immunosorbent assay that resulted in an inverse measure of testosterone concentration in the saliva samples. Although the ranges of testosterone concentrations differed slightly between the four test kits used, a comparison of the sample concentrations to the known concentrations of the serially diluted testosterone standards on each plate verified that testosterone levels of all samples were within expected ranges. Following the assay process, the B/B0 concentrations were divided by the concentrations of a known higher-testosterone standard within the same plate so that samples would be more comparable across assay plates. In order to represent higher testosterone levels as higher values, the concentration of each sample was then subtracted
from 10 to derive a relative testosterone level. The resulting scores ranged from 7.90 to 9.93. All analyses involving testosterone in this study use the aforementioned relative testosterone scores.

Pre-Manipulation Measures

There were no differences between treatment groups on pre-manipulation measures of stigma consciousness, social dominance orientation, math identification, personality dominance (i.e., the Simple Adjective Test), testosterone level, math assessment performance, or ACT math score ($p$'s > .05). This lack of significant differences suggests that any post-manipulation differences are due to stereotype activation treatments.

Manipulation Checks

Three questions assessed whether participants' beliefs in gender-related math stereotypes and concern for gender stigmatization were affected by the three stereotype activation treatments. Responses to the question, "To what extent do you believe that one sex is better at math than the other" varied by treatment condition, $F(2, 117) = 3.64$, $p = .03$, $\eta^2 = .06$. The low threat group rated females as relatively better at math than both the high threat group, $t(72) = 2.41$, $p = .02$, $d = .56$, and the control group, $t(72) = 2.80$, $p < .01$, $d = .65$ (see Table 2). Levels of agreement with the two additional statements: "I am concerned that the researcher will judge women, as a whole, based on my performance on this test" and "The researcher will think that women, as a whole, have less math ability if I did not do well on this test," did not vary significantly across treatments ($p$'s > .05).
Table 2

**Manipulation Checks**

<table>
<thead>
<tr>
<th>Stereotype Treatment</th>
<th>I am concerned that the researcher will judge women based on my performance on this test.(^\wedge)</th>
<th>The researcher will think that women have less math ability if I did not do well on this test.(^\wedge)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>High Threat</td>
<td>4.46(^b)</td>
<td>.95</td>
</tr>
<tr>
<td>(n=41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>4.42(^b)</td>
<td>.82</td>
</tr>
<tr>
<td>(n=43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Threat</td>
<td>4.91(^a)</td>
<td>.63</td>
</tr>
<tr>
<td>(n=33)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Scale ranged from 1 (men are better) to 9 (women are better).
\(^\wedge\) Scale ranged from 1 (strongly disagree) to 9 (strongly agree).

The one significant manipulation indicator suggests that the low threat participants were cognizant of the low-threat manipulation message of ambiguity regarding innate differences in math ability between men and women. The lack of difference between the high threat and control groups may be due to the subtler wording of the high threat condition that implied that men's math scores, not necessarily math abilities, were higher than women's for undetermined reasons.

The lack of significant differences on the remaining two manipulation check
variables does not necessarily indicate that the stereotype activation treatments were
unsuccessful, but merely demonstrates that participants' reported concern for the
researcher's judgment was unaffected by the treatments.

Post-Manipulation Measures

Three additional questions further explored differences in participants'
perceptions of the post-manipulation assessment. The stereotype treatments had no main
effects on participants' nervousness, $F(2, 117) = 1.75, p = .18, \eta^2 = .03$, effort, $F(2, 117) = 1.50, p = .23, \eta^2 = .03$, or experienced difficulty, $F(2, 117) = 2.51, p = .09, \eta^2 = .04$
(see Table 3).

Table 3

<table>
<thead>
<tr>
<th>Stereotype Treatment</th>
<th>How nervous were you during the math performance task?*</th>
<th>How would you rate the amount of effort you put into completing the math performance task?^</th>
<th>Compared to the first math assessment, how difficult was the second math assessment?#</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Threat ($n=41$)</td>
<td>4.46, 2.31</td>
<td>7.58, 1.26</td>
<td>4.61, 1.56</td>
</tr>
<tr>
<td>Control ($n=43$)</td>
<td>5.37, 2.37</td>
<td>7.07, 1.74</td>
<td>5.21, 1.21</td>
</tr>
<tr>
<td>Low Threat ($n=33$)</td>
<td>4.54, 2.67</td>
<td>7.09, 1.46</td>
<td>5.09, .95</td>
</tr>
</tbody>
</table>

* Scale ranged from 1 (not nervous at all) to 9 (very nervous).
^ Scale ranged from 1 (no effort at all) to 9 (as much effort as I could).
# Scale ranged from 1 (much easier) to 9 (much more difficult).
Hypothesis Testing

Hypothesis 1a

ANCOVA analyses examining differences in each assessment two performance indicator controlling for the corresponding assessment one performance indicator and ACT math scores showed that the only main effect of the four treatment conditions that approached statistical significance was the average length of time participants took before answering a question incorrectly, $F(2, 95) = 2.94, p = .06, \eta^2 = .06$. Notably, high threat participants spent about 17 seconds with each incorrectly answered item, whereas low threat and control participants spent less than 15 seconds with each incorrectly answered item. None of the remaining performance indicators (i.e., corrected score, number of items answered per minute, correct answer latency) varied significantly by treatment condition ($p's > .3, \eta^2's < .03$; see Table 4).

In order to further test the hypothesis, new performance scores were calculated for each assessment using only "difficult" items that were correctly answered by 85 percent or fewer of participants who attempted them. Calculating these difficult item performance scores permits an analysis of performance under more cognitively taxing conditions (i.e., solving more difficult problems). In assessment one, 18 difficult items were identified. In assessment two, 22 difficult items were identified. Using the same scoring technique mentioned previously, a corrected score was derived for each participant using only these difficult items. Furthermore, the average latency for correctly answered difficult items and the average latency for incorrectly answered difficult items were calculated.
Table 4

Post-manipulation Performance Indicator Estimated Means by Treatment (Controlling for Pre-Test Indicator and Math ACT Score)

<table>
<thead>
<tr>
<th></th>
<th>Performance Indicators</th>
<th>Difficult Item Performance Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>( SE )</td>
</tr>
<tr>
<td>Hi Thrt ( n=41 )</td>
<td>43.04a</td>
<td>1.75</td>
</tr>
<tr>
<td>Corrected Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td># correct - (# wrong / 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control ( n=43 )</td>
<td>44.26a</td>
<td>1.72</td>
</tr>
<tr>
<td>Lo Thrt ( n=33 )</td>
<td>43.32a</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>Hi Thrt ( n=41 )</td>
<td>5.22</td>
</tr>
<tr>
<td></td>
<td>Control ( n=43 )</td>
<td>5.54</td>
</tr>
<tr>
<td></td>
<td>Lo Thrt ( n=33 )</td>
<td>5.35</td>
</tr>
<tr>
<td>Average Time Spent</td>
<td>Hi Thrt ( n=41 )</td>
<td>11.95a</td>
</tr>
<tr>
<td>with Correctly</td>
<td>Control ( n=43 )</td>
<td>11.32a</td>
</tr>
<tr>
<td>Answered Items</td>
<td>Lo Thrt ( n=33 )</td>
<td>11.84a</td>
</tr>
<tr>
<td>(in seconds)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Time Spent</td>
<td>Hi Thrt ( n=35 )</td>
<td>17.35a</td>
</tr>
<tr>
<td>with Incorrectly</td>
<td>Control ( n=43 )</td>
<td>14.33b</td>
</tr>
<tr>
<td>Answered Items</td>
<td>Lo Thrt ( n=29 )</td>
<td>14.35b</td>
</tr>
<tr>
<td>(in seconds)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Pairs of assessment indicators were compared using parameter estimates. Indicators in the same column that do not share a subscript differ at \( p < .05 \).

* Items answered per minute was not calculated in difficult item analysis.
Again, corrected assessment scores were unaffected by stereotype treatments. However, the average latency for correctly answered items was significantly affected by treatment condition, $F(2, 107) = 5.75, p = .004, \eta^2 = .10$; low threat and high threat groups took over 15 seconds, on average, to correctly answer difficult items, and the control group spent just over 13 seconds to correctly answer difficult items. Furthermore, the data suggest that the average latency for incorrectly answered difficult items may have also been affected by the treatments, $F(2, 92) = 2.43, p = .09, \eta^2 = .05$, resulting in incorrect difficult item latencies of over 20 seconds for high threat participants and incorrect difficult item latencies of less than 17 seconds for control and low threat participants (see Table 4). No other difficult item performance indicators varied significantly by treatment condition ($p$'s > .3, $\eta^2 < .03$).

**Hypothesis 1b**

Because this hypothesis predicts that stigma consciousness affects only low threat participants (i.e., those for whom the negative stereotype was exposed as an exaggerated claim), a hierarchical regression model separately analyzed each treatment's performance indicators first entering the corresponding assessment one performance indicator, then entering math ACT scores, and finally entering stigma consciousness as predictors (for correlations, see Appendix I).

Specifically, stigma consciousness improved assessment two corrected scores among low threat participants $F(1, 28) = 9.65, p = .004$, but did not positively influence corrected scores among high threat or control participants (see Table 5).
Table 5

Summary of Regression Analysis for Assessment One Corrected Scores, ACT Math Scores, and Stigma Consciousness Predicting Assessment Two Corrected Scores

<table>
<thead>
<tr>
<th>Model</th>
<th>Corrected Score</th>
<th>Difficult Item Corrected Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
</tr>
<tr>
<td><strong>High Threat</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R² = .80 &amp; .65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assmnt One Corr. Score</td>
<td>1.70</td>
<td>.21</td>
</tr>
<tr>
<td>ACT Math Score</td>
<td>.78</td>
<td>.40</td>
</tr>
<tr>
<td>Stigma Consciousness</td>
<td>-3.16</td>
<td>1.81</td>
</tr>
<tr>
<td>R² Change Due to Stigma Con.</td>
<td></td>
<td>.02^</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R² = .66 &amp; .57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assmnt One Corr. Score</td>
<td>2.05</td>
<td>.29</td>
</tr>
<tr>
<td>ACT Math Score</td>
<td>-.36</td>
<td>.61</td>
</tr>
<tr>
<td>Stigma Consciousness</td>
<td>1.18</td>
<td>2.07</td>
</tr>
<tr>
<td>R² Change Due to Stigma Con.</td>
<td></td>
<td>.003</td>
</tr>
<tr>
<td><strong>Low Threat</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R² = .71 &amp; .61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assmnt One Corr. Score</td>
<td>1.50</td>
<td>.26</td>
</tr>
<tr>
<td>ACT Math Score</td>
<td>.42</td>
<td>.78</td>
</tr>
<tr>
<td>Stigma Consciousness</td>
<td>7.46</td>
<td>2.40</td>
</tr>
<tr>
<td>R² Change Due to Stigma Con.</td>
<td></td>
<td>.10**</td>
</tr>
</tbody>
</table>

Note. ^ p < .10  * p < .05  ** p < .01
The same effect of stigma consciousness emerges when examining the assessment two difficult item corrected scores, $F(1, 28) = 7.89, p = .009$. Stigma consciousness did not significantly predict other assessment two performance indicators or difficult item performance indicators ($\beta$'s > .1).

**Hypothesis 1c**

Hierarchical regression analyses examining the impact of pre-manipulation stigma consciousness on assessment one performance were conducted first entering math ACT scores into step one and then entering stigma consciousness scores as step two (for correlations, see Appendix I). Results show that stigma consciousness levels had no effect on pre-manipulation performance indicators ($\Delta R^2 < .01, p$'s > .10). Math ACT scores positively related to assessment one corrected scores and items answered per minute, and math ACT scores negatively related to average time spent with correctly answered items (see Table 6). Math ACT scores did not predict average time spent with incorrectly answered items ($p > .10$).

The same regression model was used to examine the effects of math ACT scores and stigma consciousness on difficult-items performance. ACT math score remained a notable predictor of corrected scores and average time spent with correctly answered items (see Table 7). Stigma consciousness had no effect on corrected scores or average time spent with correctly answered items. Neither math ACT scores nor stigma consciousness affected average time spent with incorrectly answered items among difficult items.
### Table 6

**Summary of Regression Analysis for ACT Math Scores and Stigma Consciousness Predicting Assessment One Performance Indicators**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Corrected Score ((R^2 = .29))</th>
<th>Items Answered Per Minute ((R^2 = .12))</th>
<th>Avg Time Spent with Correctly Answered Items ((R^2 = .09))</th>
<th>Avg Time Spent with Incorrectly Answered Items ((R^2 = .01))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(SE)</td>
<td>(B)</td>
<td>(\beta)</td>
<td>(SE)</td>
</tr>
<tr>
<td>ACT Math Score</td>
<td>1.18</td>
<td>.18</td>
<td>.53 **</td>
<td>1.03</td>
</tr>
<tr>
<td>Stigma Consciousness</td>
<td>1.15</td>
<td>.80</td>
<td>.12</td>
<td>.22</td>
</tr>
</tbody>
</table>

*Note.* \(^p < .10\)  \(*p < .05\)  \(**p < .01\)

### Table 7

**Summary of Regression Analysis for ACT Math Scores and Stigma Consciousness Predicting Assessment One Difficult-Item Performance Indicators**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Corrected Score ((R^2 = .22))</th>
<th>Avg Time Spent with Correctly Answered Items ((R^2 = .12))</th>
<th>Avg Time Spent with Incorrectly Answered Items ((R^2 = .03))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(B)</td>
<td>(SE)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>ACT Math Score</td>
<td>.62</td>
<td>.12</td>
<td>.45 **</td>
</tr>
<tr>
<td>Stigma Consciousness</td>
<td>.95</td>
<td>.52</td>
<td>.16 ^</td>
</tr>
</tbody>
</table>

*Note.* \(^p < .10\)  \(*p < .05\)  \(**p < .01\)
Hypothesis 2

An ANCOVA analysis showed that levels of post-manipulation stigma consciousness were unaffected by the stereotype manipulations when controlling for pre-manipulation stigma consciousness, $F(2, 117) = 1.54, p = .22, \eta^2 = .03$ (see Table 8).

Table 8

<table>
<thead>
<tr>
<th>Stereotype Treatment</th>
<th>Post-Manipulation Stigma Consciousness Means by Treatment (Controlling for Pre-Manipulation Stigma Consciousness)</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Threat ($n=41$)</td>
<td></td>
<td>4.91</td>
<td>.13</td>
</tr>
<tr>
<td>Control ($n=43$)</td>
<td></td>
<td>5.22</td>
<td>.12</td>
</tr>
<tr>
<td>Low Threat ($n=33$)</td>
<td></td>
<td>5.11</td>
<td>.14</td>
</tr>
</tbody>
</table>

Hypothesis 3

A treatment interaction variable was computed for each of the dominance measures (i.e., Social Dominance Orientation, Simple Adjective Test, Standardized Testosterone Levels) resulting in two interaction variables for each dominance measure (e.g., High Threat Social Dominance Orientation and Low Threat Social Dominance
Orientation). The control interaction variables were not entered in order to avoid multicollinearity in the model. A series of four hierarchical regressions were used to examine the main effects and treatment interaction effects of each dominance measure on each of the four performance indicators. At step one of each regression, math ACT scores, assessment one indicators, and the three main effect variables were entered into the regression model. At step two, the three high threat interaction variables and the three low threat interaction variables were entered into the model (for correlations, see Appendix I). Results showed that for each performance indicator, the interaction model was not a better predictor than the main effect model alone ($\Delta R^2 = .10$ to .01, $p$'s > .20). Additionally, none of the dominance measures' main effects predicted performance indicators at a significant level ($|\beta| = .15$ to .006, $p$'s > .10)

Research Question Analyses

Research Question One

An ANCOVA was used to determine whether levels of dominance were affected by the experimental conditions. Neither SDO scores, $F(2, 117) = .20, p = .82, \eta^2 = .004$, nor testosterone levels, $F(2, 79) = .54, p = .58, \eta^2 = .01$, varied according to the treatment group when controlling for pre-manipulation levels of each measure (see Table 9).
Table 9
*Post-Manipulation Means of Social Dominance Orientation and Testosterone Level (Controlling for Pre-Manipulation Levels of Each)*

<table>
<thead>
<tr>
<th>Social Dominance Orientation</th>
<th>Testosterone Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
</tr>
<tr>
<td>High Threat</td>
<td>2.87</td>
</tr>
<tr>
<td>Control</td>
<td>2.94</td>
</tr>
<tr>
<td>Low Threat</td>
<td>2.99</td>
</tr>
</tbody>
</table>

Research Question Two

No measures of dominance significantly correlated with pre-test measures of stigma consciousness. However, two measures of dominance were significantly correlated with post-test measures of stigma consciousness. The pre-manipulation Simple Adjective Test was negatively correlated with overall post-manipulation levels of stigma consciousness, average within-cell $r = -.20$, $p = .04$ (see Table 10). Upon closer examination, the correlation between these two measures is marginally significant when examining the only those within the *control* condition, $r = -.28$, $p = .07$ or the *high threat* condition, $r = -.24$, $p = .13$. However, the same two measures are far less correlated when examining the *low threat* condition only, $r = -.002$, $p = .99$. Thus, higher levels of personality dominance (as measured by the SAT) moderately correlated with lower stigma consciousness scores only when participants were not exposed to the *low threat*
treatment.

Furthermore, although post-test measures of social dominance orientation were not correlated with post-test measures of stigma consciousness overall, average within-cell $r = -.07, p = .46$, these measures did negatively correlate in the control condition, $r = -.32, p = .04$. The same measures did not significantly correlate in the high threat condition, $r = .26, p = .10$, or the low threat condition, $r = -.15, p = .41$.

Table 10

Average Within-Cell Correlations Between Stigma Consciousness and Dominance Measures.

<table>
<thead>
<tr>
<th></th>
<th>Pre-manip StigCon</th>
<th>Pre-manip SocDom</th>
<th>Pre-manip Simple Adj Test</th>
<th>Pre-manip Test</th>
<th>Post-manip Test</th>
<th>Post-manip StigCon</th>
<th>Post-manip SocDom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest StigCon</td>
<td>1.00</td>
<td>.04</td>
<td>-.09</td>
<td>-.13</td>
<td>-.13</td>
<td>.69**</td>
<td>-.06</td>
</tr>
<tr>
<td>Pretest SocDom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.82**</td>
</tr>
<tr>
<td>Simple Adj Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
</tr>
<tr>
<td>Pre-manip Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.82**</td>
</tr>
<tr>
<td>Post-manip Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.03</td>
</tr>
<tr>
<td>Post-manip StigCon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.20</td>
</tr>
<tr>
<td>Post-manip SocDom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Note. Correlations in this table may slightly differ from those in Appendix I because this table reports average within-cell correlations. N's for each comparison ranged from 117 to 79.*

* $p < .05$ ** $p < .01$
Finally, another important finding to note is that none of the measures of "dominance" used in this research correlated with one another (see Table 10). Although correlations between the social dominance orientation measures and testosterone levels were high relative to other correlations ($r = .14$ to $.21$), these correlations were not significant at the $p < .05$ level.
CHAPTER 4

DISCUSSION

Findings

The traditionally expected "lower score" associated with stereotype threat phenomena did not emerge among high threat participants in this study. Notably, however, this study did reveal a subtle difference between the high threat and non-threatened participants (i.e., control and low threat participants) on one of the four performance indicators. Specifically, participants under stereotype threat took about two seconds longer to incorrectly answer assessment items than did the other participants. Considering the relatively easy nature of the assessment tasks, overall, it seems reasonable to assume that the incorrectly answered items on each assessment were more difficult items that required more thought and consideration. Perhaps if more items presented in this study's assessments were more difficult, the cognitive depletion caused by the stereotype threat would be more evident in assessment scores. Such evidence may suggest that people are slower to complete more challenging tasks following exposure to a relevant, negative stereotype. This does not explain why high and low threat participants took longer to evaluate correctly answered difficult items, though. This slowed response to difficult questions may be a result of more deliberate consideration caused by the wording of the high and low threat treatments. Regardless of the reason, for the effect, it is reinforced by the design of the study; participants were randomly assigned to treatment conditions and each treatment group's performance indicators were
nearly identical before the manipulations were administered. Only following stereotype activation manipulations did performances begin to differ. The activation of a relevant, negative stereotype caused *high threat* individuals to slow down in evaluating incorrectly answered assessment items.

This evidence supports the idea that stereotype threat effects may manifest, in part, because of slower responses to more difficult items. In the assessments given in this study, neither scores nor items answered per minute varied by treatment condition. This lack of effect may be due to the relatively easy assessment items. Items on these assessment required only basic arithmetic abilities, and participants merely evaluated whether each equality was correct or incorrect. The ease of evaluation and lack of numerous possible responses may have permitted many respondents to complete most of the assessment items with little cognitive strain. If this were true, then any cognitive load generated by the stereotype threat may not have been enough to interfere with the evaluation of most question items. Additionally, the items answered per minute may not have shown the stereotype threat effect because participants answered many more items correctly than incorrectly. Thus, the small increase in latency for incorrectly answered items may have been too small to have significantly affected the items answered per minute. The lack of effect among these performance indicators does suggest that apparent stereotype threat-induced deficits are not due to inaccessible knowledge or self-handicapping (Stone, 2002) and may be due to cognitive load. The slowed responses to incorrectly assessed items are similar to the increased task-evaluation latencies found in Frantz and colleagues' (2004) IAT investigation of stereotype threat phenomena.
The Role of Stigma Consciousness

Research by Pinel (1999) has suggested that awareness of one's own stigmatized status (i.e., stigma consciousness) may mediate the effects of stereotype activation on task performance. This study found that debunking a stigmatizing stereotype affected higher stigma conscious participants. Stigma consciousness significantly enhanced post-manipulation assessment scores among low threat participants; that is, following the low threat message, the scores of participants higher in stigma consciousness improved. The debunking of a negative stereotype lead higher stigma conscious participants to perform better on assessment score, perhaps relieving stereotype threat effects affecting them prior to the low threat treatment.

These findings regarding stigma consciousness are a first glimpse into the complexities of stereotype threat phenomena. Instead of modifying the evaluation latencies (as the treatment conditions did), stigma consciousness significantly modified assessment scores following a low threat condition. The findings suggest that the actual abilities of highly stigma conscious individuals to correctly evaluate math items are enhanced by information that questions the validity of stigmatizing stereotypes. This increased ability may be a sign that highly stigma conscious participants were suffering from self-imposed stereotype threat effects prior to the low threat treatment, although the findings of no performance differences between high and low stigma conscious participants on assessment one make this less certain. Regardless, the simple awareness of one's own stigmatized status greatly enhanced task performance following the debunking of a stigmatizing stereotype.
Pinel's (1999) contention that stigma consciousness is the prime mediator between stereotype activation and task performance further suggests that higher levels of stigma consciousness predict reduced levels of task performance in the absence of any explicit stereotype activation. The only overall effect of stigma consciousness evident in this study was the positive influence of stigma consciousness on difficult-item scores and its negative influence on correctly answered item latencies for difficult items. More highly stigma conscious individuals answered more difficult items correctly and more quickly and were not, as predicted, inhibited by the high levels of stigma consciousness.

No evidence supported the notion that stigma consciousness alone inhibited pre-manipulation task performance. It may be that more explicit stereotype cues are necessary for stigma consciousness to be activated as a performance mediator; the initial assessment task may not have been perceived as a cue to math-gender stereotypes. Alternatively, it may be that the effect of stigma consciousness is small and more difficult items are necessary for the effect to observably manifest on stigmatizing tasks. Previous findings suggesting that stigma consciousness moderates stereotype threat effects (Brown & Pinel, 2003) used relatively difficult assessment items to reveal the relationship. The easy nature of this study's items may not have adequately taxed participants' abilities and may, in fact, have resulted in an ironic inflation of scores due to participants' eagerness to complete more items more quickly.

Furthermore, no evidence suggests that stigmatizing stereotype activation or refutation affected reported levels of post-manipulation stigma consciousness. These findings are not surprising when considering the findings discussed above. Stigma
consciousness, it seems, is not directly mediating the relationship between stereotype activation and performance, but is instead simply moderating the effects of stereotype activation (or debunking) on task performance. With these findings, there is little evidence that stigma consciousness alone is a sufficient or necessary factor for stereotype threat effect manifestation.

The Role of Dominance

Recently, evidence has emerged that identifies dominance-related measures such as testosterone (Josephs et al., 2003) and social dominance orientation (Philipp & Harton, 2005) as moderators of stereotype threat effects. More specifically, the findings suggest that higher levels of dominance lead to even greater performance decrements when under stereotype threat. Although empirical evidence demonstrates that supplemental steroid injections do not affect working memory capacity in women (Janowsky, Chavez, & Orwoll, 2000), findings in stereotype threat research suggest that high testosterone levels may deplete women's cognitive resources following a stigmatizing stereotype activation (Josephs et al., 2003).

Overall, no evidence from this study supports the role of Social Dominance Orientation, personality dominance (i.e., SAT), or testosterone as moderators of stereotype threat effects. Even among difficult items within the assessment, no effect of dominance was evident on any performance indicators. Perhaps the effects of dominance emerge only on tasks that are difficult and more convincingly diagnostic of advanced domain ability. For example, in both Philipp and Harton (2005) and Josephs et al. (2003), the tasks used to assess performance were a finite set of challenging questions drawn
from already established standardized tests of quantitative ability. The present study's task, in contrast, permitted participants to complete any number of relatively easy items in a fixed period of time. Also, the author developed all items in the present study. Although the assessments in this study were stated to be diagnostic of ability, the assessments may not be perceived as diagnostic of domain ability if the items comprising the assessments seemed noticeably easy and seemingly trivial. Regardless, the findings of this study do not support the notion that dominance moderates stereotype threat effects.

Some research has suggested that dominance levels (testosterone in particular) are susceptible to change in response to social cues such as changes in status hierarchies (Jeffcoate et al., 1986). In this study, stereotype activation did not affect post-manipulation dominance measures (i.e., Social Dominance Orientation and testosterone). The lack of change in this study may indicate a variety of phenomena. One possible explanation for this null finding is that the stereotype manipulations did not elicit the social information necessary for dominance levels to change. The isolation of wearing headphones during the assessments may have diminished the competitive tendencies of more dominant individuals.

Another consideration is that salivary testosterone levels in females are approximately one-seventh of those in males and the variance of females' testosterone levels is approximately one-quarter of males' variance (Dabbs & Mohammed, 1992). With this in mind, the measure of female testosterone may have been too imprecise to accurately measure small amounts of change due to the stereotype manipulation. This is imprecision may be exacerbated because samples were collected in the late afternoon,
when testosterone levels are known to be lower than earlier times of the day (Dabbs & Mohammed, 1992). Although the assays in this study were conducted as directed by Salimetrics instructions, the assays were not conducted by professional technicians and were susceptible to confounding factors including variations in plate incubation time, exposure of plates to multiple laboratory environments, and less precise measurement of chemicals due to simple inexperience. Finally, even visually unperceivable levels of blood contamination from oral micro-trauma (e.g., recent brushing of teeth) may artificially raise salivary testosterone levels for an hour or more following the trauma (Granger et al., 2004). Although none of the saliva samples tested in this study appeared to have blood contamination, no attempt was made to screen participants regarding oral hygiene or oral trauma. Thus, some participants' measured testosterone levels may be exaggerated due to blood contamination and, therefore, be less predictive of testosterone-dependent behaviors.

Although salivary testosterone measured by an enzyme immunoassay results (EIA) is highly correlated with serum testosterone, the correlation between these two measures in women \((r = .38 \text{ to } .48)\) is notably lower than in men \((r = .80 \text{ to } .85;\) Salimetrics, 2004). Because serum free-testosterone is so poorly correlated with EIA measured salivary testosterone, computer-generated estimates suggest that salivary testosterone measured by EIA may underestimate testosterone-dependent behaviors by as much as 29.99%, greatly reducing the statistical power of analyses using salivary measures of testosterone (Granger et al., 2004; Shirtcliff, Granger, & Likos, 2002). This underestimation may further explain why this study's methods of investigation found no
effects of testosterone on performance. Although the findings of Josephs and colleagues (2003) were derived using less powerful statistical techniques (ANCOVA and t-test analyses), the effect of testosterone may have been more pronounced because the study more explicitly primed a variety of relevant stigmas by requiring participants to actively respond to the stigmatizing statements (e.g., "In math classes, I often feel that others look down on me because of my gender."). In addition, the statistical power of testosterone measures may have been enhanced in Josephs's study because professional laboratory technicians assayed the saliva samples used in statistical analyses. Future research must consider such power issues when attempting to use salivary testosterone assays.

The pre-manipulation levels of Simple Adjective Tests (SAT) were inversely related to post-manipulation Stigma Consciousness scores among high threat and control participants. Furthermore, post-manipulation Social Dominance Orientation scores also rose as post-manipulation Stigma Consciousness decreased only among control participants. It is difficult to understand these findings in light of the previously hypothesized relationships.

In the case of the SAT correlation, it may well be that more dominant individuals were more attuned to the purpose of the study and as a result were more guarded in their responses to the Stigma Consciousness Questionnaire. Alternatively, low levels of personality dominance may be indicative of women who hold more traditional gender-role ideals, a trait indicative of more highly stigma conscious women (Pinel, 1999). SDO's relationship with Stigma Consciousness may exist for the same reason--highly stigma conscious individuals are simply lower in dominance. If this is true, the
relationship may not have emerged among pre-manipulation scores due to the shorter scales used in pre-manipulation assessments. It is also likely that a more complex interaction between individual dominance measures and stigma consciousness exists outside the scope of this study.

Finally, it is very important to note that none of the three "dominance" measures used in this study correlated with the other measures of "dominance." Previous research suggested that SAT scores tend to be higher among higher testosterone women (Grant & France, 2001) and stereotype threat findings suggested that testosterone (Josephs et al., 2003) and SDO (Philipp & Harton, 2005) similarly moderated stereotype threat effects among women. In the present study, however, none of these relationships were evident. Certainly the imprecise measurement of testosterone mentioned previously might be to blame for the non-significant correlations between testosterone measures and the self-response measures. If this were the only problem, however, the measures of SDO and SAT should still be correlated to one another--this is not the case. An important lesson from these findings is that the psychological measures of "dominance" are not necessarily measuring a common, latent construct. Instead, different types of dominance may be gauged by these measures; a common dominant disposition may not underlie each measure. Future research investigating the effects of "dominance" will do well to be specific about the measure being used and resist the temptation to equivocate regarding dominance constructs.
Summary of Findings

In all, the design of this study permitted a number of analyses not previously attempted in stereotype threat research. The pre-test/post-test design permitted an in-depth investigation of both the mechanisms that lead to stereotype threat phenomena and the psychological consequences of stereotype activation beyond performance effects. Such a design permits easy modeling of stereotype activation's effects on both task performance and additional post-manipulation factors of interest while accounting for pre-manipulation covariates.

In addition to the design of the study, the computer administration of the study permitted high internal validity across numerous testing sessions. Because nearly equal proportions of each treatment group were present in each testing session and the same experimenter and research assistant administered each session, there is little worry that any differences between testing sessions confounded the findings of this study. Moreover, the random assignment of treatment condition throughout the testing room made it difficult for the experimenter to know which computers were administering each treatment, and the computer guided instructions reduced external influences on participants' assessment performances and focused their attention on the stereotype manipulations.

This study found that stereotype activation caused changes in the task evaluation latencies of incorrectly answered items among threatened individuals without any differences in reported anxiety—a finding that suggests that cognitive load is affected independent of anxious feelings. Also, the differential influence of stigma consciousness
on stereotype-induced underperformance had a different influence on performance than mere stereotype activation. The investigation also pointed to the important role that debunking a stereotype makes to high stigma conscious participants.

Future research would do well to continue using pre-test/post-test designs to more fully model the mechanisms that lead to stereotype threat phenomena. It is too tempting for psychologists to test individual moderators without examining the differential effects these moderators have across different levels of stereotype activation. More importantly, however, different types of performance indicators should be used to test stereotype threat effects in order to more clearly identify the causes of underperformance. As with real academic performances, experimental assessments should make the timed nature of assessments salient to participants and measure response latencies to assessment items. Assessment score is only one indicator of task performance and indicators such as overall score may not reveal smaller performance decrements caused by decreased cognitive load. In many cases, it may be that purported "moderators" of stereotype threat effects are merely demonstrating different main effects on performance that compound the apparent underperformance effects of the stereotype activation.

Theoretical Implications

This study has shed additional light on the theoretical mechanisms that may govern stereotype threat phenomena. The evidence generated by this research suggests that stereotype threat effects were adequately manifested by the treatment conditions; the participants stigmatized by the activated stereotype (i.e., females) suffered deficits in performance (i.e., slower evaluation responses) that may have been mediated by task-
irrelevant cognitive activity. This research offers support for the notion that some type of cognitive activity is primed by stereotype activation and that cognitive activity leads to slower evaluations of task items, especially more difficult task items (i.e., items answered incorrectly and difficult items answered correctly). The lack of differences between participants' reported nervousness, effort, and experienced difficulty further suggest that stereotype threat effects are not necessarily mediated by threat-induced anxiety or self-handicapping.

Important considerations must be made regarding this lack of difference, though. For one, participants exposed to the high threat condition may have underestimated their effort relative to control and low threat participants. In order to account for this possibility, subtler, less face-valid measures of effort and anxiety (e.g., heart rate, measures of cognitive load) could be used that are sensitive to attempts of underestimation. A second consideration is that if participants under threat did not experience the second assessment as more difficult, the cognitive activity elicited by stereotype activation may not have been salient to participants. Stereotype activation may have triggered unconscious cognitive activity that interfered with cognitive resources unperceived by participants. If such a phenomenon is occurring, the defining features of stereotype threat mechanisms must be reassessed. Only future research will tell whether the cognitive distractions are driven by conscious or unconscious mechanisms.

Evidence is less supportive about whether the model proposed in chapter one is adequate for explaining stereotype threat phenomena. First, the model assumes that a person must be aware of stigmas associated with his or her group memberships before
stereotype threat effects will occur (i.e., perception factors). It is argued that such awareness is generated through highly identifying with the stigmatized group and is aided by identifying with the domain of the stigma, resulting in stigma consciousness. The evidence generated in this study does not support the contention that stigma consciousness is necessarily related to people's awareness of gender stereotypes; belief in a gender-related math stereotype did not covary with stigma consciousness ($r = .04, p = .67$).

Perceptions of the stereotype's plausibility seem to have been somewhat affected by the low threat treatment. Specifically, higher levels of stigma consciousness increased participants' assessment scores following the debunking message. Stigma consciousness did not moderate scores among any of the other treatment conditions, suggesting that stigma consciousness is not a sufficient factor for defining stereotype perception, but may enhance attention to messages debunking relevant stigmatizing stereotypes. There was no evidence that more highly stigma conscious individuals are more prone to stereotype threat effects without explicit stereotype activation.

The lack of findings regarding dominant dispositions as predictors of stereotype susceptibility suggest that these factors may be subtler in moderating stereotype threat effects or may be irrelevant to the actual manifestation of stereotype threat phenomena. Additionally, the lack of correlation between any of the dominance measures suggests that each measure was describing unrelated factors that were not indicators of a single, latent dominance factor.
Finally, the findings in this research support previous claims that stereotype threat phenomena are most evident among more cognitively taxing tasks and fail to appear when tasks are easily accomplished with minimal effort. The latency information, in particular, demonstrated that stereotype threat-induced deficiencies manifest for participants regardless of their domain identification. Among more difficult tasks, it may well be that these increased response latencies are the reason for apparent stereotype decrements in performance score.

Limitations

Although a number of important findings have emerged from this research, it is important to consider a few of the shortcomings of this study. Specifically noted below are limitations regarding the stereotype manipulations employed and the study’s external validity.

Manipulations

Although this research had a number of advantageous design features that permitted novel analyses of stereotype threat effects, a few comments are warranted regarding the lack of more change among the manipulation check variables. The most obvious explanation for the null findings of manipulation checks is that none of the treatments had any effect on participants and the lack of manipulation check significance is indicative of poor stereotype threat treatments. In fact, the means by which treatments were administered differed from most previous research; that is, the manipulations were administered by a disembodied male voice instead of a physically present agent. Such a
treatment could indeed be less effective due to the reduced immediacy of a recording and the perceived reduced publicity of the stereotype information.

Additionally, the wording of the debunking treatment left open the possibility that there are gender differences in at least some math abilities rather than totally dispelling this notion or specifying the types of mathematical tasks in which women do perform better than men. Consideration must be given to the possibility that the "low threat" condition served to precipitate stereotype threat among some participants in that the wording of the treatment acknowledged the possibility of male math superiority. Although plausible, this explanation is not fully adequate in explaining the shortcomings of the manipulations. First, reduced performance consistent with stereotype threat effects was evident among latencies of incorrectly answered items. Second, participants in the low threat condition rated men and women as more equal in math ability than either control or high threat participants.

Another possible reason for the failure of the manipulation checks is that the questions were inadequate measures of the manipulations. No questions were posed regarding the purpose of the study or the apparent differences between assessment one and assessment two. Instead, the manipulation checks make theoretical assumptions that the treatments would elicit different levels of belief in gender/math stereotypes and alter participants' concern for what the experimenter might be thinking. It may be that these questions were inadequate measures of the true qualities that lead to differences among the treatment groups.
Future research should consider using more fundamental manipulation checks, for example, asking participants to recall the stated treatment or the purpose of the assessment. Researchers should avoid manipulation check questions that assume a theoretical model; questions used as manipulation checks should simply assess participants' awareness of the treatments.

**Experimental Validity**

Although the findings of this study are important to understanding many types of stereotype threat phenomena, it must be noted that the population of this study consisted mainly of young, Midwestern, European-American females who have the good fortune to be attending college. Although it can be assumed that this population is especially appropriate when considering the effects of gender-related math stereotypes on math performance, it may be that many of the factors found here to affect performance are unique to this population and are not necessary for mechanisms of other stereotype threat phenomena. For example, the influence of stigma consciousness on stereotype threat effects may only exist in more formal contexts (e.g., college classrooms) and may be less influential in real world settings (e.g., job performance). For these reasons it will be very important for these findings to be replicated in future research among new groups with unique stereotypes and tasks.

**Testosterone Measures**

Finally, it is important to mention that the ELISA procedure used to measure testosterone in women may have resulted in this study being underpowered to detect testosterone effects. Although assay procedures were conducted according to
specifications, the low correlation between salivary testosterone and serum levels among women is known to be small and visually unnoticeable blood contamination may have artificially increased testosterone levels for an unknown number of participants. Although researchers may desire to measure women's testosterone via salivary assay in the future, it is important to calibrate power estimates based on the known under-estimation of the technique and to screen participants for possible oral trauma that may lead to blood-contaminated saliva samples.

Summary

This research is a small first step in improving our understanding of stereotype-induced underperformance. Whether such effects should be called by the name "stereotype threat" is of some dispute; little evidence demonstrates that people feel threatened by the stigmatizing stereotype. However, social-cognitive evidence increasingly demonstrates that cognitive mechanisms are triggered by social stereotypes and these cognitive mechanisms may diminish task performances of stigmatized individuals. Factors such as stigma consciousness may be found to moderate stereotype threat effects. Yet it is important to investigate the exact effect of purported moderators through systematically varying moderator levels and rigorously testing their mechanisms instead of only piecing together mechanisms from various empirical findings.

Stereotype threat effects are often cited as the reason for many performance disparities between different social groups. More and more, educators and social scientists alike are looking for the mechanisms that moderate these effects so as to diminish potential under-performance on important assessments that have future
implications for students. This research has found that stereotype inoculations may be especially important for people who are acutely aware of the stigmas associated with their identity. Students are regularly bombarded with stigmatizing messages from their social world about their limited capabilities and innate shortcomings. Although many educators may attempt to induce a sense of efficacy among students with messages of endless potential and capabilities, a more important message for stigmatized students may be to highlight the erroneous beliefs behind group stigmas. By debunking the stereotypes that students hold about themselves, educators may go far in improving their performance and giving them a sense of ability that they cannot give themselves.
REFERENCES


APPENDIX A

PRE-TEST INFORMED CONSENT FORM
UNIVERSITY OF NORTHERN IOWA -- HUMAN PARTICIPANTS REVIEW
INFORMED CONSENT

Project Title: Domain Assessment PRE-TEST
Name of Investigator(s): Michael Philipp

Invitation to Participate: You are invited to participate in a research project conducted through the Department of Psychology at the University of Northern Iowa. The University requires that you give your signed agreement to participate in this project. The following information is provided to help you make an informed decision whether or not to participate.

Nature and Purpose: This research is designed to assess particular characteristics that may partially influence how well people perform on different assessment tools (e.g., math tests). Ultimately, the information provided by participants will aid the researcher in selecting participants for future studies.

Explanation of Procedures: You will be given 4 questionnaires to complete for this study. The questionnaires will ask you to describe yourself by rating your perceptions of how other's judge you, rating your preferences for social standards, rating your attitudes toward math, English, and the general academic domain, and rating the extent to which particular words describe you well. The questionnaire will also ask you to report your gender. You will also be asked if you would like to be contacted to participate in later research that uses the responses that you gave in this study. If you wish to participate in later research you will be asked to provide contact information and a personal identification number.

Discomfort and Risks: The researchers foresee no long term discomfort or risks from participation in this study.

Benefits: There are no direct benefits to participating in this research. You will receive one half of a research credit for your participation.

Confidentiality: Information obtained during this study which could identify you will be kept strictly confidential. All contact information will tied to the data by the personal identification number provided. However, all contact information will be destroyed once contact has been successfully made. The summarized findings with no identifying information may be published in an academic journal or presented at a scholarly conference.

Right to Refuse or Withdraw: Your participation is completely voluntary. You are free to withdraw from participation at any time or to choose not to participate at all, and by doing so, you will not be penalized or lose benefits to which you are otherwise entitled.

Questions: If you have questions about the study or desire information in the future regarding your participation or the study generally, you can contact Michael Philipp at 319- 273-2303 or Helen Harton at the Department of Psychology, University of Northern Iowa 319-273-2235. You can also contact the Office of the Human Participants Coordinator, University of Northern Iowa, at 319-273-2748, for answers to questions about rights of research participants and the participant review process.

Agreement: I am fully aware of the nature and extent of my participation in this project as stated above and the possible risks arising from it. I hereby agree to participate in this project. I acknowledge that I have received a copy of this consent statement. I am 18 years of age or older.

Participant's Signature ___________________________ Participant's Printed Name ___________________________ Date __________
Signature of Investigator ___________________________ Date __________ Signature of Advisor ___________________________ Date __________
APPENDIX B

STIGMA CONSCIOUSNESS QUESTIONNAIRE

PINEL (1999)
Using the following scale, please indicate the letter that best describes how much you agree with each of the statements below. Larger numbers represent stronger levels of agreement.

\[
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\mid & & & & & & & & \\
\text{Strongly} & \text{Strongly} \\
\text{Disagree} & \text{Agree} \\
\end{array}
\]

---

_R_ Stereotypes about women have not affected me personally.

_R_ I never worry that my behaviors will be viewed as stereotypically female.

_____ When interacting with men, I feel like they interpret all of my behaviors in terms of the fact that I am a woman.

_R_ Most men do not judge women on the basis of their gender.

_R_ My being female does not influence how men act with me.

_R_ I almost never think about the fact that I am female when I interact with men.

_R_ My being female does not influence how people act with me.

_____ Most men have a lot more sexist thoughts than they actually express.

_R_ I often think that men are unfairly accused of being sexist.

_____ Most men have a problem viewing women as equals.

---

Note: "R" denotes an item that was reverse scored.
A one (Strongly Disagree) through seven (Strongly Agree) scale was used on the pre-test version of this questionnaire.
APPENDIX C

SOCIAL DOMINANCE ORIENTATION SCALE

PRATTO, SIDANIUS, STALLWORTH, & MALLE (1994)
Using the following scale, please indicate the number that best describes how much you agree with each of the statements below. Larger numbers represent stronger levels of agreement.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly Agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Some groups of people are simply inferior to other groups.
2. We should strive to make incomes as equal as possible.
3. In getting what you want, it is sometimes necessary to use force against other groups.
4. Group equality should be our ideal.
5. It's OK if some groups have more of a chance in life than others.
6. We should do what we can to equalize conditions for different groups.
7. We would have fewer problems if we treated people more equally.
8. To get ahead in life, it is sometimes necessary to step on other groups.
9. No one group should dominate in society.
10. If certain groups stayed in their place, we would have fewer problems.
11. It's probably a good thing that certain groups are at the top and other groups are at the bottom.
12. Inferior groups should stay in their place.
13. All groups should be given an equal chance in life.
14. Sometimes other groups must be kept in their place.
15. It would be good if groups could be equal.

Note: "R" denotes an item that was reverse scored.

A one (Strongly Disagree) through seven (Strongly Agree) scale was used on the pre-test version of this questionnaire.
APPENDIX D

MATH IDENTIFICATION QUESTIONNAIRE

SMITH & WHITE (2001)
Indicate the degree to which you agree with each of the statements below by writing numbers in the space that follows each statement. Use the following scale to indicate agreement:

Write in smaller numbers to indicate less agreement (zero indicates no agreement) and larger numbers indicate more agreement (nine indicates complete agreement).

1 2 3 4 5 6 7 8 9

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree</td>
<td>Agree</td>
</tr>
</tbody>
</table>

I have positive attitudes regarding my abilities in English

I learn things quickly in Math classes.

English is one of my best subjects.

Math is one of my best subjects.

I have better than average Math abilities.

I have better than average abilities in the Social Sciences.

I value being a student.

I have positive attitudes regarding my abilities in Social Sciences

Being good at English is important to me.

I dislike Math

Academics are an important and/or necessary part of my life.

I enjoy Math-related subjects.

I enjoy Social Science-related subjects.

I would like to take a job in a Math-related field.

I identify Math with a sense of who I am.

Being a student is important to me.

Being good at Math is important to me.

I get good marks in Math.

I have always done well in English.

I’m hopeless in Math classes.
I get good grades in English.

I do poorly on tests of English.

Note: Items in bold were used to derive a math identification score.

"R" denotes an item that was reverse scored.

A zero (Do Not Agree) through nine (Completely Agree) scale was used on the pre-test version of this questionnaire.

* denotes items added to the original MIQ.
APPENDIX E

FINAL STUDY INFORMED CONSENT FORM
UNIVERSITY OF NORTHERN IOWA – HUMAN PARTICIPANTS REVIEW
INFORMED CONSENT

Project Title: Mathematical-Domain Assessment Study
Name of Investigator(s): Michael Philipp

Invitation to Participate: You are invited to participate in a research project conducted through the Department of Psychology at the University of Northern Iowa. The University requires that you give your signed agreement to participate in this project. The following information is provided to help you make an informed decision whether or not to participate.

Nature and Purpose: This research is concerned with evaluating the many factors that affect people when taking tests of mathematical ability. Ultimately, the information provided by participants will aid researchers in the development of less-biased assessment procedures in mathematics and related domains.

Explanation of Procedures: Two tests of arithmetic ability will be administered and you will also be asked to provide saliva samples twice during the study. These saliva samples will be used to measure current hormone levels that may affect test performance. You will then be asked to complete items that ask you to rate your agreement with statements about social attitudes and perceptions of stigmas. You will also be asked to respond to questions regarding demographics and the quality of the study, itself. Finally, we will request to retrieve academic information related to standardized test performance. Participation will take approximately 60 minutes.

Discomfort and Risks: The researchers foresee no long-term discomfort or risks from participation in this study. The difficulty level of particular sample test problems may cause you mild anxiety as you work on solving the problems. No other discomforts are foreseen.

Benefits: There are no direct benefits to participating in this research. You will receive one research credit for your participation.

Confidentiality: Information obtained during this study that could identify you will be kept strictly confidential. This consent form will remain separate from any information you provide during the study. Saliva samples will be used ONLY to assess hormone levels and will be destroyed after testing. The summarized findings with no identifying information may be published in an academic journal or presented at a scholarly conference.

Right to Refuse or Withdraw: Your participation is completely voluntary. You are free to withdraw from participation at any time or to choose not to participate at all, and by doing so, you will not be penalized or lose benefits to which you are otherwise entitled.

Questions: If you have questions about the study or desire information in the future regarding your participation or the study generally, you can contact Michael Philipp at 319-273-3114 or Helen Harton at the Department of Psychology, University of Northern Iowa; 319-273-2235. You can also contact the Office of the Human Participants Coordinator, University of Northern Iowa, at 319-273-2748, for answers to questions about rights of research participants and the participant review process.

Agreement: I am fully aware of the nature and extent of my participation in this project as stated above and the possible risks arising from it. I hereby agree to participate in this project. I acknowledge that I have received a copy of this consent statement. I am 18 years of age or older.

Participant's Signature ___________________________ Participant's Printed Name ___________________________ Date ____________

Signature of Investigator ___________________________ Date ____________ Signature of Advisor ___________________________ Date ____________
APPENDIX F

SIMPLE ADJECTIVE TEST

GRANT (1992)
For each word listed, check the box next to the word if you would answer “Yes” to the question,

“Do you quite often feel _______?”

<table>
<thead>
<tr>
<th>Activated</th>
<th>Domineering</th>
<th>Pain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admired</td>
<td>Egotistical</td>
<td>Powerful</td>
</tr>
<tr>
<td>Aggressive</td>
<td>Enjoyment</td>
<td>Protected</td>
</tr>
<tr>
<td>Alert</td>
<td>Feeble</td>
<td>Proud</td>
</tr>
<tr>
<td>Arrogant</td>
<td>Free</td>
<td>Rejected</td>
</tr>
<tr>
<td>Awed</td>
<td>Friendly</td>
<td>Responsible</td>
</tr>
<tr>
<td>Bold</td>
<td>Frustrated</td>
<td>Sad</td>
</tr>
<tr>
<td>Bored</td>
<td>Guilty</td>
<td>Self-satisfied</td>
</tr>
<tr>
<td>Capable</td>
<td>Happy</td>
<td>Serious</td>
</tr>
<tr>
<td>Carefree</td>
<td>Helpless</td>
<td>Shamed</td>
</tr>
<tr>
<td>Concentrating</td>
<td>Humiliated</td>
<td>Sheltered</td>
</tr>
<tr>
<td>Confused</td>
<td>Influential</td>
<td>Shy</td>
</tr>
<tr>
<td>Controlling</td>
<td>Inhibited</td>
<td>Startled</td>
</tr>
<tr>
<td>Cruel</td>
<td>Insecure</td>
<td>Strong</td>
</tr>
<tr>
<td>Crushed</td>
<td>Inspired</td>
<td>Timid</td>
</tr>
<tr>
<td>Deactivated</td>
<td>Joyful</td>
<td>Triumphant</td>
</tr>
<tr>
<td>Defeated</td>
<td>Kind</td>
<td>Troubled</td>
</tr>
<tr>
<td>Depressed</td>
<td>Lonely</td>
<td>Useful</td>
</tr>
<tr>
<td>Despairing</td>
<td>Masterful</td>
<td>Vigorous</td>
</tr>
<tr>
<td>Dignified</td>
<td>Mighty</td>
<td></td>
</tr>
<tr>
<td>Distressed</td>
<td>Nauseated</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX G

FINAL STUDY MANIPULATION QUESTIONNAIRE
How nervous were you during the math performance task?

1 —— 2 —— 3 —— 4 —— 5 —— 6 —— 7 —— 8 —— 9
Not At All Nervous —— Very Nervous

How would you rate the amount of effort you put into completing the math performance task?

1 —— 2 —— 3 —— 4 —— 5 —— 6 —— 7 —— 8 —— 9
No Effort At All —— As Much Effort As I Could

Compared to the first math assessment, how difficult was the second math assessment?

(1 —— 2 —— 3 —— 4 —— 5 —— 6 —— 7 —— 8 —— 9)
-4 -3 -2 -1 0 +1 +2 +3 +4
Much Easier About the Same Much More Difficult

Do you feel that your performance on the math problems was a good reflection of your true math ability?

YES (1) NO (2)

To what extent, if any, do you believe that one sex is better at math than the other?

(1 —— 2 —— 3 —— 4 —— 5 —— 6 —— 7 —— 8 —— 9)
-4 -3 -2 -1 0 1 2 3 4
Men are Much Better No Difference Women are Much Better
To what extent do you agree or disagree with the following statements:

I am concerned that the researcher will judge women, as a whole, based on my performance on this test.

1 —— 2 —— 3 —— 4 —— 5 —— 6 —— 7 —— 8 —— 9
Strongly Disagree Strongly Agree

The researcher will think that women, as a whole, have less math ability if I did not do well on this test.

1 —— 2 —— 3 —— 4 —— 5 —— 6 —— 7 —— 8 —— 9
Strongly Disagree Strongly Agree

Note: Scales appearing in **bold** are those seen by participants during the session. In some cases, the scale used by participants was differently coded into the dataset. In such cases, the scale listed in parentheses demonstrates the coding scheme used to write responses.
APPENDIX H

SALIVA ASSAY PROCEDURES
SALIVA TESTOSTERONE ASSAY PROCEDURES

SAMPLE PREPARATION PROCESS

1) Freeze all samples. Immediately before testing, bring all samples to room temperature.

2) Vortex samples.

MATERIALS: + Vortex

3) Transfer 200 µL from cryovials to centrifuge vials. Cross label appropriately.

MATERIALS:
+ Cross labeling schematic
+ 200 µL pipette (yellow tips)
+ vial holding apparatus
+ 400 µL centrifuge vials (appropriate for 3000rpm centrifuging)

TESTOSTERONE ASSAY PROCESS

1. Centrifuge samples at 3000 rpm for 15 minutes to clarify.

MATERIALS: + centrifuge

2. Replace H1 and H2 positions w/ nontreated wells. Add 50µL of assay diluent (pink liquid) to H1 and H2. Add 50µL of assay diluent (pink liquid) to the two zero standard wells G1 and G2.

MATERIALS: + 50 µL pipette (yellow tips)

3. Set up six standard serial dilution tubes.

#1. 200 µL standard (small vial, clear top)
#2. 150 µL assay diluent (pink liquid)- then add 100 µL from #1 & MIX
#3. 150 µL assay diluent (pink liquid)- then add 100 µL from #2 & MIX
#4. 150 µL assay diluent (pink liquid)- then add 100 µL from #3 & MIX
#5. 150 µL assay diluent (pink liquid)- then add 100 µL from #4 & MIX
#6. 150 µL assay diluent (pink liquid)- then add 100 µL from #5 & MIX

>> Add 50 µL to A1 & A2
>> Add 50 µL to B1 & B2
>> Add 50 µL to C1 & C2
>> Add 50 µL to D1 & D2
>> Add 50 µL to E1 & E2
>> Add 50 µL to F1 & F2
4. Add 50 µL of each saliva sample to the wells (chart locations of samples very thoroughly).

**MATERIALS:**
+ 50 µL pipette (yellow tips)

5. Combine 7 µL of enzyme conjugate (small vial w/ pink sticker) w/ 18 mL of assay diluent (pink liquid). Immediately mix & add 150 µL of the diluted conjugate to each well using multichannel pipettor.

**MATERIALS:**
+ 7 µL pipette (yellow tips)
+ 1 mL beaker
+ 20 mL beaker
+ multichannel reservoirs
+ 150 µL multichannel pipettor

6. Stir 60 minutes on plate rotator @ 500 rpm (at room temp)

**MATERIALS:**
+ plate rotator @ 500 rpm

7. Combine 12 mL wash buffer (large clear bottle) w/ 108 mL distilled water.

**MATERIALS:**
+ 20 mL beaker
+ 150 mL beaker

8. Flip plate to remove liquid from wells. Add 300 µL of diluted wash buffer to each well using multichannel pipettor. Swirl gently, flip plate. Repeat wash 3 additional times. Blot plate upside down to finish.

**MATERIALS:**
+ multichannel reservoirs
+ 150 µL multichannel pipettor (yellow tips)
+ Paper towels for blotting
9. Add 200 µL of TMB solution (brown bottle) to each well using multichannel pipettor.

**MATERIALS:**
- multichannel reservoirs (20mL)
- 200 µL multichannel pipettor (yellow tips)

10. Stir 5 minutes (plate rotator @ 500 rpm), then incubate in dark for 25 minutes.

**MATERIALS:**
- plate rotator @ 500 rpm
- dark covering

11. Add 50 µL of stop solution (small clear bottle) to each well using multichannel pipettor.

**MATERIALS:**
- multichannel reservoirs (5 mL)
- 50 µL multichannel pipettor (yellow tips)

12. Stir (plate rotator @ 500 rpm) 3 minutes, or until all green color has turned yellow.

**MATERIALS:**
- plate rotator @ 500 rpm

13. Wipe bottom of plate w/ moist cloth, then w/ dry cloth. Read A450 in plate reader. Read w/in 10 minutes of adding stop solution. (correction at 492 to 620 is desirable)

**MATERIALS:**
- moist cloth
- dry cloth
- plate reader
APPENDIX I

CORRELATION MATRIX
Table II

**Correlations Between Study Variables**

<table>
<thead>
<tr>
<th></th>
<th>Pretest Stigma Consciousness</th>
<th>Pretest Social Dominance Orientation</th>
<th>Pretest Math Identification</th>
<th>Simple Adjective Test</th>
<th>Post-manip Stigma Consciousness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest Stigma Consciousness</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Social Dominance Orientation</td>
<td>.02</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Math Identification</td>
<td>-.06</td>
<td>.15</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple Adjective Test</td>
<td>-.08</td>
<td>.09</td>
<td>-.05</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Post-manip Stig Con</td>
<td>.68</td>
<td>-.02</td>
<td>.00</td>
<td>-.19</td>
<td>1.00</td>
</tr>
<tr>
<td>Post-manip Soc Dom Orientation</td>
<td>-.05</td>
<td>.81</td>
<td>.14</td>
<td>.08</td>
<td>-.06</td>
</tr>
<tr>
<td>Post-manip Math ID</td>
<td>-.07</td>
<td>.14</td>
<td>.91</td>
<td>-.08</td>
<td>.08</td>
</tr>
<tr>
<td>Pre-Manipulation Test Level</td>
<td>-.11</td>
<td>-.12</td>
<td>.02</td>
<td>.09</td>
<td>-.02</td>
</tr>
<tr>
<td>Post-Manipulation Test Level</td>
<td>-.12</td>
<td>-.22</td>
<td>-.16</td>
<td>.14</td>
<td>-.04</td>
</tr>
<tr>
<td>Assessment One Score</td>
<td>.13</td>
<td>.00</td>
<td>.29</td>
<td>-.15</td>
<td>.31</td>
</tr>
<tr>
<td>Assessment One Items Answered Per Minute</td>
<td>.16</td>
<td>.03</td>
<td>.11</td>
<td>.00</td>
<td>.24</td>
</tr>
<tr>
<td>Assessment One Avg. Latency for Correct Items</td>
<td>-.14</td>
<td>-.01</td>
<td>-.15</td>
<td>-.01</td>
<td>-.21</td>
</tr>
<tr>
<td>Assessment One Avg. Latency for Incorrect Items</td>
<td>-.08</td>
<td>.05</td>
<td>.12</td>
<td>.08</td>
<td>-.12</td>
</tr>
<tr>
<td>Assessment One Difficult Item Score</td>
<td>.16</td>
<td>.00</td>
<td>.22</td>
<td>-.15</td>
<td>.33</td>
</tr>
</tbody>
</table>

*Note.* N's for each comparison ranged from 117 to 79. **Bolded** correlations are significant at $p < .05$
<table>
<thead>
<tr>
<th>Assessment One Avg. Latency for Correct Difficult Items</th>
<th>Pretest Stigma Consciousness</th>
<th>Pretest Social Dominance Orientation</th>
<th>Pretest Math Identification</th>
<th>Simple Adjective Test</th>
<th>Post-manip Stigma Consciousness</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.18</td>
<td>-0.92</td>
<td>-0.19</td>
<td>-0.03</td>
<td>-0.24</td>
<td></td>
</tr>
<tr>
<td>Assessment One Avg. Latency for Incorrect Difficult Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.07</td>
<td>-0.06</td>
<td>0.05</td>
<td>-0.02</td>
<td>-0.09</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.19</td>
<td>0.04</td>
<td>0.23</td>
<td>-0.09</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Items Answered Per Minute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.18</td>
<td>0.07</td>
<td>0.12</td>
<td>0.03</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Correct Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.12</td>
<td>-0.08</td>
<td>-0.14</td>
<td>-0.05</td>
<td>-0.19</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Incorrect Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.16</td>
<td>0.06</td>
<td>0.09</td>
<td>-0.01</td>
<td>-0.16</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Difficult-Item Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>-0.01</td>
<td>0.20</td>
<td>-0.09</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Correct Difficult Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.04</td>
<td>-0.14</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Incorrect Difficult Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.14</td>
<td>0.09</td>
<td>0.14</td>
<td>-0.03</td>
<td>-0.13</td>
<td></td>
</tr>
<tr>
<td>Math ACT</td>
<td>0.01</td>
<td>0.03</td>
<td>0.52</td>
<td>-0.01</td>
<td>0.11</td>
</tr>
</tbody>
</table>

*Note.* N's for each comparison ranged from 117 to 79. **Bolded** correlations are significant at *p* < .05
<table>
<thead>
<tr>
<th>Post-manip Social Dominance Orientation</th>
<th>Post-manip Math Identification</th>
<th>Pre-Manipulation Testosterone Level</th>
<th>Post-Manipulation Testosterone Level</th>
<th>Assessment One Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-manip Soc Dom Orientation</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-manip Math ID</td>
<td>.09</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Manipulation Test. Level</td>
<td>-.21</td>
<td>.05</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Post-Manipulation Test. Level</td>
<td>-.25</td>
<td>-.12</td>
<td>.80</td>
<td>1.00</td>
</tr>
<tr>
<td>Assessment One Score</td>
<td>-.02</td>
<td>.40</td>
<td>-.02</td>
<td>-.06</td>
</tr>
<tr>
<td>Assessment One Items Answered Per Minute</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment One Avg. Latency for Correct Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment One Avg. Latency for Incorrect Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment One Difficult- Item Score</td>
<td>-.03</td>
<td>.33</td>
<td>-.01</td>
<td>-.09</td>
</tr>
<tr>
<td>Assessment One Avg. Latency for Correct Difficult Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** N's for each comparison ranged from 117 to 79. **Bolded** correlations are significant at \( p < .05 \)
<table>
<thead>
<tr>
<th>Assessment One Avg. Latency for Incorrect Difficult Items</th>
<th>Post-manip Social Dominance Orientation</th>
<th>Post-manip Math Identification</th>
<th>Pre-Manipulation Testosterone Level</th>
<th>Post-Manipulation Testosterone Level</th>
<th>Assessment One Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.05</td>
<td>0.04</td>
<td>-0.09</td>
<td>0.02</td>
<td>-0.31</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.02</td>
<td>0.33</td>
<td>0.02</td>
<td>-0.03</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Items Answered Per Minute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.09</td>
<td>0.14</td>
<td>-0.06</td>
<td>-0.08</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Correct Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.12</td>
<td>-0.14</td>
<td>0.07</td>
<td>0.14</td>
<td>-0.55</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Incorrect Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.06</td>
<td>0.09</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Difficult-Item Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.11</td>
<td>0.30</td>
<td>0.17</td>
<td>0.10</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Correct Difficult Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.06</td>
<td>0.03</td>
<td>0.03</td>
<td>0.06</td>
<td>-0.33</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Incorrect Difficult Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>0.15</td>
<td>-0.07</td>
<td>-0.12</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>Math ACT Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>0.56</td>
<td>0.18</td>
<td>0.17</td>
<td>0.53</td>
<td></td>
</tr>
</tbody>
</table>

Note. N's for each comparison ranged from 117 to 79.

**Bolded** correlations are significant at $p < .05$

(table continues)
<table>
<thead>
<tr>
<th>Assessment One Items Answered Per Minute</th>
<th>Assessment One Avg. Latency for Correct Items</th>
<th>Assessment One Avg. Latency for Incorrect Items</th>
<th>Assessment One Difficult-Item Score</th>
<th>Assessment One Avg. Latency for Correct Difficult Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-.92</td>
<td>.40</td>
<td>-.56</td>
<td>-.51</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>---------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Assessment One Items Answered Per Minute</td>
<td>Assessment One Avg. Latency for Correct Items</td>
<td>.40</td>
<td>1.00</td>
<td>.58</td>
</tr>
<tr>
<td>1.00</td>
<td>-.56</td>
<td>-.20</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note. N's for each comparison ranged from 117 to 79.

**Bolded** correlations are significant at $p < .05$
<table>
<thead>
<tr>
<th></th>
<th>Assessment One Items Answered Per Minute</th>
<th>Assessment One Avg. Latency for Correct Items</th>
<th>Assessment One Avg. Latency for Incorrect Items</th>
<th>Assessment One Difficult-Item Score</th>
<th>Assess. One Avg. Latency for Correct Difficult Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assess. One Avg. Latency for Correct Difficult Items</td>
<td>-.70</td>
<td>.81</td>
<td>.32</td>
<td>-.40</td>
<td>1.00</td>
</tr>
<tr>
<td>Assess. One Avg. Latency for Incorrect Difficult Items</td>
<td>-.56</td>
<td>.39</td>
<td>.94</td>
<td>-.26</td>
<td>.32</td>
</tr>
<tr>
<td>Assessment Two Score</td>
<td>.70</td>
<td>-.61</td>
<td>-.27</td>
<td>.75</td>
<td>-.47</td>
</tr>
<tr>
<td>Assessment Two Items Answered Per Minute</td>
<td>.84</td>
<td>-.71</td>
<td>-.42</td>
<td>.52</td>
<td>-.50</td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Correct Items</td>
<td>-.80</td>
<td>.72</td>
<td>.40</td>
<td>-.50</td>
<td>.51</td>
</tr>
<tr>
<td>Assess. Two Avg. Latency for Incorrect Items</td>
<td>-.37</td>
<td>.27</td>
<td>.44</td>
<td>-.07</td>
<td>.18</td>
</tr>
<tr>
<td>Assessment Two Difficult-Item Score</td>
<td>.55</td>
<td>-.50</td>
<td>-.15</td>
<td>.72</td>
<td>-.38</td>
</tr>
<tr>
<td>Assess Two Avg. Latency for Correct Difficult Items</td>
<td>-.67</td>
<td>.57</td>
<td>.36</td>
<td>-.30</td>
<td>.37</td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Incorrect Difficult Items</td>
<td>-.31</td>
<td>.23</td>
<td>.43</td>
<td>-.04</td>
<td>.14</td>
</tr>
<tr>
<td>Math ACT Score</td>
<td>.34</td>
<td>-.31</td>
<td>-.10</td>
<td>.45</td>
<td>-.30</td>
</tr>
</tbody>
</table>

*Note. N's for each comparison ranged from 117 to 79. Bolded correlations are significant at p < .05*
<table>
<thead>
<tr>
<th></th>
<th>Assessment One Avg. Latency for Incorrect Difficult Items</th>
<th>Assessment Two Score</th>
<th>Assessment Two Items Answered Per Minute</th>
<th>Assessment Two Avg. Latency for Correct Items</th>
<th>Assessment Two Avg. Latency for Incorrect Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment One Avg. Latency for Incorrect Difficult Items</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment Two Score</td>
<td>-0.32</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment Two Items Answered Per Minute</td>
<td>-0.42</td>
<td>0.82</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Correct Items</td>
<td>0.42</td>
<td>-0.76</td>
<td>-0.93</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Incorrect Items</td>
<td>0.29</td>
<td>-0.31</td>
<td>-0.52</td>
<td>0.40</td>
<td>1.00</td>
</tr>
<tr>
<td>Assessment Two Difficult- Item Score</td>
<td>-0.24</td>
<td>0.90</td>
<td>0.60</td>
<td>-0.56</td>
<td>-0.20</td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Correct Difficult Items</td>
<td>0.30</td>
<td>-0.56</td>
<td>-0.81</td>
<td>0.84</td>
<td>0.43</td>
</tr>
<tr>
<td>Assessment Two Avg. Latency for Incorrect Difficult Items</td>
<td>0.21</td>
<td>-0.24</td>
<td>-0.46</td>
<td>0.38</td>
<td>0.90</td>
</tr>
<tr>
<td>Math ACT Score</td>
<td>-0.16</td>
<td>0.50</td>
<td>0.36</td>
<td>-0.07</td>
<td>-0.07</td>
</tr>
</tbody>
</table>

*Note.* N's for each comparison ranged from 117 to 79. **Bolded** correlations are significant at α < 0.05.

(table continues)
<table>
<thead>
<tr>
<th></th>
<th>Assessment Two Difficult-Item Score</th>
<th>Assessment Two Avg. Latency for Correct Difficult Items</th>
<th>Assessment Two Avg. Latency for Incorrect Difficult Items</th>
<th>Math ACT Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment Two Difficult-Item Score</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assess. Two Avg. Latency for Correct Difficult Items</td>
<td>-.40</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assess. Two Avg. Latency for Incorrect Difficult Items</td>
<td>-.10</td>
<td>.41</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Math ACT Score</td>
<td>.45</td>
<td>-.22</td>
<td>-.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Note. N's for each comparison ranged from 117 to 79.*

**Bolded** correlations are significant at *p* < .05