ABSTRACT

TILLMAN, DAVID RAY. Domain Identification, Stereotype Threat, and Mathematics Test Performance: Analysis of TIMSS 2007 by Gender and Race. (Under the direction of Dr. Jessica Decuir-Gunby.)

Stereotype threat literature suggests that domain identification increases susceptibility to stereotype threat. For instance, girls who are highly identified with math should be particularly at risk of performance decrements due to stereotype threat. In analysis of the TIMSS 2007 student surveys and scores, interest in mathematics was negatively correlated with performance. For girls and African-American students, this negative correlation was highly significant in ways that support theoretical literature and previous studies regarding identification and ST. In addition, the dimensions of domain identification were shown to have different relationships with test performance, suggesting that perhaps domain identification should be better measured and not treated as a single factor in relationship with stereotype threat.

Domain Identification, Stereotype Threat, and Mathematics Test Performance:
Analysis of TIMSS 2007 by Gender and Race

by
David Ray Tillman

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements of the Degree of
Doctor of Philosophy

Curriculum & Instruction

Raleigh, North Carolina

2012

APPROVED BY:

_____   _____
Heather Davis, PhD   Thomas Hess, PhD

_____   _____
Jessica Decuir-Gunby, PhD   Margareta Pop, PhD
Chair of Advisory Committee

DEDICATION

To Mary Haven and Teagan, your grandparents were my examples of hard work and

lifelong learning—this dissertation is part of my example to you.

BIOGRAPHY

David Tillman earned both his BA in English and his MEd at Campbell University. Presently, he teaches both Rural Health and Health Policy Management within the Master of Science in Public Health program in Campbell University College of Pharmacy and Health Sciences. He also directs most of the department's community outreach and service-learning projects.

David is the father of two children—Mary Haven and Teagan. Most days when David isn't thinking about social injustice you will find him running barefoot around his neighborhood or playing his guitar on his porch.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Background

High-stakes tests have never played a more significant role in public education in the United States. In response to the requirements for federal funding (including the competitive funding of the recent Race to the Top initiative), state and local education agencies use standardized testing programs to make determinations about student promotion, site-specific and district-wide sanctions, and teacher evaluation. Since the reauthorization of the Elementary and Secondary Education Act in 2000, scores from these tests have also been disaggregated and analyzed according to subgroups in an effort to bring underperforming minority groups into the spotlight of education. In order to maintain educational opportunity for all students, state education agencies and test developers design tests to be "culture-reduced," "culture-fair," or even "culture-free" (Cole & Zieky, 2001). That is, effort is made to minimize test bias, or the extent to which minority groups are disadvantaged by the unessential content and language of tests. However, over the past two decades, stereotype threat theory has posited an additional means by which some subgroups of students may be systematically disadvantaged. Whereas the pursuit of a culture-free test is largely an endeavor of test developers prior to the test's use in classrooms and schools, stereotype threat theory calls attention to social psychological factors during testing and highlights the need for responsible interpretation of the scores after testing.

Although less overt than a culturally-biased test question, stereotype threat is a social psychological predicament wherein the performances of members of a stigmatized group are impaired by the burden of a pervasive, relevant negative stereotype. Under the conditions of threat, a test-taker—who identifies with some group for which there exists a widely known negative stereotype related to the test—may experience evaluation apprehension beyond test anxiety. Even more, the members of stigmatized groups that are theoretically most susceptible to threat are those who also most identify with the task's domain. Tragically, the vanguard of individuals within the stereotypically inferior group who might be most likely to produce stereotype-defying performances is also most susceptible to the threat of the stereotype during the performance of the task—precisely because the task is self-definitional. In our efforts to assess the extent to which "no child is left behind" by public education, stereotype threat literature broadens culture-fair test design to include identifying and minimizing situational primes for stereotype threat during testing, identifying and intervening with students who are most susceptible to threat, and interpreting test data fairly and responsibly.

Statement of the Problem

In experiments, stereotype threat has been produced with a broad range of primes—including subtle practices that are common to most testing situations. Since the earliest theoretical and experimental work around stereotype threat (Steele & Aronson, 1995; Steele, 1997), the extent to which one is susceptible to threat conditions

has been understood to be moderated by the centrality of the task's domain to the identity of the individual. That is, as the importance of performance in a particular domain increases, the member of a stigmatized group has a correlate increase in sensitivity to threat conditions.   However, recent studies have presented evidence that seems to question whether domain identification is, in fact, a moderating variable for stereotype threat (e.g., Biek, 2009).

Purpose

By definition, stereotype threat is a social-psychological phenomenon in which a member of a stigmatized group experiences performance decrements in the completion of a task due to concerns with avoiding the reinforcement of stereotypes related to the task. The performance decrements caused by stereotype threat have been hypothesized to be an important part of the academic underperformance of both minorities (e.g., Kellow & Jones, 2007; Steele & Aronson, 1995) and women in mathematics (e.g., Lesko & Corpus, 2006; Schmader, 2002). Additionally, stereotype threat researchers have suggested that the phenomenon is most likely to negatively impact those members of stigmatized groups who are most identified with the domain of performance (Steele, 1997). That is, female students and minority students for whom mathematics is most central to their identities are most likely to experience performance decrements under conditions of stereotype threat.

This study analyzes the results of a standardized mathematics test and the correlate responses from a student questionnaire to determine in what ways the

students' self-reporting of interest in mathematics, perceptions of mathematical abilities, and perceptions of the usefulness of mathematics are related to performance on the mathematics test. In a premise similar to Smith and White (2001), if there are significant between-groups differences in the correlations between these dimensions of domain identification and test performance, the results would suggest that stereotype threat may have affected the students' test performance and would provide additional evidence for the role of domain identification as a moderator of stereotype threat.

Significance of Study

If stereotype threat literature has relevance for real-world testing situations, then highly identified students who identify with stigmatized groups are likely to experience performance decrements on relevant tests. This phenomenon should affect individuals well before the oft-studied undergraduate population and, therefore, could influence the educational trajectories of many students through altered curricular pathways, college acceptance, major selection, and even school completion (Osborne & Walker, 2006; Faulkner, 2010). Using a large sample of an understudied population-- namely students in middle school--this study will contribute to the literature by providing evidence for establishing developmental contexts in which concerns about stereotype threat and even considerations of domain identification are appropriate. Additionally, using the results of a school-administered test, the study will contribute to the literature on stereotype threat by offering data from non-laboratory contexts that

are close to the real-world classroom conditions in which stereotype threat is thought to be most problematic.

Furthermore, if it can be shown that identification with mathematics is negatively correlated with mathematics performance for students who are members of stigmatized groups, then the study might also suggest that test scores must be interpreted in light of performance decrements that are related to stereotype threat. The disaggregation of test score data and the use of subgroup analysis for high stakes testing programs might be neglecting the very real and important differences in testing conditions—even for students who are sitting next to one another in a classroom. Even more, the interpretation of individual test scores for African Americans and girls might need to account for stereotype threat or risk underestimating the aptitude of the individual. One might imagine that performances on mathematics tests under threat conditions could be keeping qualified African-American and girls out of gifted programs in elementary school, early algebra classes in middle school, higher maths in high school, certain prestigious colleges, and math-centric degree programs. While this study would not establish precisely how scores could be more accurately interpreted (in light of threat), this large-scale analysis of the Trends in Mathematics and Science Study (TIMSS) data could provide important evidence of the need for such a program of research into cultural-fair test interpretation.

CHAPTER TWO

Introduction

If standardized tests are to be culturally unbiased, then the language of tests must be designed to minimize minority cultural disadvantage, but also the testing environment and the interpretation of test scores must account for systematic prejudice. Stereotype threat theory presents the possibility that well-known negative stereotypes may create unfair testing environments and decrease performance for stigmatized groups for a broad range of tests. The difficulty of the theoretical problem starts with the fact that the biased conditions are hypothesized to be the result of psychological burden and, therefore, students in the same room are experiencing very different testing environments. Further complicating the matter is the evidence of individual differences in the experience of stereotype threat. However, a careful reading of the theory and evidence behind stereotype threat acknowledges the individual differences and explains how some students would be expected to experience the threat conditions differently. This chapter particularly reviews the moderating influence of domain identification on the experience of stereotype threat. Beginning with an overview of the relationship between domain identification and stereotype threat, the chapter that follows will define domain identification and consider the development of identification with domains, both generally and specifically with respect to identification with mathematics for African-American students and girls. Then, the theorized moderating influence of domain identification

on stereotype threat is explored and compared to the evidence in the literature of the hypothesized moderation.

Domain Identification as an Understudied Aspect of the Popular Theory

In the past two decades, perhaps nothing else in the social psychology of education has captured the attention of the general public like stereotype threat theory. Stereotype threat has been featured in articles in popular magazines (e.g., *Newsweek* report by Begley, 2000), television news programs (e.g., *20/20* feature by Stossel & Kendall, 2006), and even in nonfiction bestsellers (e.g., an extended discussion in *Blink* by Gladwell, 2005). Though often misunderstood by the public as a wholesale explanation of achievement gaps between groups of students (for a critique of the popular misunderstandings, see Sackett, Hardison, & Cullen, 2004), stereotype threat is an intriguing hypothesis partly because, as a situational threat, it seems less intractable than other explanations for the underperformance of minority groups. In fact, stereotype threat theory provides an explanation for psychological impediments to performance especially for the "vanguard" members of stigmatized groups[1] who are otherwise the most likely to be successful.

That is, stereotype threat is understood to be a situational quandary in which the performance of a member of a stigmatized group is impaired by the existence of a negative stereotype which is relevant to the task. According to the classic stereotype

---

[1] Here, stereotype threat will be discussed as a phenomenon affecting members of stigmatized groups. While interesting laboratory designs have created a stereotype threat situational phenomenon for groups that are without a history of stigmatization (e.g., Leyens et al., 2000; Martiny et al., 2011), it is difficult to imagine a naturally occurring and meaningful example of stereotype threat for anyone other than a member of a stigmatized group.

threat theory, this threat condition is most likely to affect those individuals who are most invested in the task or who are "identified" with the domain of relevance. Shortly following their original research into stereotype threat (Steele & Aronson, 1995), both Steele, in a theoretical article (1997), and Aronson, in chapter-length explication of stereotype threat theory (Aronson, Quinn, & Spencer, 1998), underscored the fundamental relationship between identification with the domain and the experience of stereotype threat. For example, female students who are most invested in mathematics should also be most vulnerable to stereotype threat during math-related tasks. It was suggested by both of these theoretical explanations that a likely consequence of chronic experience of stereotype threat was disidentification with the domain as self-protection (Aronson, Quinn, & Spencer, 1998; Steele, 1997).

While articulations of the theory of stereotype threat have been largely consistent in including domain identification as a critical factor, the evidence presented by researchers regarding the nature of this relationship between domain identification and stereotype threat has been less consistent. According to a recent meta-analysis by Nyugen and Ryan (2008), domain identification was found to be a significant moderator of threat. Indeed, many studies have presented evidence for the moderation of threat in exactly the manner suggested by the original researchers (e.g., Lawrence, Marks, & Jackson, 2010). However, other studies have found the relationship between domain identification and stereotype threat to be insignificant (e.g., Hess & Hinson, 2006) and some researchers have even suggested that the domain identification aspects of

8

stereotype threat theory are misleading (Biek, 2006, 2009). Far more often, stereotype threat research ignores this aspect of the phenomenon and neglects to measure or account for domain identification—even though it is considered to be fundamental in classic stereotype threat theory.

Defining Domain Identification

The notion of domain identification begins in a basic understanding of a separation of self-concept and self-esteem. Here, self-concept refers to one's own sense of competence. In response to a given self-concept, self-esteem is the sum of emotional responses to this appraisal of ability. While in the most basic formulation a low self-concept might result in lowered self-esteem, the various domains of human activity have unequal influence on the self. While slight failures in some domains can be devastating to the self, complete failures in other domains might have little or no impact on the self. Consider, for instance, an individual who is unable to whistle. It is easy to imagine that the complete failure to produce a whistle would have negligible impact to individual's self-esteem. Whistling is a low-value domain. On the other hand, even relatively small differences in high-value domains like physical attractiveness, intelligence, or athleticism might have weighty consequences for one's self-esteem.

According to some researchers, differences between the impacts of performances in particular domains to the global self-esteem are stable and predictable. In particular, Marsh (2008) provides evidence that nomothetic modeling of self-concepts' influence on self-esteem can be as predictive as individual-importance

models. Indeed, as Marsh points out, evidence for individual-importance effects remains "elusive." Even so, for well over a century, most self-esteem researchers have espoused the theory of between-persons differences in the valuing of domains due to its intuitive appeal.

As first articulated by William James (1890/1968), the individual is understood to consider many possible selves and then select those upon which to "stake one's salvation." The consideration of other possible selves necessarily involves the construction of alternative hierarchies of domains of activity. Shared cultural values and social structures give some stability to the hierarchies across individuals—particularly within cultural groups—however, from a Jamesian perspective, there are meaningful individual differences in the valuing of domains. Using the Jamesian theory of differential importance, we might easily make sense of one person's laughter in admitting to a crowd that she "can't carry a tune in a bucket" while accounting for another individual's tears after receiving minor critical feedback from judges in a vocal competition. For the former, failure to sing well had little consequence for her self-esteem. For the latter, as a competitor in a vocal competition, performance in the musical domain was highly valued and feedback related to performance in the domain can greatly impact self-esteem. The ease of generating this example—or James' "puny fellow" or the countless other examples from experience or imagination—results in the aforementioned intuitive appeal of the Jamesian theory of differential importance.

Domain identification is predicated on this theory of the differential valuing of domains in the formation of global self-esteem.

Extending the individual-importance hypothesis, identification with domains is a strategic, self-protective psychological process involving the valuing and devaluing of domain-specific self-concepts in relation to self-esteem. This process of identifying with domains strategically is rooted in the evolution of the Jamesian perspective into the symbolic interactionist theory of self-esteem. The symbolic interactionist theory asserts that individuals receive, attend to, and evaluate social feedback before that feedback impacts global self-esteem (see Osborne & Jones, 2011 for a more detailed discussion of the symbolic interactionist theory of self-esteem as it related to domain identification). In a number of ways, individuals might contextualize, devalue, or even simply be inattentive to rejection feedback in strategic protection of self-esteem. Common theories of self-protective psychological processes would include attribution theory, self-handicapping, and discounting. Somewhat similarly, when an individual perceives patterns of acceptance feedback from performances in a particular domain, the individual might identify with the domain, or make the domain more central to the global self-concept and, subsequently, the global self-esteem. Conversely, if the individual perceives a pattern of rejection feedback from performances in a domain, the individual might disidentify with the domain, minimizing the impact of incompetence in the domain for the global self-esteem.

While this process of identifying or disidentifying with a domain might be understood to be a consequence of social feedback, Osborne and Jones (2011) emphasize that the process is "iterative" and "cyclical" (p. 152). That is, changes in the degree to which an individual is identified with a particular domain alter the individual's subsequent performances in the domain. Such a connection between domain identification and domain-specific outcomes has been drawn between academic identification and academic motivation and outcomes (Osborne & Jones, 2011; Osborne & Rausch, 2001), vocational identification and productivity (Kanungo, 1979; Rabinowitz & Hall, 1977), and identification with fathering and involvement in child-rearing (Pasley, Futris, & Skinner, 2002). In each of these examples, increases in identification with the domain are associated with increases in the frequency of desirable behaviors and desirable outcomes within the domain. Equally, disidentification is associated with increases in undesirable behaviors and decreases in desired outcomes (Morgan & Mehta, 2004; Nosek, Banaji, & Greenwald, 2002).

Domain Identification and Development

Even in a basic Jamesian sense—in which an individual selects from possible selves a commitment-worthy, salvation-staked identity—a coherent identity means movement from identity diffusion to identity achievement. That is, without entering into the specifics of any particular model of identity development, individuals must be understood to engage in a process of exploration and commitments around identity (Marcia, 1966). While research is scarce concerning the development of domain

identification in particular, it seems reasonable to assert that commitments to specific domains of activity occur within the broader contexts of identity formation.

Erikson (1968) explains that the development of a coherent identity is the chief objective of the adolescent stage of development. While Marcia (1966) presented the states of identity achievement in terms of individual differences, Waterman (1982) augmented Marcia's model by suggesting a stable directionality (from identity diffusion to identity achievement) and also situating that process in adolescence. Subsequent research has offered validating evidence to support these basic tenets of Waterman's understanding that adolescence is largely a process of moving from the exploration of identities to commitment to identities (Meeus, 1996; van Hoof, 1999; Berzonsky & Adams, 1999; Kroger, 2007).

Hence, while some meaningful questions might be asked about the extent to which younger children are developing commitments to domains, early adolescence may be the earliest period at which one might rightly begin to consider an individual to be identified with a particular domain. Even with early adolescence, it is much more likely that individuals are engaged primarily in exploration rather than in commitment-making. In these individuals, researchers might best understand the development of identification with domains by considering early emerging indications of commitment.

One of the earliest markers for understanding the centrality of a particular domain to the identity of an individual may be the development of interest. Hidi and Renninger (2006) introduce a model of interest development that suggests that

sustained individual interest grows out of repeated experiences with situational

interest. Contrasting with other models of interest development, this model defines

interest apart from expertise. As such, even very young children can be identified as

having (or developing) an interest and specifically suggests that this interest is an

important early indicator of "investment" or identification in a domain. As Hidi and

Reninger point out, however, this development of interest can be either be enhanced or

disrupted by external forces—especially classroom experiences.

Academic Identification and African American Youth

Stigmatization is an important external force which may have an impact on the

development of interest and, ultimately, identification with a domain. The general

discussion of the relationship between domain identification and stigmatization is

exemplified with consequential specificity in the case of African-American youth and

mathematics. In 2011, African-American workers represented 11% of the total labor

force (US Department of Labor, 2012). However, in careers that require high levels of

competency in mathematics, African-Americans are grossly underrepresented. For

example, in "computer and mathematics," African-American workers represented 6.9%

and, in "architecture and mathematics," African-Americans represented only 5.2% of

the labor force (US Department of Labor, 2012). This underrepresentation, while

disturbing, should not be surprising as African-American students have continually

scored well below all other racial categories on standardized mathematics tests. Even

with all of the attention educators have given to "closing" important achievement gaps,

on the National Assessment of Educational Progress (NAEP) the average scale score of African-American students has lagged 31-41 points behind the average scale score of White counterparts on every administration of the test for more than twenty years (US Department of Education, 2011).

Several years prior to the pioneering work on stereotype threat theory, Steele (1992) proposed that the underperformance of African-American students might be explained to some degree by the fact that many African-American students do not identify with schooling. Perhaps for some students, performance in schooling is not perceived as relevant enough to their adult models and other daily realities to be motivating (Steele, 1992). Other researchers have advanced theories that African-American students associate schooling with White values and subsequently disidentify with academics in an effort to resist assimilation (Fordham & Ogbu, 1986).

Perhaps even more intriguing, however, is research that suggests that African-American students that are highly identified with particular academic domains or with schooling are also likely to underperform (Lawrence, Marks, & Jackson, 2010) or withdraw from schooling (Osborne & Walker, 2006). Theoretically, the cyclical relationship between identification, motivation, performance, and feedback seems to function in ways that ultimately push African-American identities outside of academic domains. That is, although researchers have documented that African-American boys have comparable science- and math-related career aspirations as White peers in early adolescence (Riegle-Crumb, Moore, & Ramos-Wada, 2010), by late adolescence these

groups' realities are very different in ways that result in reduced selection of college majors in mathematics and science and, consequently, underrepresentation in related careers (Alliman-Brisett & Turner, 2010). According to Alliman-Brisett and Turner (2010), these racially defined discrepancies persist as expressions of both personal and institutional racism such that dispositional racial prejudices make academic achievement more difficult and, with the perception of achievement burdened by racism, African-American youth find it more difficult to identify with the White male dominated domain of mathematics.

Identification with Mathematics and Female Youth

Similarly to the manner in which racist stereotypes work to prevent African-American students' achievement in and identification with mathematics and to subsequently reduce access of African-Americans to the lucrative career opportunities in science, technology, engineering, and mathematics (STEM), gender stereotypes function to limit the STEM-career participation by women. However, while mathematics underperformance and disidentification with mathematics appear to be cyclically reinforcing for African-American students, the stigmatization related to gender and mathematics aptitude persists in the face of essentially equivalent levels of academic achievement. Decades ago, researchers were able to note evidence of female underperformance relative to their male counterparts (Hyde, Fennema, & Lamon, 1990). More recently, however, evidence suggests that gender differences on mathematics assessments no longer exist (Hyde, Lindberg, Linn, Ellis, & Williams, 2008;

Hyde & Mertz, 2009). Still, female college students are less likely than equally capable male students to select mathematics majors or careers that require advanced mathematics skills (Xie & Shauman, 2003; Stangor & Sechrist, 1998).

For female students, identification with mathematics may be compromised by persistent—though anachronistic—stereotypes of mathematics and science as masculine domains. Experimental data suggests that cuing feminine identity immediately prior to engaging a mathematics task is enough to create a threatening testing environment (Beilock, 2008). Even with the progress of girls in today's schools, a feminine identity still seems to some degree incompatible with mathematics. A recent study presents evidence that a contributing factor in the preservation of feminine disidentification with mathematics is the stereotype endorsement of one's mother (Tomasetto, Alparone, & Cadinu, 2011). By extension, these findings suggest that important social structures reinforce stereotypical and essentialized gender differences that are no longer evidenced in the actual assessment data.

Domain Identification and Stereotype Threat

For women and African Americans in mathematics, identification with the domain is complicated by racist and sexist psychological barriers. For African-Americans, identification with school (generally) and mathematics (in particular) is confounded by both social messaging and significantly depressed achievement, as compared with White peers (Alliman-Brissett & Turner, 2010). Somewhat differently, girls in school seem to achieve (particularly through middle school) very comparably to

their male counterparts and whatever small differences in achievement may remain fail to be significant enough to explain the wide gaps in participation in STEM majors in college and in STEM careers in adulthood (Simon & Farkas, 2008; Xie & Shauman, 2003). A theoretical framework that may help to explain the relationships between domain identification and stigmatizing social feedback for both African-Americans and women is stereotype threat.

Stereotype Threat. Stereotype threat argues that members of a stigmatized group tend to experience performance decrements on tasks in which there are task-relevant stereotypes that are not the result of inability or general performance anxiety. During the completion of the task, members of stigmatized groups must contend with the possibility of being viewed through the lens of the stereotype, with the struggle against being stereotyped, and with the possibility of reinforcing the stereotype with poor performance. Thus, stereotype threat is a social-psychological predicament in which members of stigmatized groups must contend not only with the anxiety already associated with performance of the task, but also with these additional psychological burdens.

Importantly, the individual need not endorse or internalize the stereotype in the experience of stereotype threat (Steele & Aronson, 1995). Stereotype threat could be experienced even if one believes the stereotype to be untrue both for the individual and for the stigmatized group as a whole. Rather than requiring the endorsement of the stereotype, stereotype threat conditions are produced simply by triggering awareness

of a task-relevant stereotype about a group to which the individual belongs. Indeed,

Steele (1997) explained that the negative stereotypes are threats "in the air," which

permeate social contexts regardless of the belief systems of those directly involved.

Due to the pervasiveness of social stigma, most members of stigmatized groups are

already aware of the negative stereotypes by early adolescence (Devine, 1989).

As such, stereotype threat is a situational quandary in which the "threat in the

air" is made relevant to the task at hand. In experiments, stereotype threat researchers

have been able to prime threat in a variety of ways—including overt and contrived

methods as well as subtle ways that mirror relevant real-world conditions. Among the

more artificial and blatant manipulations of threat conditions, researchers have asked

participants to read scientifically written articles that assert that some outgroup is

superior (Aronson et al., 1999) or to watch gender-typed television commercials

(Davies, Spencer, Quinn, & Gerhardstein, 2002), have asked children to color

stereotype-reinforcing images (Neuville & Croizet, 2007; Ambady, Shih, Kim, &

Pittinsky, 2001), and have cued the stereotype as a part of the written or oral task

directions (Osborne, 2007; Steele & Aronson, 1995). As might be expected, participants

that were assigned to "high-threat" condition tended to perform significantly below

participants in comparison groups; that is, individuals who were also members of the

stigmatized group but were in the "low-threat" condition as well as outgroup members

from both high- and low-threat conditions.

Perhaps surprisingly, much less overt manipulations—such as merely having the participants indicate group membership, such as race or gender, at the outset of an academic assessment (Croizet & Claire, 1998; Steele & Aronson, 1995), completing tasks that are characterized as being diagnostic of ability (Bell, Spenser, Iserman, & Logel, 2003; Gonzales, Blanton, & Williams, 2002; Marx, Stapel, & Mueller, 2005), or simply having women complete mathematics tasks in the presence of men (Huguet, & Regner, 2007; Inzlicht & Ben-Zeev, 2000)—have been demonstrated to produce similarly deleterious impacts on performance. These more subtle manipulations suggest that Steele was correct about the powerful ubiquity of the "threat in the air" and that, by simply calling attention to identification with a stigmatized group, group members are likely to perceive the negative stereotypes as salient. Importantly, many of the more subtle manipulations of threat use experimental manipulations that resemble or recreate relevant classroom and testing conditions.

If the variety of stereotype threat manipulations suggests an all-encompassing breadth of potentially threatening situational contexts, the spectrum of groups and conditions in which stereotype threat effects have been tested suggests that stereotype threat is a situational phenomenon that nearly any individual might experience given the necessary conditions (see Footnote 1). Research has shown, for instance, underperformance due to stereotype threat in individuals throughout the life span: elementary and middle school students (Good, Aronson, & Inzlicht, 2003; Huguet & Regner, 2007), high school students (Keller, 2002; Osborne & Walker, 2006), college

students (e.g., Osborne, 2007; Spencer, Steele, & Quinn, 1999) and even older adults

(Hess, Auman, Colcombe, & Rahhal, 2003). While often researched with regard to

educational contexts, stereotype threat has also been observed outside academic

domains, such as in athletics (Chalabaev, Sarrazin, Stone, & Cury, 2008; Stone,

Sjomeling, Lynch, & Darley, 1999) and in job performance evaluations (Von Hippel, Issa,

Ma, & Stokes, 2011).

      Domain identification as a moderator[2] of stereotype threat. With the

original theory of stereotype threat, Steele clearly associates the experience of

stereotype threat with the "vanguard" of individuals within the stigmatized group who

are most highly identified with the domain of relevance (1997). Since stereotype threat

is by definition a situational predicament in which a negative stereotype is made

relevant for the task-at-hand, the ubiquitous stereotype only becomes self-threatening

to the extent that performance on the task is self-relevant. Hence, stereotype threat is

an acute high-threat situational response to a chronic, generalized condition of

stigmatization. Individuals should, therefore, only experience this acute threat

condition in the performance of tasks within domains that are highly valued for self-

perception. As Steele (1997) explained:

      For the domain identified, the situational relevance of the stereotype is

      threatening because it threatens diminishment in a domain that is self-

---

[2] Many authors inaccurately use moderating and mediating interchangeably. Throughout this article, domain identification is discussed as a moderating variable, not as a mediator of stereotype threat. For a discussion of the distinctions between mediators and moderators, see Baron & Kenny (1986). For an extended discussion of the mediators of stereotype threat, see Smith (2004) and Schmader, Johns, & Forbes (2008).

definitional. For the less domain identified, this recognition is less threatening

or not threatening at all, because it threatens something that is less self-

definitional. (p. 617)

In a more developed modeling of the social-psychological constructs involved in

the experience of stereotype threat, Schmader, Johns, and Forbes (2008) suggest that

performance decrements under stereotype threat conditions are the result of a

"cognitive imbalance" between the triadic concept of self, concept of group, and concept

of ability domain. That is, the situational factors of stereotype threat create

disequilibrium in the relationships between identification with the domain and

identification with the group. While something like self-handicapping can be helpful to

some individuals for managing this imbalance, those individuals for whom performance

in the domain is especially central to the self cannot avail themselves of self-

handicapping strategies (Schmader et al., 2008). Even discounting strategies, whereby

a member of a stigmatized group might protect the self by discounting the validity of

evaluative feedback as biased, are limited as a realistic strategy if the feedback is highly

trusted or consequential for the given domain (Schmader et al., 2001). If the member of

the stigmatized group is to continue to identify with the domain and pursue the

rewards of the domain, the individual needs stereotype-defying performance under the

self-threatening conditions. However, this task is ongoing and impossible, as scores of

stereotype-defying performances cannot dispel the "threat in the air" and a single

failure could reinforce it as self-relevant—particularly as one advances to higher levels (or groups) of expertise (Steele, 1997).

Extending Schmader et al. (2008) triadic model, domain identification exists alongside group identification and task difficulty as important moderators of stereotype threat. Moderating variables are those variables which affect both the intensity and directionality of the relationship between independent variables (membership in a stigmatized group under stereotype threat conditions) and dependent variables (test performance) (Baron & Kenny, 1986). One way in which the individual could reconcile the cognitive imbalance is to reduce identification with the stigmatized group. In other words, the individual essentially suggests that I am unlike my group and therefore will be able to perform in this domain. Racial identification has been shown to be a moderator of stereotype threat for minority groups (Davis, Aronson, & Salinas, 2006). Similarly, studies have reported that women performing tasks under stereotype threat conditions are likely to disavow feminine traits on surveys taken after the priming of the threat (Pronin, Steele, & Ross, 2003; Von Hippel, Walsh, & Zouroudis, 2011). In addition to "identity separation" (Von Hippel et al., 2011) or "bifurcation of identity" (Pronin et al., 2003) as a response to the experience of stereotype threat, identification with the stigmatized group is likely to be iterative and cyclical in relationship to performance of tasks under threat. Those individuals who are highly identified with the group are more susceptible to stereotype threat conditions (Armenta, 2010; Bergeron, Block, & Echtenkamp, 2006; Marx, Staples, & Muller, 2005; Schmader, 2002).

For example, Armenta (2010) reported that highly ethnically identified Latino participants experienced performance decrements on a mathematics task under stereotype threat conditions. Interestingly, this study (Armenta, 2010) also reports that highly ethnically identified Asian participants performed better than comparison groups when racial identity was activated prior to performance—exemplifying perhaps the stereotype boost (Shih et al., 2002) or stereotype lift (Walton & Cohen, 2003) phenomenon. On the other hand, the performances of Asian participants and Latino participants who were measured as low in ethnic identification were not sensitive to the manipulation of stereotype threat (Armenta, 2010). The results of this study suggest that group identification has both responsive and moderating relationships to stereotype threat.

Additionally, group identification seems to interact with domain identification in the experience of stereotype threat by individuals. In a study that used changes in the level of identification with gender as the dependent variable, Pronin et al. (2003) found that only those female participants who self-reported as highly identified with the domain were likely to diminish the importance of stereotypically feminine traits under high stereotype threat conditions. That is, in a more extended design than Armenta (2010), Pronin and colleagues report that the most susceptible individuals to stereotype threat have a convergence of strong identification with both their stigmatized group and the relevant domain . Differences in the centrality of either

group membership or domain identification may help to explain individual differences in response to stereotype threat.

Going even further, an individual's identification with a particular domain may be affected by the centrality of group membership to the individual's identity. In a study that examined the relationship between identification with gender and mathematics identity, Nosek et al. (2002) found that those female participants who were highly identified with gender had predictably lower levels of mathematics identity. As the authors explain:

> The stronger the association is between self and the group (gender identity), the greater the extent to which individual preferences (attitudes) mirror the expectations of the group (stereotypes), even when those preferences appear to be freely chosen. (p. 54)

In related findings that highlight the socially-constructed nature of group membership, Bergeron et al. (2006) report that women who identify not only as less feminine, but actually in stereotypically masculine ways tend to be buffered from the negative effects of stereotype threat and respond to threat conditions more like highly identified male participants.

At yet another level of complexity, domain identification seems to be highly interactive with task difficulty with regard to the individual effects of stereotype threat. According to a study by O'Brien and Crandall (2003), individuals under stereotype threat tend to actually perform better on easy or moderately difficult tasks under threat

conditions. The individuals seem to be primed by the need to produce a stereotype-defying performance and the threat conditions, in turn, have a motivational benefit for performance (O'Brien & Crandall, 2003). However, recent research suggests that the relationship between stereotype threat and task difficulty is complicated by domain identification. The performance boost on easy or even moderately difficult tasks under stereotype threat conditions is only found in those individuals who are highly identified with the domain (Keller, 2007; Lawrence, Marks, & Jackson, 2010). With those less identified with the domain, an exactly opposite pattern is reported as these individuals experience a slight performance boost on difficult tasks in stereotype nullification conditions and performance decrements on easier tasks under stereotype threat (Keller, 2007; Lawrence, Marks, & Jackson, 2010). Apparently, when the difficulty of the task stretches the self-concept, the deleterious effects of stereotype threat manifest only for those individuals who are highly identified with the task's domain.

Of course, another self-protective response in managing the cognitive imbalance under conditions of stereotype threat is for the individual to disidentify with the domain. In this case, the centrality of the domain-specific self-concept to the self-esteem is diminished in response to the endless effort to produce not merely competent performances, but stereotype-defying performances. The burdens of this task, which Steele (1997) refers to as Sisyphean, and the probable resultant performance decrements result in decreasing appraisal of one's ability and the opportunity for success in the domain. Therefore, domain identification theoretically has both

antecedental (as a moderator of threat) and consequential (through self-protective disidentification, over time) relationships to the experience of performance under of stereotype threat conditions.

Review of the Evidence of Domain Identification as a Moderator of Threat

Beginning with the seminal study on stereotype threat, Steele and Aronson (1995) named identification with the domain as an important antecedental feature for stereotype threat. Though not mentioned explicitly by Steele and Aronson (1995) and not fully developed theoretically until two years later (Steele, 1997), Steele (1999) asserted that participants in this first series of studies on the impact of threat conditions on the performance of tasks for minority undergraduates were selected based on high identification with the verbal domain (Steele & Aronson, 1995). The fourth study in the series even included a brief measure of identification with the domain (Steele & Aronson, 1995). Interestingly, the identification scales were completed after the experience of the threat conditions, and a MANOVA exploring relationships between the level of identification yielded insignificant results. Nevertheless, the theorized relationship between domain identification and stereotype threat remained a fixture of the program of research for Steele and Aronson.

Domain identification as a selection criterion. A common way in which domain identification continues to factor into stereotype threat research is that domain identification is used a selection criterion for participants. Since theoretically only those highly identified individuals would be susceptible to the effects of stereotype

threat, studies employed a range of strategies for excluding those who were less

identified with the domain of relevance. For instance, in the first of two experiments,

Aronson et al. (1999) used a combination of self-reporting on a Likert-type scale and a

standardized score threshold as an indication of highly identified individuals. Similarly,

Spencer, Steele, and Quinn (1999) used responses to two Likert-type scale items, a

standardized test score (85th percentile on the SAT-M), and completion of a calculus

class with a grade of at least a "B" as evidence of sufficient levels of mathematics

identification for female participants.

Following these early examples, a number of studies continue to use either scale

items, very high demonstrations of performance in the domain, or a combination of the

two as a part of participant selection (e.g., Ben-Zeev, Fein, & Inzlicht, 2005). Even when

not explicitly measured or factored into the study design, the assumptions of a

relationship between domain identification and stereotype threat can be established

through the researchers' choices. For instance, Good, Aronson, and Harder (2008) used

participants in a rigorous calculus course because the course was a "pipeline" for

majors in hard sciences or mathematics. Clearly, the researchers in this study were

concerned with including only participants for whom performance in the domain

should be self-relevant. In a slight variation on this way of including domain

identification as a selection variable, Williams (2006) was not able to observe the

hypothesized stereotype threat effects and offered the participants' insufficiently low

identification with the domain as a probable explanation for the absence of the expected phenomenon.

Domain identification as a variable in stereotype threat studies. While classic stereotype threat experiments have used domain identification as a selection variable, fewer studies have included measures of domain identification as a variable in the study. Some studies incorporate data regarding domain identification into the data analysis and control for identification when examining interactions between stereotype status and threat condition on performance (Carr & Steele, 2009; Osborne, 2007). Other studies have sought to explore specifically the interaction between the level of domain identification and stereotype threat. These studies are perhaps most helpful in considering the theorized status of domain identification as a moderator of stereotype threat.

In experiment 2 of Aronson et al. (1999), the researchers used Likert-type scales to create a measure of identification with mathematics to test possible interaction effects with threat conditions. Two groups of participants were used for this part of the analysis. The third of participants who most highly identified in the self-reporting on scale items were compared to the third of participants who were least identified on the scale items. Notably, the researchers considered these two groups to be "highly identified" and "moderately identified" since both groups were students in a rigorous mathematics course at a selective university and had mean scores reflecting—even with the lowest third—higher identification than might be assumed from a broader

population. In comparing these two groups, Aronson et al. (1999) found a significant interaction between threat condition and level of mathematics identification, with those participants in the "highly identified" group performing worse on the mathematics task in the high threat condition. Interestingly, the moderately identified group actually performed better under the threat condition.

Similarly, a program of research by Smith and colleagues has reported that stereotype threat effects—and perhaps even stereotype boost effects—are moderated by domain identification. In an early study, Smith and White (2001) found that highly identified male participants did significantly better on a difficult mathematics test than highly identified female participants. At the same time, there was no significant difference between the performance of low-identified men and women in the study. While this interaction between identification and gender on performance echoes the result expected in the stereotype threat theory and in earlier studies, a subsequent study finds that domain identification might also moderate a related phenomenon— stereotype boost. According to Smith and Johnson (2006), men who were low in identification with the domain "choked" under the pressure of the positive-gender stereotype (that men are superior in mathematics). Alternatively, these men performed much better when the stereotype was nullified. In this manner, men who are low in mathematics identification are exact opposites from highly math-identified women and experience perfectly negative mirrored responses to stereotype threat conditions.

In another of these important studies, Leyens et al. (2000) created experimental conditions under which a non-stigmatized group (white males) could be made to experience stereotype threat by evoking a stereotype that men have difficulty processing affective information. The researchers created this stereotype threat condition in which males were primed for this stereotype by test directions that characterized the task as one in which men supposedly performed less proficiently than women. Then, participants completed verbal and affective tasks as well as a post-experimental questionnaire, in which they indicated (on Likert-type scale items) the degree to which they were identified with the affective domain. According to the researchers, "threatened male participants performed more poorly the more they were identified with the task domain" (Leyens et al., 2000, p.1195). This relationship was found to exist in precisely the manner suggested in Steele's theoretical formulation (1997). Interestingly, the study found the opposite to be true of the female participants, whose performance increased in relationship to their degree of identification with the domain.

In a study designed to test the extent to which individuals might seek to distance themselves from stereotype relevant group characteristics, Pronin, Steele, and Ross (2003) included, in Study 2, a measure of domain identification. Using a 4-item scale, domain identification was demonstrated to have a significant interaction with stereotype threat condition. That is, female students who were highly identified with mathematics and received the stereotype priming (reading stereotype-confirming

articles prior to the performance of a mathematics task) were more likely to "bifurcate" their gender identity by disavowing those characteristics most relevant for the gender stereotypes. This study highlights the aforementioned disequilibrium between identities created by stereotype threat, as suggested by Schmader and Johns (2003).

Another study that dealt with the complexity of self and looked at domain identification also attempted to explore the interventions possible within the complexities of identities. In their study, Gresky, Ten Eyck, Lord, and McIntyre (2005) looked at the extent to which priming a diversified self-concept might alleviate stereotype threat effects. By having participants draw concept maps that either included many aspects of self, just a few, or none, the researchers then primed stereotype threat in a classic design and analyzed the value of the diversity of self-concept as an intervention. Results were analyzed according to gender and level of domain identification. There was a significant interaction between gender and level of domain identification on task performance; however these effects were minimized for highly identified women who had drawn complex self-concept maps (Gresky et al., 2005). In this study, Gresky and colleagues present unique results by demonstrating the moderating influence of domain identification on stereotype threat but also providing an example for intervening in this relationship without diminishing domain identification.

Somewhat similarly, Lesko and Corpus (2006) designed a study that included domain identification in a test of whether participants in high stereotype threat

conditions would be more likely to discount the validity of the task. Importantly, though a pre-performance domain identification measure (2-item, Likert-type) was used, the scores were not used as a selection criterion and the sample ranged widely in degree of mathematics identification. The study found that there was a significant interaction between mathematics identification and stereotype threat condition for female participants such that highly identified female participants in the high-threat condition were more likely to discount the test, as compared to those in the low-threat condition and those participants who were less identified with mathematics.

Interestingly, this study also included a post-test measure of domain identification to explore whether a stereotype threat condition might affect the participants' self-reporting of identification with the domain. This attempt to measure a near immediate disidentification yielded no significant results. Using a rare, non-adult sample, Keller (2007) included domain identification in a stereotype threat study that also analyzed the difficulty of test items. Using high school students, participants were asked to perform both easy and difficult mathematics tasks in either a high-threat or low-threat condition. Female students who were highly identified with mathematics (as defined by higher than median scores on a 2-item, Likert-type scale) were more likely to experience performance decrements on difficult items under stereotype threat conditions. In addition, these highly-identified and highly threatened female participants were found to be more successful on easier items as compared to highly-identified female participants in the low stereotype threat condition. This significant

interaction suggests that individuals might experience a useful motivational boost from stereotype threat on tasks that are easily managed. This hypothesis is further bolstered in the study's reporting of a similar, yet opposite interaction effect with the female participants who were less identified with mathematics. These participants seemed to benefit from the stereotype threat condition on difficult items, but not on the easier ones. This finding appears to be consistent with previous research that shows the stereotype threat phenomenon to negatively affect performance on tasks that are difficult (O'Brien & Crandall, 2003).

By contrast, Lawrence, Marks, and Jackson (2010) report evidence that stereotype threat negatively impacts the performance of African-American undergraduates with high domain identification, even on moderately difficult tasks. Using scores from a four-item survey (10-point response scale) to measure domain identification, this study treated the degree of domain identification as a continuous variable and conducted a regression analysis with stereotype threat condition on the number of correct and the accuracy of responses. In the high threat condition, as the identification with the domain for a participant increased, both the number of correct responses and rate of accuracy decreased. No such predictive value of domain identification was evidenced in the low-threat conditions. This study complements Keller (2007) by addressing the concern that perhaps the dichotomized easy and hard test items created a ceiling effect for items that were too easy. In this study, the participants engaged test items that were considered to be of moderate difficulty and

yet the performance decrements for highly identified participants was akin to the patterns already documented for difficult test items. Importantly, Keller (2007) and Lawrence et al. (2010) were targeting different stigmatized groups (female students on mathematics tasks versus African-American students with verbal tasks) and dealt with the domain identification measures very differently.

Studies that question domain identification as a moderator of threat. In the decades since the original research, the studies chronicled in the preceding section have supported the classic theoretical relationship between stereotype threat and domain identification with evidence from testing these assumptions. However, it would be misleading to fail to mention that the body of evidence is far from unanimous in this support. Indeed, a number of studies have investigated domain identification as a moderating variable in stereotype threat research and found little evidence for the relationships that Steele and Aronson (1995) have suggested.

In probably the most prominent critiques of stereotype threat, a series of articles by Cullen and colleagues present evidence of an absence of stereotype threat effects and include domain identification as a feature of their studies. In the first article, Cullen, Hardison, and Sackett (2004) use domain identification as a selection criterion. In keeping with previous studies that use demonstration of high ability as a proxy for identification, the authors used standardized scores as a substitute for a measure of domain identification and reasoned that predicted stereotype threat effects would occur at the high end of the distribution of scores. Using regression analyses that

compared SAT performance to grades and ASVAB performance to job performance

data, Cullen, Hardison, and Sackett (2004) reported that stereotype effects were absent

for the high ability—and, according to the researchers, highly identified test-takers.

In a follow-up study, Cullen, Waters, and Sackett (2006) answer criticism of the

use of score distributions as a proxy for identification and include a more explicit

comparison between levels of identification. For this study, the authors used intended

major in post-secondary studies as a measure of identification. That is, those

individuals who marked on the Student Descriptive Questionnaire (SDQ) during the

administration of the SAT that they were "fairly certain" or "very certain" about an

intended major in mathematics or a math-related discipline were considered by the

study to be "math-identified." In a somewhat dizzying study design in which SAT-M is

used to predict English GPA in college, the authors report a lack of evidence for

stereotype threat effects among the highly math-identified female students. Part of the

reason for the awkward study design is the researchers' desire to test stereotype threat

theory under more real-world, rather than laboratory, conditions. The conclusions of

the two articles is that testing stereotype threat as moderated by domain identification

yields no evidence of significance outside of laboratory environments.

Quite the opposite, Biek has offered a series of studies that critique domain

identification, but in ways that expand the theoretical reach of stereotype threat. In

both his dissertation (2006) and in subsequent studies presented at conferences (e.g.,

2009), Biek reports finding evidence of stereotype threat, but no moderating effect of

domain identification. In these studies, Biek uses the Domain Identification Measure (Smith & White, 2001) as an explicit assessment of domain identification. In Study 1 of his dissertation, Biek found no evidence that highly identified participants experienced greater performance decrements under stereotype threat conditions. Rather, Biek reports that stereotype threat negatively impacted the performance of female participants on mathematics tasks across the range of identification with mathematics. In addition, those female participants who were highly identified with mathematics performed at a higher level (controlling for SAT-M) than lower-identified participants in both threat and non-threat conditions. A similar finding was reported in a conference presentation in which Biek (2009) again used the DIM and found no significant interaction between gender, threat condition, and identification level for participants completing a mathematics task. Biek argues that these findings suggest a broader potential population of individuals who are susceptible to stereotype threat.

Whereas most of stereotype threat literature focuses on relatively young populations—with most samples comprised of children, adolescents, or young adult undergraduates—Hess and colleagues have published a series of articles that explore the stereotype threat phenomenon in older adults (Hess, Auman, Colcombe, & Rahhal, 2003; Hess, Hinson, & Statham, 2004; Hess & Hinson, 2006). In these studies, the domain of relevance is memory-related tasks and the negative stereotype that aging adults have unreliable memories.   In an early study, the researchers reported a significant interaction between the centrality of memory to the participants' identity

and their susceptibility to threat conditions (Hess et al. 2003). However in the follow-up studies, which used similar designs and the same measure of domain identification, Hess and colleagues found no corroborating evidence of the moderating effect of domain identification on stereotype threat (Hess, Hinson, & Statham, 2004; Hess & Hinson, 2006). Hess and Hinson (2006) suggested that perhaps the null findings in the subsequent studies are related to the high degree of domain identification by the group as a whole.

One other study that reports insignificant relationships between domain identification and stereotype threat is also focused on a rarer stereotype—stereotypes about computer competence. In a study with undergraduate participants, McFarland, Lev-Arey, and Ziegert (2003) collected domain identification information using a brief, 3-item self-reporting questionnaire. In analysis, there was no significant correlation between race, condition, and identification with the domain (computer science).

It is important to note that these findings are in addition to the majority of stereotype threat studies that neglect to include domain identification in the study design altogether. If a study is to document a stereotype threat effect in a group of participants, but neither uses some measure of domain identification as a selection criterion for participants nor covaries domain identification in the analysis of data, then the evidence of stereotype threat is essentially from a sample that must be assumed to represent a range of domain identification. Such a result might be interpreted to undermine the importance of domain identification as a moderator of stereotype threat.

Difficulty in Measuring Domain Identification in Stereotype Threat Research

As mentioned previously, many studies incorporate domain identification into the study design by using domain identification as a selection criterion for study participants. Other studies have looked at the interaction between domain identification and the stereotype threat condition as a part of the study design. In both cases, researchers must discriminate between those participants who are highly identified with the domain and those for whom the domain is of less importance to self-esteem. In stereotype threat research, studies have used a variety of approaches in trying to make this discrimination. Unfortunately, a clearly defined and agreed upon method of measuring identification with a domain is not yet a part of the literature. The difficulty of establishing evidence for the relationship between domain identification and stereotype is entangled in the range of approaches to measuring domain identification.

Some studies, particularly early ones, used measures of ability as the primary indication of identification with the domain. These researchers use test scores as a proxy for identification. For example, in two important early studies of stereotype threat (Aronson et al., 1999; Spencer et al., 1999) participants with very high mathematics SAT scores are considered to be highly identified with math. These scores were complemented by simple self-reporting on scale items that measure the degree to which the individual feels that "I am good at math" and that being good at mathematics is "important" (Aronson et al., 1999; Spencer et al., 1999). According to logic of

their use of these test scores and the responses to the self-efficacy scales, individuals

who are able to perform very well in a domain are likely to have already invested

themselves in order to build competence and are also likely to further value the domain

because of their abilities.

However, other researchers point out that identification with a domain and self-

efficacy should not be equated. Ryan and Ryan (2005) argue that self-efficacy stands

apart from "mathematics investment." Similarly, Biek (2006) suggests that there exists

a real distinction between self-efficacy and "true identity." For both of these

researchers, identification with a domain is not only the sense that one is competent in

a domain, but also that one has something at "stake" (in the Jamesian sense) with

performance in the domain. In Steele's seminal work (1997), the facets or dimensions

of domain identification are suggested in a series of questions:

> Are the rewards of the domain attractive or important? Is an adequate
>
> opportunity structure available? Do I have the requisite skills, talents, and
>
> interests? Have others like me succeeded in the domain? Will I be seen as
>
> belonging in the domain? Will I be prejudiced against in the domain? Can I
>
> envision wanting what this domain has to offer? And so on. (p. 616)

Hence, subsequent researchers have suggested that measures of domain identification

must assess not only competence and valuing but also interest and a sense of belonging

(Cullen et al., 2006). Test scores, even when complemented with a couple of scale

items, have little hope of measuring the multi-faceted phenomenon evoked in Steele's questions.

Another approach to the issue of measuring domain identification is using the selection of a course or course of study as a proxy for identification (Cullen et al., 2006; Good, Aronson, & Harder, 2008; Spencer et al., 1999). Compared to performing on a test, selecting a major or even electing to engage the challenge of a particularly rigorous course reflects more completely the complexity of identification with a domain. In choosing a course of study, the individual is choosing to make an investment that likely correlates to a sense of competence, interest, and belonging (Cullen, Waters, & Sackett, 2006).

While these authors admit that idiosyncratic external forces like parental influence might impact these choices in ways that have little to do with identification (Cullen, Waters, & Sackett, 2006), there are broader and more systematic external variables that impact these choices in ways that are consequential for the study of stereotype threat. The selection of individual courses and courses of study are shaped powerfully early in schooling by factors like teacher recommendations, which are well beyond the interest and ability of the individual (Faulkner, 2010). Indeed, the social structures and prejudices of most relevance to the phenomenon of stereotype threat are likely to shape the trajectories of individuals very early in their development (Faulkner, 2010). Additionally, these choices would tend to indicate disidentification that might be partially a consequence of stereotype threat phenomenon (Steele &

Aronson, 1995; Osborne & Walker, 2006). That is, by using something like the selection

of a college major as a proxy for identification with the domain of mathematics,

researchers are likely to be vastly under-representing those individuals for whom

stereotype threat has been an influential phenomenon.

Hence, a more direct method of measuring domain identification would be

preferable—particularly for understanding the extent to which identification acts as a

moderator of stereotype threat. Some studies use responses on Likert-type scales that

are designed to measure identification with the domain. Ranging from just two items

(Aronson et al., 1999; Keller, 2007; Spencer et al., 1999) to seven items (Ben-Zeev et al.,

2005), these measures allow participants to self-report their domain identification.

Most often, these items have not been through a rigorous and systematic scale

development process. Theoretically and psychometrically sound identification

measures for assisting researchers in mapping the interactions between identifications

with domain and other factors would be more valuable for understanding individual

differences in response to stereotype threat (Smith & White, 2001).

A few researchers have developed scales that have been used to measure

identification in the stereotype threat literature. Voelkl's (1996, 1997) Identification

with School measure, Osborne's (1997a; Osborne & Walker, 2006) School Perceptions

Questionnaire (SPQ), Smith and White's (2001) Domain Identification Measure (DIM),

and Dixon & Hultsch's (1984) Metamemory in Adulthood (MIA-Ach) have all been used

in stereotype threat research to measure the participants' degree of identification with the domain.

The Identification with School measure developed by Voelkl (1996, 1997) was introduced into the stereotype threat literature through the experiments of Osborne (2007). Voelkl's scale is intended to measure the extent to which an individual values schooling and feels a sense of belonging in the school environment. Indeed, these aspects of schooling could reflect some of the facets of domain identification outlined in Steele's series of questions (1997). However, in Osborne's (2007) use of the scale in a stereotype threat study, he complemented the use of the Identification with Academics measure with the SPQ in order to also account for "the extent to which an individual defined the self through academic performance, the centrality of the academic self to the overall self, and the extent to which academic outcomes impact the self in substantial ways" (Osborne & Jones, 2011, p. 142).

In order to account for these additional components of academic identification, the SPQ includes a total of 16 items and has been incorporated into a program of research by Osborne and colleagues. Initially published in a 13-item version that assessed academic identification in a sample of community college students, the SPQ was shown to be strongly related to academic outcomes--including grade point average and dismissal from coursework for academic reasons (Osborne, 1997). Subsequently, the measure was incorporated, again along with Voelkl's Identification with Academics measure, to predict the likelihood that high school students would complete schooling

(Osborne & Walker, 2006). Though not specifically a study of stereotype threat phenomenon, the longitudinal study showed that minority students who were higher in academic identification (according to the SPQ) were actually less likely to complete their programs (Osborne & Walker, 2006). Finally, in the only study to incorporate the use of the SPQ in stereotype threat research, Osborne (2007) covaried scores on the SPQ in analyzing physiological indications of anxiety across threat conditions.

Importantly, some researchers have questioned whether "academics" is a proper domain with which to measure domain identification. In a series of articles, Lawrence and colleagues suggest that valuing academics generally is very different from identification with a specific domain—like identification with mathematics (Lawrence & Charbonneau, 2009; Lawrence & Crocker, 2009). The findings from these studies suggest that measuring identification with academics broadly might result in spurious relationships with identification and threat conditions (Lawrence & Crocker, 2009).

Another scale, the Domain Identification Measure (DIM), was specifically developed and used to measure domain identification in stereotype threat literature. In contrast to the SPQ, the DIM developed by Smith and White (2001) addresses specific domains within academics. The 16-item DIM includes separate items for English and mathematics such that the tool can be used to measure the level of identification within those specific academic domains. In the original research with the tool, Smith and White found that highly identified women performed worse than highly identified men on a mathematics task, yet there was no statistical difference between the performance

of low-identified male and female participants.   In subsequent studies, the DIM has

been used to test stereotype threat effects and has produced results that both reinforce

(Gresky et al., 2005; Smith & Johnson, 2006) and question (Biek, 2006, 2009) the

moderating role of domain identification in the experience of stereotype threat. A

version of the DIM has also been developed to assess identification with computer

science, though the measure was not directly used in a stereotype threat study (Smith,

Morgan, & White, 2005).

A final scale used to measure domain identification in stereotype threat research

is the Metamemory in Adulthood Questionnaire (MIA-Ach). This scale was used in the

aforementioned line of research by Hess and colleagues regarding stereotype threat

and memory in older adults. The MIA-Ach is another self-reporting questionnaire that

was developed by Dixon and Hultsch (1984). Over the last three decades, the MIA has

been used in a variety of psychological studies. Since most stereotype threat research is

conducted with younger populations in connection to schooling, the MIA-Ach will have

limited usefulness to most stereotype threat researchers.

Even beyond the types of scales used, other differences in the measurement of

domain identification in stereotype threat research add complexity to the synthesis of

evidence about the nature of the relationship between domain identification and

stereotype threat. One important difference is whether the scores from the scale items

are used as a continuous variable or are dichotomized for ANOVA tests. Many

researchers (e.g., Biek, 2006, 2009; Keller, 2007; Lesko & Corpus, 2006; Smith & White,

2001) use median splits to dichotomize the scores of participants into "high" and "low" levels of identification. The aforementioned study by Aronson and colleagues (1999) used extreme scores to define the "highly identified" as opposed to the moderately identified. On the other hand, Lawrence et al. (2010) used responses to identification scale items as a continuous variable and conducted regression analyses on the identification data and performance data. Citing MacCullum, Zhang, Preacher, and Rucker (2002), these researchers point out that the dichotomization of the data lumps together individuals that differ widely in level of identification, while separating those individuals whose scores are slightly below the median from those scores which are slightly above the median—even though these individuals are likely similar with respect to level of domain identification (Lawrence, Marks, & Jackson, 2010).

In addition to differences in measures used and in the treatment of the data collected, differences in the timing of the collection of the domain identification data further complicates comparisons--even amongst those stereotype threat studies which explicitly measure and compare a range of levels of domain identification. Most studies measure the degree of domain identification on scale items just prior to the performance of the stereotype-relevant task (e.g., Lawrence et al., 2010). However, a number of studies have included collection of identification data after the completion of the task (Biek, 2006; Lesko & Corpus, 2006; Leyens et al., 2000). Leyens and colleagues (2000) found that the manipulation of threat conditions had an impact on level of identification, in that participants indicated lower levels of identification with the

domain in high threat conditions. In subsequent studies, however, this effect of threat

on the self-reporting of identification with domain has not been replicated (Biek, 2006;

Lesko & Corpus, 2006). Interestingly, a test designed to measure the impact of

stereotype threat conditions on post-performance responses to domain identification

scale items resulted in insignificant evidence of stereotype threat (Biek, 2006). In

discussion of the results of the study, the author suggests that perhaps the common

practice of probing participants' domain identification just prior to task performance

might have an impact on priming stereotype threat (Biek, 2006).

These challenges in accurately and reliably measuring domain identification—

including the challenges in establishing what factors should be measured, validating

scales for measurement, and using the data from self-reported questionnaires

consistently—must be met before the relationship between domain identification and

stereotype threat in multiple studies can be compared and interpreted. Unfortunately,

with all of the differences in what researchers are measuring and how researchers are

measuring, it is difficult to endeavor to resolve the broader questions about the nature

of domain identification as a moderator of the stereotype threat phenomenon.

Summary

In the original articulation of stereotype threat theory, Steele (1997) argued that

the phenomenon is likely to be most detrimental to those individuals who are most

highly identified with the relevant domain. The relationship between domain

identification and stereotype threat is such a fundamental assumption of the theory that

from the seminal study (Steele & Aronson, 1995) onward domain identification continues to be used as a selection criterion for participants in stereotype threat studies. However, few studies have actually gathered evidence for the interaction between the centrality of the domain to the individual's identity and the experience of stereotype threat. In the limited number of studies that have explored this interaction, domain identification seems to function as a moderating variable for the experience of stereotype threat. With a few notable exceptions (Biek 2006, 2009; Cullens et al., 2006), highly identified individuals have been shown to be more likely to experience performance decrements when stereotype threat is primed. In addition, domain identification interacts dynamically with other moderating variables--such as task difficulty and identification with the relevant group—to shape the individual's response to threat conditions. For example, highly identified individuals tend to perform even better on easy or moderately difficult tasks under stereotype threat conditions. Also, highly identified individuals are more likely to disavow group traits in the presence of threat.

CHAPTER THREE

Research Design

The performance decrements caused by stereotype threat have been hypothesized to be an important part of the academic underperformance of both minorities (e.g., Kellow & Jones, 2007; Steele & Aronson, 1995) and women in mathematics (e.g., Lesko & Corpus, 2006; Schmader, 2002). Additionally, stereotype threat researchers have suggested that the phenomenon is most likely to negatively impact those members of stigmatized groups who are most identified with the domain of performance (Steele, 1997). That is, female students and minority students for whom mathematics is most central to their identities are most likely to experience performance decrements under conditions of stereotype threat. This study analyzes the results of a standardized mathematics test and the correlate responses from a student questionnaire to determine in what ways the students' self-reporting of interest in mathematics, beliefs about mathematical abilities, and beliefs about the usefulness of mathematics are related to performance on the mathematics test. In a premise similar to that of Smith and White (2001), if there are significant between-groups differences in the correlations between these dimensions of domain identification and test performance, the results would suggest that stereotype threat has affected the students' test performance and would provide additional evidence for the role of domain identification as a moderator of stereotype threat.

Through a variety of experiments, researchers have presented evidence that even minimal priming can make a ubiquitous stereotype relevant for the task-at-hand. Several of the more subtle approaches used in controlled stereotype threat experiments have created stereotype threat conditions with manipulations that are common-place for standardized testing—including characterizing a test as diagnostic of ability in the test directions (e.g., Bell, Spenser, Iserman, & Logel, 2003) and having members of stigmatized groups perform tasks in the presence of outgroup members (Huguet & Regner, 2007; Inzlicht & Ben-Zeev, 2000). This study will analyze the results of mathematics tests given in the Trends in International Mathematics and Science Study (TIMSS). The administration of the TIMSS avoids some pervasive testing practices that have been shown to cue stereotype threat, including having test-takers indicate race and gender just prior to testing (e.g., Stricker & Ward, 2004). However, the test directions include statements that cue social comparisons by informing students that the test is being used to compare the performance of students in different countries. In addition to this characterization of the test as useful for comparison, the test is administered in heterogeneous racial and gendered groups. These two elements of the administration, while minimal and certainly typical of school-based standardized testing, are theoretically adequate for producing stereotype threat conditions for members of groups about which negative stereotypes about mathematics performance exist.

Much of the stereotype threat literature works from an understanding that both African-American students and girls are groups that are susceptible to stereotype threat. By comparing the results of selected survey items and the mathematics test items between groups—boys and girls, African-American and White—significant differences could suggest evidence of stereotype threat. Since males and White students should not experience stereotype threat, females and African-American students should have stronger negative correlations between identification and test performance by comparison (Smith & White, 2001). Additionally, since domain identification should—in the absence of stereotype threat—be thought to have no detrimental impact on performance, then negative relationships between level of identification and performance on the mathematics tests might be taken as evidence of stereotype threat effects.

There is some evidence of a waning existence of negative stereotypes about the performance of girls in mathematics. For example, Biek (2006) reports that the majority of participants in his dissertation study failed to produce the hypothetically pervasive stereotype of feminine inferiority in mathematics. By comparing the relationships between mathematics identity and mathematics performance for female subgroups in addition to African-American subgroups, the analysis might suggest whether indeed--to adapt Steele's (1997) suggestion that stereotype threat is a "threat in the air" --the air is clearing for women in mathematics.

Furthermore, the analysis of performance between groups is with an important population—children in early adolescence. Both the development of domain identification and the experience of stereotype threat are understudied in childhood and early adolescence. As Biek (2006) pointed out, studying these phenomena in younger school-aged populations is fraught with ethical concerns—including the possibility of introducing novel information concerning problematic stereotypes. While several studies have cleverly minimized the ethical risks and demonstrated stereotype threat effects even in elementary school students (e.g., Neuville & Croizet, 2007), none of these studies have looked at the relationship between domain identification and performance decrements in children under threat conditions. This study avoids the ethical dilemmas of studying stereotype threat in early adolescence by functioning only as post-experiential analysis; the hypothetical stereotype threat condition is believed to have already been created in the administration of the tests themselves and the secondary data analysis could bring new understanding without the possibility of increasing risk to the tested youth. By analyzing relationships between the mathematics identity and mathematics performance of 8th grade students, the study may provide the opportunity to contribute to the understanding of interactions between identification with the mathematics and performance on mathematics items as individuals develop. Besides being understudied, these stages of development are even more critical for the linkage between stereotype threat and broader social ills than the oft-studied undergraduate population—since many of the theoretical consequences of

stereotype threat (e.g., academic disidentification) have likely profoundly impacted

vulnerable groups prior to adulthood (Osborne & Walker, 2006).

Table 1
Research Questions and Correlate Hypotheses

Research questions	Hypotheses
1. What is the relationship between female students' identification with mathematics and their performance on mathematics tests under possible stereotype threat conditions?	H1. For female students, there will be a significant negative correlation between survey items measuring identification with mathematics and test performance.
2. What is the relationship between African-American students' identification with mathematics and their performance on mathematics tests under possible stereotype threat conditions?	H2. For African-American students, there will be a significant negative correlation between survey items measuring identification with mathematics and test performance.
3. Are there meaningful differences between the domain identification-test performance relationship for members of nonstigmatized outgroups as compared with members of stigmatized groups?	H3. The relationships between domain identification and test performance for Whites and boys will not be significantly negative and may even be positive.
4. Are there meaningful differences in the relationships between the dimensions of domain identification and the sub-domains of mathematics?	H4. For girls and African-American students, there will be a stronger negative correlation between survey items measuring identification with mathematics and more complex sub-domains, like algebra and geometry.

Sampling

The data used in this study is taken from a recent administration of the TIMSS

(2007). Because of the cultural relevance of stereotypes and some variability in the use

of the student questionnaires in other countries, the scores used in this analysis will be

limited to students in the United States. Over the previous administrations of the test,

school districts from 13 states have participated in the TIMSS (Connecticut, Idaho, Illinois, Indiana, Maryland, Massachusetts, Michigan, Missouri, North Carolina, Oregon, Pennsylvania, South Carolina, and Texas). While not a comprehensive reflection of the United States, the sampling selection process for the TIMSS ensures that the schools, classrooms, and students selected are demographically representative of the nation (for a detailed explanation of the sampling procedure, see Williams et al. 2009).

An important consideration for the design of this study is ensuring a sample size that is large enough to allow for the requisite analyses. Even with the exclusion of data from participants outside of the United States, the sample is large enough to conduct both exploratory factor analysis, testing for factorial invariance, and structural equation modeling of the latent variables on the test performance. Due to the fact that between groups comparisons are essential to the research design, it is important that the sample sizes of relevant subgroups also be large enough to allow for factor analysis and modeling of disaggregated groups. With the listwise deletion of those with missing data, the total sample size was 7050 students. As one might expect, the gender categories each represented roughly half of the sample, with 3578 boys and 3472 girls. Additionally, the size of each of the racial subgroups was adequately large for factor analysis, with 900 African-American students and 3752 White students.   These racial categories included both girls and boys. With thousands of scores in each of the pertinent subgroups of students, the data is sufficiently large for separate modeling and subsequent comparisons of racial and gender subgroups.

Instrumentation

Both the mathematics assessments (and science assessments) and the student questionnaires used in TIMSS 2007 were developed by The International Association for the Evaluation of Educational Achievement (IEA). The items on the mathematics and science assessment are intended to measure topics or content that students are expected to know and also the cognitive skills that students should be developing (Williams et al., 2009). As the *TIMSS 2007 US Technical Report* explains:

> In TIMSS 2007 the cognitive domains are *knowing*, *applying*, and *reasoning* and are common to each grade and subject area. The mathematics topical or content domains covered at grade four are *number, geometric shapes and measures*, and *data display*. At grade eight, the mathematics content domains are *number*, *algebra*, *geometry*, and *data and chance*. (Williams et al., 2009, p.64)

For TIMSS 2007, a total of 353 items are included in the assessment booklets, but these items are presented in a rotated block design so that each student encounters mathematics and science items, though no single student responds to all of the items (Williams et al., 2009). The integrated mathematics and science content are intended to require 90 minutes for 8th grade students. Calculators are permitted for use during the assessment by 8th graders (Williams et al., 2009).

The student questionnaire has evolved with each administration and has included different items over time. As with the mathematics and science assessment

items, each item on the student questionnaire is field tested and evaluated prior to use

(Williams et al., 2009). Again, according to the *2007 TIMSS US Technical Report*:

> The questionnaires requested information to help provide a context for the
>
> performance scores, focusing on such topics as students' attitudes and beliefs
>
> about learning, their habits and homework, and their lives both in and outside of
>
> school. (Williams et al., 2009, p. 66)

The student questionnaires are bound within the assessment booklets and students

responded to the items in the questionnaire after the testing period for the mathematics

and science assessment.

Data Collection

The data files for the scores for the mathematics assessment and the

corresponding survey results are publicly available in a suitable format for the analyses

required for this study. The variables of interest include the plausible values for the

performance on the mathematics assessment, student-reported race and gender, and

student responses to particular items on the questionnaire that should be useful in

measuring relevant aspects of domain identification.

Racial groups for comparison will be "Black or African American" and "White."

While some data suggests that Latinos are also subject to stigmatization around

mathematics (Schmader & Johns, 2003), there is less clarity about the pervasiveness of

relevant stereotypes with this group—especially as compared to the prototypical

groups of interest in math-related stereotype threat research, female and African-

American students. Additionally, due to variations in the reporting of race across administrations, data about the Latino group would be difficult to establish and compare.

According to stereotype threat researchers, domain identification means not only a sense of self-efficacy, but also interest in the domain and beliefs that the rewards of the domain are important (Biek, 2006; Cullens, Waters, & Sackett, 2006; Steele, 1997). Several questions on each administration of the TIMSS student questionnaire address these dimensions of domain identification. In measuring interest in mathematics, responses to questions about "liking" and whether or not mathematics is "boring" will be included for analysis. To measure beliefs about efficacy or abilities, responses to items that ask students to indicate whether they "do well" in mathematics or whether "mathematics is more difficult for me than for many of my classmates." Finally, to assess the degree to which students find the rewards of the domain to be useful or important, the questionnaire asks students—particularly in 8th grade—to rate the extent to which they agree with statements that characterize mathematics as "important to everyone's life" or necessary for accessing a preferred career or college (see Table 2 for the exact items used).

Table 2

Selected Items from TIMSS 2007 Student Questionnaire

I usually do well in mathematics.
I would like to take more mathematics in school.
Mathematics is more difficult for me than for many of my classmates.
I enjoy learning mathematics.
Mathematics is not one of my strengths.
I learn things quickly in mathematics.
Mathematics is boring.
I like mathematics.
I think learning mathematics will help me in my daily life.
I need mathematics to learn other school subjects.
I need to do well in mathematics to get into the university or college of my choice.
I need to do well in mathematics to get the job I want.

Data Analysis

In the first stage of data analysis, the large dataset will be cleaned, paying particular attention to missing data. Contrary to what might be expected with a large dataset, less than 5% of the cases are missing data on the variable of interest. Therefore, cases with missing data will simply be deleted listwise.

Then, a randomly-selected half of each subset of the sample will be used for factor analysis to establish the proxy measures for the dimensions of domain identification. Using SPSS, exploratory factor analysis will substantiate these dimensions of domain identification measured by selected items from the student questionnaire. That is, with minimal cross-loadings, survey items will group in ways analogous to theoretical components of domain identification—including beliefs about

ability in mathematics, interest in mathematics, and beliefs about the usefulness of mathematics. In the second phase of data analysis, the remaining half of each subset of the sample will be subjected to a confirmatory factor analysis—again using SPSS(AMOS)--to verify factor invariance. By constraining the factor structure according to the model that emerged from EFA and then running the model for each of the comparison groups, resultant multi--group replicability will indicate that the factor structure is stable enough for between-groups comparisons.

Following the factor analysis, the entire sample will be used in a structural equation model that relates each dimension of domain identification (interest, self-efficacy, and usefulness) to mathematics performance (see Figure 1). The variable "mathematics performance" will include each of the five imputed plausible values per student. Due to the rotated-block design, each student completes too few items for the assessment to render a content scale score for each student. Hence, the SEM will include each of the five plausible values reported for each student in the sample.

Interest in Math

Mathematics

Performance

Perceptions of Ability
in Math

Perceptions of the
Usefulness of Math
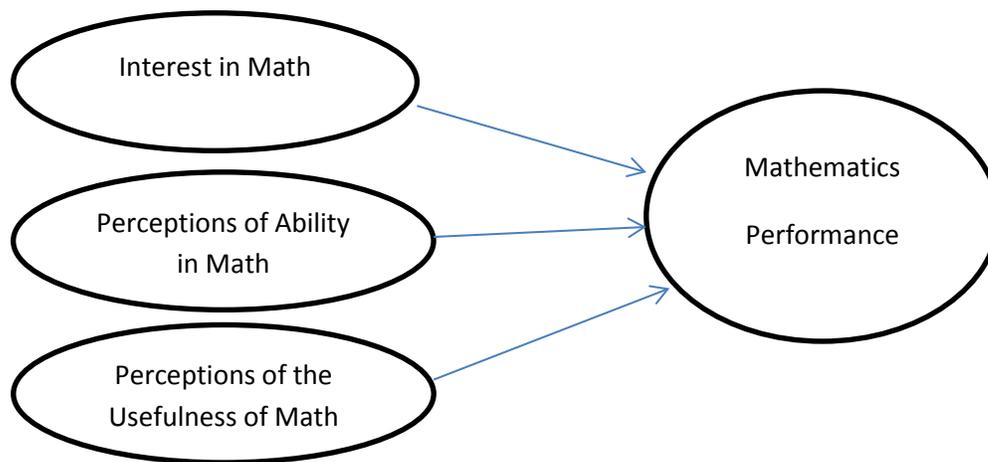
Figure 1. *Theoretical Model for Identification and Performance*

A structural equation model will be used to compare the relationship between the latent constructs (dimensions of domain identity) and mathematics performance across groups that would theoretically be susceptible to stereotype threat (i.e., women and minority students) and those groups that would not experience stereotype threat (especially White males). If a construct loads significantly negatively for a stigmatized group, but not for non-stigmatized group, the results might be interpreted as indications of the influence of stereotype threat on performance—due to the interaction between group membership and the relationships between constructs (Smith & White, 2001). The use of SEM will avoid the simple dichotomization of domain identification data that is typical of many stereotype threat studies (e.g., Biek, 2006, 2009; Keller,

2007; Lesko & Corpus, 2006; Smith & White, 2001), but is criticized by others (e.g.,

Lawrence, Marks, & Jackson, 2010). Additionally, the structural equation modeling may

illumine relationships between aspects of domain identification and test performance

under threat conditions. Beliefs about one's ability in mathematics might be still

positively correlated with test performance while interest in mathematics loads

negatively.

Validity/Reliability

During the first phase of data analysis, the exploratory factor analysis will be

used to establish the dimensions of domain identification that are measured with the

items for the student questionnaire. The validity of the factor structure from the EFA

will be demonstrated in the minimal cross-loading of items between factors. Also,

communalities for each item and eigenvalues for each factor will be reported to

estimate the amount of variance actually being explained by the factor structure for

each item and for the set of items, respectively.

Additionally, the confirmatory factor analysis will provide evidence for factor

stability across the groups. Multiple group confirmatory factor analysis will assess

whether a set of observed measures reflects a latent variable equally well across the

racial and gender groups in the sample. By constraining the model for confirmation,

factor loadings are required to be equal for each group. The observed items are

understood to be valid across the groups if the model fit (when constrained) is not

significantly worse.

In determining to what extent the suggested relationships in the theoretical model are compatible with the real data, the validity of the structural equation modeling will be established using the chi-square (χ^2) test, the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Normed Fit Index (NFI) values (as suggested by Chen et al., 2001). Only after the model is validated by these tests of the model fit is it appropriate to interpret the loadings of each construct as trustworthy and meaningful.

CHAPTER FOUR

Establishing the Factor Structure of Survey Items

Prior to analyzing the relationship between the dimensions of domain

identification with performance on the mathematics tests by the groups of interest, an

exploratory factor analysis established the appropriateness of the selected survey items

for measuring the latent variables in the theoretical model. According to the theoretical

model there are at least three dimensions of domain identification which should be

present even in middle-school aged individuals. These three dimensions of domain

identification are interest, perceived self-efficacy, and the perceived usefulness of the

domain. Theoretically, under conditions of stereotype threat, any or all of the

dimensions of domain identification might load negatively on performance for

stigmatized groups—review Figure 1. The exploratory factor analysis should establish

the factor structure of the selected survey items in a manner that will allow for

comparison of the "clustering" of items to the theoretical model. Theoretically, the

following items were thought to be proxy measures of interest: "I would like to take

more mathematics in school," "I enjoy learning mathematics," "Mathematics is boring"

(reverse-scored), and "I like mathematics." These items were theoretically proxy

measures of perceived self-efficacy: "I usually do well in mathematics," "Mathematics is

more difficult for me than for many of my classmates" (reverse-scored), "Mathematics is

not one of my strengths" (reverse-scored), and "I learn things quickly in mathematics."

Finally, the following items were used as proxy measures of perceptions of the

usefulness of mathematics: "I think learning mathematics will help me in my daily life," "I need mathematics to learn other school subjects," "I need to do well in mathematics to get into the university or college of my choice," and "I need to do well in mathematics to get the job I want."

The exploratory factor analysis was conducted using maximum likelihood extraction and promax rotation—due the theoretically oblique relationships between the dimensions of domain identification. A scree plot suggests that three factors are indeed a plausible model for analyzing the responses to the selected survey items—see Figure 2.
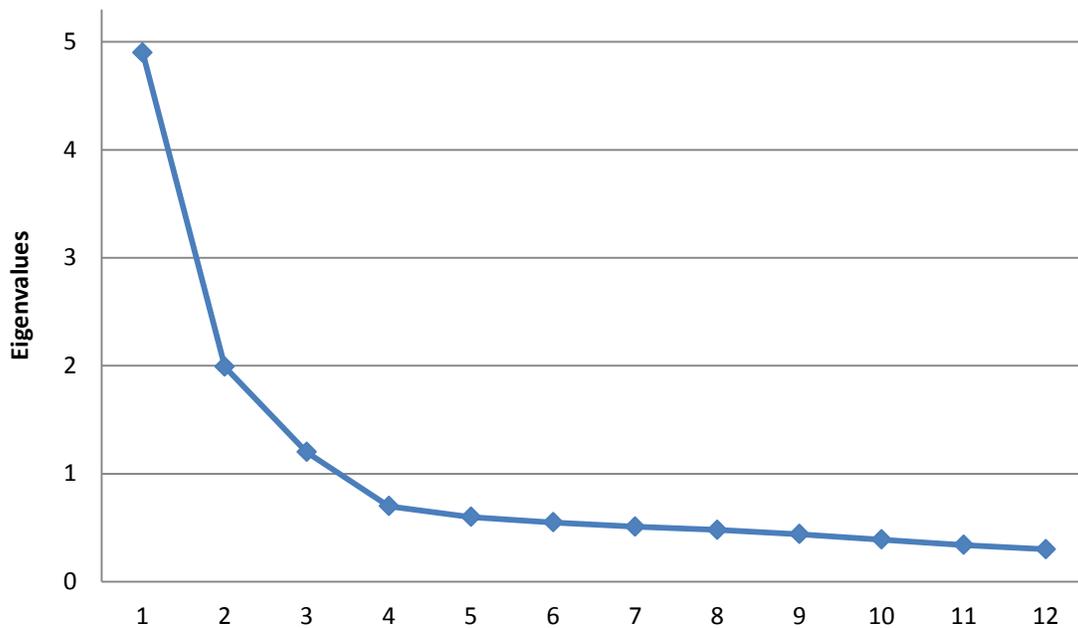

Figure 2. *Scree Plot of Factors*

Since a pattern matrix assists in parsimony by accounting for direct paths to variables, a pattern matrix was preferred over a structure matrix. With minimal cross-loadings, the theoretical factor structure was consistent with the pattern matrix in the exploratory factor analysis. As Table 3 demonstrates, the smallest factor loading was .558 after eliminating loadings of less than .300. The resulting pattern matrix suggests that the 12 selected survey items might be analyzed using three factors and, in turn, used as proxy measures for the three theoretical dimensions of domain identification. Communalities also suggest that each item contributes meaningfully to the model and that all 12 items might reasonably remain included in the analysis.

Table 3
Resulting Pattern Matrix from EFA
(n=3,025)

Item	Factor			Commun-alities
	1	2	3	
I would like to take more mathematics in school.	**.60**	.030	.14	**.47**
I enjoy learning mathematics.	**.87**	.091	.008	**.76**
Mathematics is boring.	**.74**	.090	.053	**.48**
I like mathematics.	**.87**	-.007	.142	**.82**
I usually do well in mathematics.	.008	**.74**	.093	**.64**
Mathematics is more difficult for me than for many of my classmates.	-.069	**.78**	-.083	**.52**
Mathematics is not one of my strengths.	.095	**.75**	-.078	**.61**
I learn things quickly in mathematics.	.007	**.71**	.064	**.57**
I think learning mathematics will help me in my daily life.	.178	-.057	**.56**	**.41**
I need mathematics to learn other school subjects.	.083	-.057	**.59**	**.37**

Table 3 (cont'd.)

I need to do well in mathematics to get into the university or college of my choice.	-.16	.062	**.68**	**.40**
I need to do well in mathematics to get the job I want.	-.046	.003	**.70**	**.47**

Multiple-Group Replicability of the Model

Due to the fact that the research questions require intergroup comparisons, the factor structure that emerged from the exploratory factor analysis must be demonstrated to have a good fit for each of the relevant groups. That is, each group's data must be analyzed using the three-factor structure as established in the exploratory factor analysis, then goodness-of-fit statistics will indicate whether or not the model seems to be trustworthy for each of the groups. The fit indices that will used in confirming the factor structure for all groups will be Chi-Square, Root Mean Square Error of Approximation (RMSEA), Normed Fit Index (NFI), and Comparative Fit Index (CFI).

Due to the ubiquity of Model Chi-Square as a fit statistic, the values are reported for all groups. However, in large samples, the Chi-Square statistic tends to indicate that the model should be rejected, since the calculation of Chi-Square is highly sensitive to sample size (Hooper, Coughlan, & Mullen, 2008; Bentler & Bonnet, 1980; Joreskog & Sorborn, 1993). The smallest of the groups for comparison is the African-American group, but still there were more than 900 African-American participants.

Hence, the used of RMSEA, NFI and CFI is meant to supplement the

interpretations of the Chi-Square statistic with fit indices that are more trustworthy for

large samples. RMSEA is "one of the most informative fit indices" and favors

parsimonious modeling (Hooper et al., 2008; Diamantopoulos & Siguaw, 2000).

According to Steiger (2007), an RMSEA of less than .07 indicates that the model is

properly specified. The NFI is derived from the Chi-Square statistic and, as such, is also

sensitive to sample size. However, the NFI is not recommended to be used in small

samples. The tendency of the NFI to reject the model (in samples of less than 200) is

not of concern for the sample groups in this analysis (all of which are greater than 900).

Even more completely, the inclusion of the CFI, which is a form of the NFI that works

well even with small samples, should add confidence to the interpretation of the

goodness of fit. With both eth CFI and NFI, values range from 0 to 1, with higher values

indicating better fits. Values above 0.95 indicate a good fit in both CFI and NFI (Hu &

Bentler, 1999).

Table 4
Goodness of Fit Indices for CFA

	Chi-Square	RMSEA	CFI	NFI
Boys	413.04	.049	.98	.98
Girls	418.42	.049	.98	.98
White	388.66	.046	.98	.98
African-American	129.13	.046	.98	.97

As Table 4 clearly documents, the fit indices (particularly the RMSEA, NFI and

CFI) indicate that the model should not be rejected for any of the groups of relevance

for this study. The RMSEA scored range from .046 to .049, well within the .07 cutoff

suggested by Steiger (2007). The NFI and CFI values are all above .969 which exceeds

the 0.95 minimum for model acceptance (Hu & Bentler, 1999). In addition, the relative

similarities of values between groups suggest that the model operates in the same

manner for each of the relevant groups for comparison. It is reasonable, then, to

interpret the three-factor structure that emerged from exploratory factor analysis as

stable across the groups.

Model Fit and Structural Equation Modeling Across Groups

As explained above, the exploratory factor analysis suggests that the responses

to selected survey items cluster into three factors—theoretically, the dimensions of

domain identification or the three latent variables of interest, perceptions of self-

efficacy, and perceptions of the usefulness of the domain. The confirmatory factor

analysis goes further with the survey items and confirms that the model is adequately

specified for each of the groups of relevance for this study. In order to consider this

study's research questions, however, the survey results must be analyzed in

relationship to the performance of survey respondents on the items of the mathematics

test. Goodness-of-fit indices should also be examined with respect to the structural

equation models that combine the survey items with the test performance.

Since the test items on the TIMSS are given in a rotated-block design, a set of five

plausible values is calculated for mathematics performance. The latent construct of

Mathematics Performance is therefore rendered from the observed variable of all of the

plausible values of the study taken together. Figure 3 shows the theoretical model that

includes both the dimensions of domain identification and also mathematics

performance.
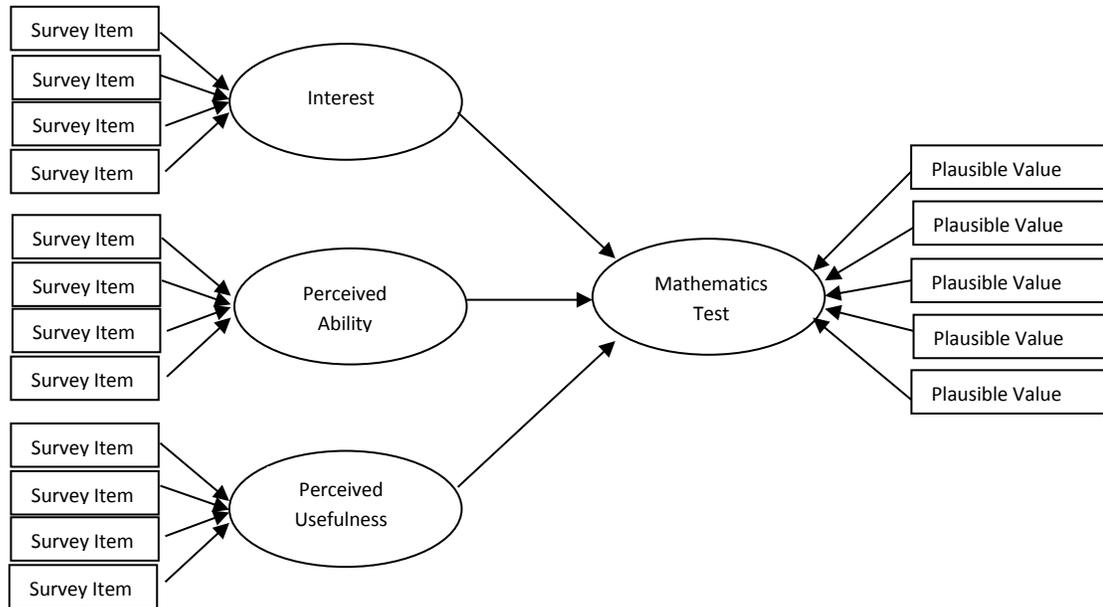
Figure 3. *Structural Equation Model*

Using the same goodness-of-fit measures as with the confirmatory factor

analysis, the structural equation model can be examined for the adequacy of model

specification across the relevant groups. Again, Chi-Square is reported due to its

traditional role in determining fit, but with the large sizes of each group the Chi-Square

statistic is balanced with the RMSEA, NFI and CFI. As Table 5 shows, the model that

examines the relationship between the dimensions of domain identification and

combined, overall mathematics performance had scores on each of the indices that far exceed the acceptable levels.

Table 5
Goodness-of-fit Indices for SEM

Group	Chi-Square	RMSEA	CFI	NFI
Boys	643.91	.038	.99	.99
Girls	708.44	.040	.99	.99
White	576.30	.034	.99	.99
African-American	268.22	.041	.99	.98

RMSEAs for each of the groups ranged from .034 to .042 (below the .07 cutoff) and none of the NFIs or CFIs measured below .98 (above the desired .95). These fit indices suggest that even though Chi-Squares might suggest rejection of the model, that indeed this result is biased by the sizes of the samples and that the model is very likely to be a good fit for the data and a parsimonious structural equation model.

Going even further, the TIMSS data includes plausible values calculated for each sub-domain of mathematical knowledge on the test—Algebra, Data & Probability, Numeracy and Geometry. Table 6 presents the fit indices of running the model with each comparison group within each of the subdomains of mathematics. The scores for each measure of the goodness-of-fit were again quite similar to one another across groups and all met the requirements for accepting the model as properly specified.

Table 6

Goodness-of-fit for SEM, Across Subdomains of Mathematics

	Chi-Square	RMSEA	CFI	NFI
Boys				
Algebra	654.36	.039	.98	.99
Data	643.50	.038	.99	.98
Numeracy	641.57	.039	.99	.99
Geometry	659.61	.040	.98	.98
Girls				
Algebra	627.84	.038	.98	.99
Data	716.19	.040	.98	.98
Numeracy	668.48	.039	.98	.99
Geometry	731.71	.041	.98	.98
White				
Algebra	268.22	.041	.98	.99
Data	582.39	.035	.99	.99
Numeracy	581.04	.035	.98	.98
Geometry	618.49	.036	.99	.99
African-American				
Algebra	233.93	.039	.97	.99
Data	241.26	.038	.97	.98
Numeracy	244.18	.038	.97	.98
Geometry	272.68	.042	.97	.98

Domain Identification and Test Performance, by Gender

The first of the research questions for this study inquires as to whether stereotype threat conditions might have existed to such an extent that the level of identification with mathematics might be negatively correlated with mathematics test performance. Central to the theoretical framework of this study is the possibility that certain dimensions of domain identification might emerge earlier in development or might have differing relationships with the performance of the task. In this case, it was hypothesized that girls should experience negative correlations between at least some

of the dimensions of domain identification, while boys might experience a positive

correlation under the same conditions.

As presented in Figure 4, the dimensions of domain identification had quite

different relationships with girls' performance on the mathematics test. Interest and

Perceived Self-Efficacy had significant standardized regression coefficients--with β = -

.32 (p<.0001) and β = .73 (p<.0001), respectively. On the other hand, Perceived

Usefulness of Mathematics had a non-significant standardized regression coefficient (β

= -.027).
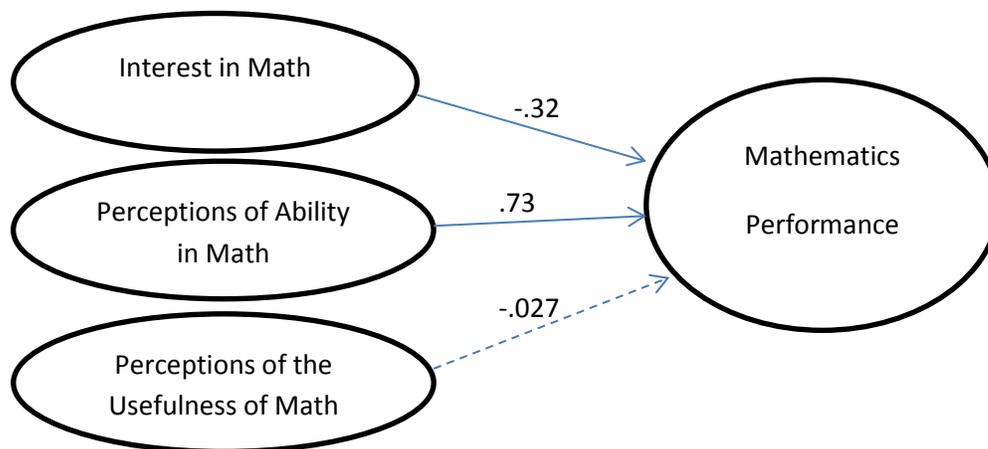
Figure 4. *Dimensions of Domain Identification and Overall Test Performance, for Girls*

Also, it is important to highlight that (though both were significant) the relationships

between Interest and Perceived Self-Efficacy with Test Performance were directionally

opposite—with Interest the only dimension of domain identification that loaded

significantly in the negative correlation, as suggested in the stereotype threat theory.

Contrarily, Perceived Self-Efficacy loaded significantly, but positively with the

performance of girls on the mathematics test.

For boys, while the theoretical model suggests that boys might experience the

test conditions during a standardized mathematics test very differently from their

stigmatized female counterparts, the data suggest that the dimensions of domain

identification actually have relationships to boys' test performance that are similar to

female participants—see Figure 5.
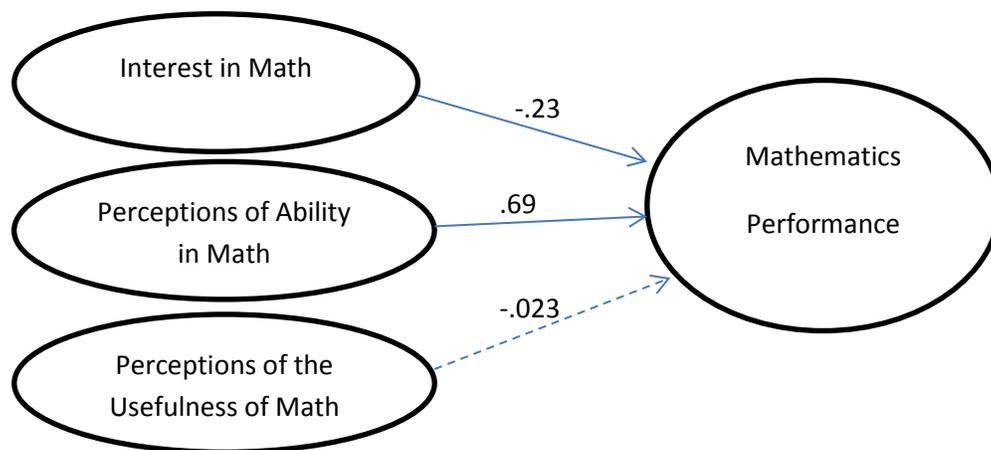
Figure 5
Dimensions of Domain Identification and Overall Test Performance, for Boys

Again with boys, Interest and Perceived Self-Efficacy had significant standardized

regression coefficients—in the case of boys, $\beta = -.23$ (p<.0001) and $\beta = .69$ (p<.0001),

respectively. In the same manner as with girls, Perceived Usefulness of Mathematics

also had an insignificant standardized regression coefficient for boys ($\beta = -.023$). Once

more, it is important to highlight that again the relationships between Interest and

Perceived Self-Efficacy with Test Performance were directionally opposite—with

Interest loading significantly negatively, in the opposition to the theoretical relationship

proposed in stereotype lift theory (significantly positively) or even simply an

insignificant relationship (as might have been implied by stereotype threat theory). As

might have been expected, Perceived Self-Efficacy loaded significantly, but somewhat

surprisingly had a slightly smaller standardized regression coefficient than for girls—

with $\beta = .69$ as compared with girls' $\beta = .73$.

Therefore, in comparing the relationships between the dimensions of domain

identification with mathematics test performance along gender lines, only the

dimension of Interest seems to be a point of real contrast. That is, although both for

girls and for boys the correlation was negative, the magnitude of the effect of Interest

was much larger for girls. Indeed, when interpreting the standardized regression

coefficient in terms of its capacity to explain variance (i.e., the square of the β), Interest

seems to explain about 10% of the variance in girls' scores as compared with only 4% of

the variance in boys' scores.

Dimensions of Domain Identification and Test Performance, by Race

The second of the research questions for the study compares the data for white

participants and African-American participants in ways that are similar to the

comparisons made by gender. Across the model the results of between-groups

comparisons for race are also very similar to gender.

In this instance, it would be hypothesized that African-American participants

under conditions of threat may experience negative correlations between domain

identification. White participants would be expected to have neutral or even positive

effects of domain identification. However, as with the gender comparisons, Interest (for

African-American participants) was the dimension of domain identification that was

significantly negatively correlated with test performance . As Figure 6 illustrates, there

was a significant positive correlation between Perceived Self-Efficacy and Test
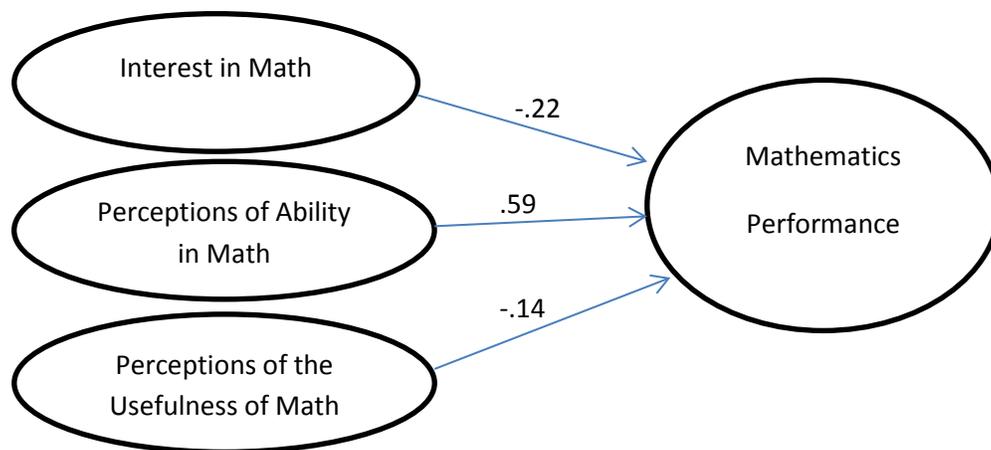
Performance.

Figure 6. *Dimensions of Domain Identification and Overall Test Performance, for African-American Students*

Interestingly, Perceived Usefulness was significant for the African-American group.

Surprisingly, of the four groups for comparison (boys vs. girls and white vs. African-

American) only with African-American participants was the Perceived Usefulness of

Mathematics significantly correlated to test performance and the correlation was

negative. This result would not have been predicted by stereotype threat theory.
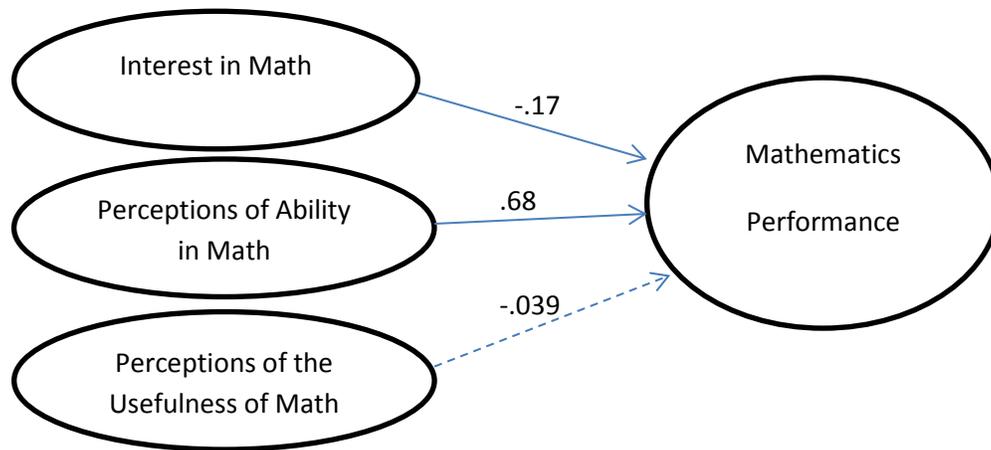
Figure 7. *Dimensions of Domain Identification and Overall Test Performance, for White Students*

The other comparisons between these two racial categories are almost identical to

comparisons between girls and boys. Again with the stigmatized group, the Interest-

Performance relationship had a slightly larger magnitude (African-American, β = -.22;

White, β = -.17) but importantly—and even surprisingly—the directionality was again

the same. The magnitude of the positive relationship between Perceived Self-Efficacy

and Test Performance was lowest for African-American participants (β = .59) and

accounted for 35% of the variance. While this is clearly a significant relationship, it falls

well below the other comparison groups whose standardized regression coefficients

with this particular dimension of domain identification ranged from .68 to .73 and accounted for 46% to 53% of the total variance.

Like with gender, in comparisons of White and African-American participants' relationships between dimensions of domain identification and mathematics test performance, only the dimension of Interest maintains a significant relationship in the theoretical direction. That is, although both for African-American participants and for White participants the correlation was negative, the magnitude of the effect of Interest was slightly larger for African-Americans in the manner suggested by stereotype threat theory.

Domain Identification and Performance on Sub-Domains of Math

The final research question for this study addressed whether domain identification might have differential effects according to the sub-domains of the mathematics test—namely Algebra, Data, Numeracy and Geometry. Theoretically, the stereotype threat effect only manifests on sufficiently difficult tasks. If there are important differences in the complexity of the mathematics within particular sub-domains, then it is conceivable that the stereotype threat phenomenon would affect performance on different parts of the test differently.

However, as Table 7 shows, one is struck not by the differences in the numbers but rather the similarities in the standardized regression coefficients across groups and across subdomains of mathematics.

Table 7

Standardized Regression Coefficients for Groups across Sub-Domains of Math

	Interest	Self-Efficacy	Usefulness
Boys			
Algebra	-.18	.70	-.020
Data	-.25	.63	-.022
Numeracy	-.14	.61	-.048
Geometry	-.19	.70	.081
All	-.23	.69	-.023
Girls			
Algebra	-.22	.71	-.017
Data	-.32	.65	-.036
Numeracy	-.19	.67	-.052
Geometry	-.32	.72	.010
All	-.32	.73	-.027
White			
Algebra	-.22	.59	-.14
Data	-.19	.61	-.029
Numeracy	-.14	.66	-.008
Geometry	-.21	.69	.071
All	-.17	.68	-.039
African-American			
Algebra	-.15	.58	-.13
Data	-.21	.49	-.11
Numeracy	-.16	.58	-.18
Geometry	-.23	.60	-.09
All	-.22	.59	-.14

It is possible to note that Numeracy seems to be affected slightly less by Interest. Also, an interesting, though statistically insignificant, aspect of the data is that Perceived Usefulness is positively correlated with performance in Geometry for three of the four groups (Boys, Girls, and White). Importantly, these notable aspects of the data are non-significant should not be overemphasized. Rather, the overwhelming conclusion of the data is that across subdomains of mathematics the dimensions of domain identification affect performance very similarly.

CHAPTER FIVE

Introduction

      The theory of stereotype threat has, since its genesis, included the belief that domain identification should function as a moderator of threat conditions. The domain identification as moderator was offered as an explanation for some of the individual differences in the presence of a stereotype threat prime. However, a paucity of evidence exists to substantiate the theoretical role of domain identification with stereotype threat. Even more, the nature of domain identification itself is an understudied social psychological construct—particularly in childhood and adolescence.

      In this study, the author investigated the nature of domain identification and its role during a standardized testing procedure that could theoretically produce stereotype threat. Using survey items that were completed along with the performance of the mathematics test, this study analyzes correlations between dimensions of domain identification (specifically Interest, Perceived Ability in the Domain, and Perceived Usefulness of the Domain) and performance on the mathematics test. By comparing groups that should theoretically be sensitive to stereotype threat conditions to non-stigmatized peer groups, differences might indicate both the presence of stereotype threat and the function of the dimensions of domain identification.

      Broadly, the data suggests that the groups actually evinced similar relationships between the dimensions of domain identification and test performance. Perceptions of

the usefulness of mathematics were not significantly related to test performance for these seventh grade students. The students' perceptions of self-efficacy were significantly and positively correlated with test performance. Interest was the only dimension of domain identification that was significantly and negatively correlated with test performance.   The directionality and magnitude of the relationship between Interest and stereotype threat for girls and African-American students suggests confirmation of the theoretical moderating role of domain identification under stereotype threat. While certainly the majority of stereotype threat studies that have included domain identification as a variable have found that domain identification functions in essentially the manner hypothesized, the reality is that the role of domain identification in moderating stereotype threat is an understudied phenomenon and most stereotype threat studies neglect to account for identification with the domain. Thus, the results of this study are of value to the burgeoning literature on stereotype threat.

Gender Differences in the Testing Conditions

Early on, stereotype threat researchers began to investigate whether girls may be susceptible to stereotype threat related to mathematics. Stereotype threat offers an alternative explanation for the gaps in performance along gender lines. This study suggests that girls may be susceptible to stereotype threat in the administration of standardized tests and this susceptibility to threat intensifies for girls who are more interested in mathematics.

In examining the modeled relationship between interest in mathematics and test performance by gender, the standardized regression coefficients for girls, -0.32, was especially interesting as more than 10% of the variance in girls' test performance could be accounted for through this variable. For girls, the relationship between Interest and mathematics test performance has more than twice as much explanatory power than does this relationship for boys, which explains only 5% of the total variance. This difference is large enough to suggest that the girls, as a group, had a very different testing experience from the boys.

Racial Differences in the Testing Conditions

From the first published studies of stereotype threat, a key focus was on the underperformance of African-Americans in academic settings. Much like with girls in mathematics, test results consistently indicate that African-American students are performing well behind their White peers. In this study, the theoretical relationship between domain identification and stereotype threat for members of stigmatized groups under threat conditions shows that African-American students experienced testing differently from White students.

The differences between African-American students and White students were smaller than the gender-based contrasts. African American students showed a stronger overall negative correlation between Interest and test performance than did their White counterparts. Again, using the squared standardized regression coefficient as a

measure of the percentage of total variance in test scores explained by the variable, for African-American students roughly 5% of the total variance as compared to less than 3% for White students. Admittedly, for both racial groups the relationship was fairly weak, but again the variable of Interest had nearly twice the explanatory power for the stigmatized minority group. Again, for African-American students the testing condition seems to be markedly different.

Considerations Across the Comparison Groups

The results of this study suggest—at least with the dimension of one's interest in the domain—that domain identification can be negatively associated with test performance under conditions of stereotype threat. Somewhat surprisingly, this negative correlation between Interest and mathematics test performance was not specific to the test-takers under threat. That is, both males and White students, for whom the stereotype threat conditions were not salient, also evidenced negative associations. However, while the relationship between Interest and test performance was similar across groups, stigmatized groups (girls and African-American students) had greater magnitude as compared with non-stigmatized groups (boys and White students).

This difference is suggestive of support for the study's hypothesis that girls and African-American students would experience negative correlations between domain identification and test performance in ways that differ meaningfully from comparison

groups. However, the correlated hypothesis about non-significant or even positive

correlation between domain identification and test performance for boys and White

students should be rejected. While stigmatized groups had stronger relationships

between the dimensions of domain identification and performance on the task, the non-

stigmatized groups had very similar relationships—including directionality. It seems as

though any stereotype threat effects may simply intensify the impact of identification

with the domain on the task performance for stigmatized groups. In other words, the

experience of stereotype threat seems to make the identification and performance

relationship different in degree, not in kind.

Additionally, the fact that the relationship was so much more significant for

gender-based comparisons than for racially-based contrasts is interesting in light of the

differences in the type of stereotype to which girls and African-Americans are subject.

For girls, the stereotype of mathematical inferiority is specific. Girls are actually

stereotyped to be even more proficient than boys in other school-related domains, such

as behavior or reading. However, for African-American students the stereotype of

mathematical inferiority is nested within a broader sense of stereotypical White

superiority in schooling. It may well be the case that athleticism is the singular school-

related domain in which White students are not stereotypically more capable than their

African American peers. Perhaps the subtle stereotype threat manipulation was in fact

too subtle to produce much of an acute effect in African-American students who were

already immersed in a school environment full of chronic and pervasive stereotypes of inferiority.

In truth, the contrasts were relatively small for both the gender comparisons and racial comparisons. But these small differences were demonstrated in the absence of active stereotype priming at all. In fact, the only two primes for stereotype threat were the presence of stereotypically superior outgroups and the presentation of a task in which performance was evaluated and comparisons were known to be made across groups. The very notion that there would be a negative correlation between measures of interest in mathematics and mathematics test performance would surely be a surprise to many test developers. But that the effect is approximately twice as significant for girls and African-American students means that the test environment might be experienced very differently according to membership in a stigmatized group—not just during stereotype threat experiments but also in the relatively commonplace and increasingly consequential administration of standardized tests in schools.

Subtests of Mathematics and the Identification-Performance Relationships

Because task difficulty is also theoretically a moderator of threat, the author concluded analysis by analyzing the relationships between the dimensions of domain identification and each of the sub-domains of mathematics on the test—Algebra, Data, Numeracy, and Geometry. Across groups and between each of the sub-domains, there was a great deal of similarity in the patterns of values quantifying the relationships

between identification with mathematics (generally) and performance on each of the subtests. While the magnitude of the relationship was noticeably smaller than the other sub-domains in ways that might suggest some confirmation of the hypothesis, the pattern was nearly identical across groups. Therefore, the hypothesis that more complex subdomains might increase the magnitude of the identification and performance relationship for stigmatized groups was unsupported.

Domain Identification as Multi-Dimensional

The results of this study also suggest that understanding the relationship between domain identification and stereotype threat may mean parsing out the dimensions of domain identification. For Steele in the seminal work on stereotype threat theory (1997), domain identification meant that one finds the rewards of the domain attractive or important, perceives the domain as having adequate opportunity, perceives one's self as possessing the requisite abilities in the domain, and several other facets or dimensions of identification. Other researchers have subsequently also included not just a sense of competence, but also interest, valuing, and having a sense of belonging in the domain. However, this study suggests that these dimensions of identification with a particular domain do not operate in a single direction. Rather, perceived ability strongly and positively correlates with performance, while interest in the domain is negatively correlated with performance. In this attempt to study the dimensions of domain identification in relationship to stereotype threat, only one latent variable, Interest, interacted with the threat condition to produce the theorized

performance decrements. Therefore, these results also suggest that combining survey items about interest, ability, and usefulness as a single construct, domain identification, may misrepresent the ways in which these different dimension impact performance under stereotype threat conditions.

Also, consideration of the multiple dimensions of domain identification may be helpful in making sense of domain identification amongst the other known moderators of stereotype threat—including group identification and task difficulty. As Schmader et al. (2008) posited, stereotype threat can be understood has a disequilibrium between these identities in the face of a task. One might imagine that individual differences with respect to the dimensions of domain identification interact with group identification in different ways. For example, perhaps interest explains differences among individuals with both high perceived ability and high group identity. Alternatively, interest may be devoid of explanatory power in the instance in which high interest and high perceived ability occur in an individual with low group identification. This study suggests that a monolithic approach to domain identification may be problematic for understanding its role in moderating stereotype threat.

Limitations

In doing secondary analysis of a data set, the study has important limitations as compared to conceivable experimental designs. One of the most important limitations of the study is the inability to control for group identification with the test-takers. In light of the triadic relationships that create stereotype threat—between group

identification, domain identification, and task difficulty—an ideal study design would consider the variations in the correlation between domain identification and test performance along the spectrum of identification with the stigmatized group. The hypothesized interaction effect would be that as race or gender becomes more central to the identity of African-American and female students (respectively), domain identification will have a stronger negative correlation with test performance. Since none of the items from the Student Questionnaire would serve as proxy measures of racial or gender identity, it is impossible to include consideration of these differences in the analysis. On the other hand, conceptually the inclusion of individuals with wide-ranging group identifications should only reduce the power of the analysis and, therefore, between groups differences would be expected to be even greater.

Also, the items selected from the Student Questionnaire are, at best, proxy measures for the dimensions of domain identification. There is a sense in which the notion of identification means more than showing interest, believing in one's own abilities, and valuing the usefulness of the domain. Domain identification requires that an individual value the domain as an important part of the self. For example, it is possible that a curious and talented young mathematics student might endorse the global value of mathematics, while staking her salvation on music or poetry. This important aspect of domain identification is only suggested in the selected items from the student surveys, not explicitly measured.

Recommendations

In the future, scholars might do well to investigate the relationships within and between the dimensions of domain identification. To begin, theorists might work to better define exactly what the dimensions might be, systematically and exhaustively. In the case of this study, the three dimensions used were merely the three for which data was available. However, future research might identify each dimension and then continue toward the development of a more standard and useful instrument for measuring domain identification. As discussed in Chapter 2, there are wide-ranging ways in which stereotype threat researchers, in particular, assess identification with the domain. While it would be premature to number the dimensions precisely at this point, surely there exists more than two meaningful dimensions of identification with the domain. Hence, the widely used two-question domain identification measure should, moving forward, be considered as too coarse to be of much value.

Even more, this work should be done across the span of human development so that we might understand if the dimensions shift in significance to the individual over time. For example, in reviewing this study, Hess (2012) pointed out that perhaps one could investigate whether one of the dimensions might act as a moderator of another in relationship to a third construct (e.g., test performance). If that relationship is found to , it is plausible that the moderation or interaction could be different as one matures. By studying the dimensions from childhood through adulthood, we might come to know better if there are predictable patterns in these relationships at specific developmental stages. For instance, Hidi and Renninger (2006) suggest that interest may be the first of

the dimensions to emerge. Through systematic study, we might better understand how these dimensions emerge and shift in importance throughout development.

Certainly one of the greatest limitations of this study is that the data was originally collected for a different purpose. However, this is also one of the present study's greatest assets, in that this work is a secondary analysis of a testing event which is a representative example of real-world testing conditions in schools. This is not the work of an experimenter in a laboratory setting with undergraduate students who are receiving psychology class credit. Instead, this is—at least to some degree—a measure of the threat actually "in the air," in the schools, and affecting our community's children.

Researchers should continue to investigate this phenomenon in the less-than-hermetically-sealed conditions of the world outside of university's psychology departments. Stereotype threat and its associations with anxiety (see Osborne, 2007) might even be considered as a public health issue and the ubiquitous threat as an environmental contaminant. Understanding of the phenomenon, like understanding any other pollutant in the air, will require sampling from the environment. This societally-produced contaminant must therefore be subject to analysis within schools, with children and adolescents, and in as much of the authentic contexts as possible. If understanding stereotype threat is confined to college students and controlled experimentation, then the unfortunate consequence may be that the most vulnerable population (namely the children belonging to stigmatized groups) may continue to succumb to the most deleterious effects of threat before even becoming eligible for

inclusion in our undergraduate samples. Even if just in secondary analyses like this

study, more attention should be given to understanding this effect earlier. In particular,

stereotype threat research should continue to be studied in the administration of

standardized tests in public schooling.

Conclusions

In sum, the major findings of the study were that girls and African-American

students experienced the greatest magnitude of performance decrements as Interest

levels increased. In light of the theoretical relationship between domain identification

and task performance, the conditions of standardized testing might be enough to trigger

stereotype threat. Also, the dimensions of domain identification functioned very

differently from one another—with Perceived Ability actually correlating with higher

scores across groups.

As suggested by this study's results, even minimal priming seems to be capable

of creating stereotype threat conditions. These conditions would be relevant for almost

every administration of a standardized, high-stakes test. The effort to produce culture-

fair tests must account for stereotype threat and the fact that in the same room and

with the same tests, students may be experiencing meaningfully different testing

environments. If, as stereotype threat theory suggests, these differences are more than

individual differences in susceptibility to evaluation apprehension or testing anxiety,

and actually occurs in patterns which predictably disadvantage particular stigmatized

groups, then it becomes imperative to better understand stereotype threat in

relationship to testing in order to better protect our vulnerable children.

REFERENCES

Alliman-Brissett, A., & Turner, S. (2010). Racism and math-based career interests,

efficacy, and outcome expectations among African American adolescents. *Journal of Black Psychology, 36,* 197-225.

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in

children: Effects of identity activation on quantitative performance. *Psychological Science, 12*, 385-390.

Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating

role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology, 16*(1), 94-98. doi:10.1037/a0017564

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When

white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*(1), 29-46.

doi:10.1006/jesp.1998.1371

Aronson, J., Quinn, D., & Spencer, S. (1998). Stereotype threat and the academic

underperformance of minorities and women. In J. Swim and C. Stangor (Eds.), *Prejudice: The target's perspective.* Academic Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in

social psychological research: Conceptual, strategic, and statistical

considerations. *Journal Of Personality And Social Psychology*, *51*(6), 1173-1182.

Begley, S. (2000, November 6). The stereotype trap. *Newsweek* pp. 66–68.

Beilock, S. L. (2008). Mathematics performance in stressful situations. *Current Directions in Psychological Science*, *17*, 339-343.

Bell, A. E., Spenser, S.J., Iserman, E., & Logel, C. E. R. (2003). Stereotype threat and women's performance in engineering. *Journal of Engineering Education, 92*, 307-312.

Bentler, P.M. & Bonnet, D.C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88(*3), 588-606.

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

Bergeron, D. M., Block, C. J., & Echtenkamp, A. (2006). Disabling the able: Stereotype threat and women's work performance. *Human Performance, 19*(2), 133-158.

Berzonsky, M. D., & Adams, G. R. (1999). Reevaluating the identity status paradigm: Still useful after 35 years. *Developmental Review, 19,* 557–590.

Biek, D. M. (2006). Stereotype threat and domain identification. Unpublished Ph.D., Cornell University, Ithaca, NY.

Biek, D. M. (2009, May). Domain identification predicts mathematics performance under stereotype threat. Poster session presented at the annual Association for Psychological Sciences Convention, San Francisco, CA.

Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal Of Experimental Social Psychology*, *45*(4), 853-859.

Chalabaev, A. N., Sarrazin, P., Stone, J., & Cury, F. O. (2008). Do Achievement Goals Mediate Stereotype Threat?: An Investigation on Females' Soccer Performance. *Journal of Sport & Exercise Psychology, 30*(2), 143-158.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. & Kirby, J. (2001). Improper solutions in Cornell University, Ithaca, NY.

Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. Journal of Educational Measurement, 38(4), 369–382.

Croizet, J. C., & Claire, T. (1998). Extending the Concept of Stereotype Threat to Social Class: The Intellectual Underperformance of Students from Low Socioeconomic Backgrounds. *Personality and Social Psychology Bulletin, 24*(6), 588-594.

Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology, 89*, 220-230.

Cullen, M. J., Waters, S. D., & Sackett, P. R. (2006). Testing stereotype threat theory predictions for math-identified and non-math-identified students by gender. *Human Performance, 19*(4), 421-440. doi:10.1207/s15327043hup1904_6

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming Images: How Television Commercials that Elicit Stereotype Threat Can Restrain Women Academically and Professionally. *Personality and Social Psychology Bulletin, 28*(12), 1615-1628.

Davis III, C., Aronson, J., & Salinas, M. (2006). Shades of threat: African-American racial

   identity as a moderator of stereotype threat. *Journal of African-American*

   Psychology, 32, 399-417.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled

   components. *Journal of Personality and Social Psychology, 56*, 5-18.

Diamantopoulos, A. & Siguaw, J.A. (2000). Introducing LISREL. London: Sage

   Publishers.

Dixon, R. A. & Hultsch, D. F. (1984). The metamemory in adulthood (MIA) instrument.

   Psychological Documents, *31*, 235-260.

Erikson, E.H. (1968). *Identity: Youth and Crisis*. NewYork: Norton.

Faulkner, V. (2010). Predicting Student Eighth Grade Mathematics Placement. (Doctoral

   dissertation, North Carolina State University). Retrieved from

   http://repository.lib.ncsu.edu/ir/bitstream/1840.16/6456/1/etd.pdf.

   for math-identified and non-math-identified students by gender. *Human*

   Performance, 19(4), 421-440.

Fordham, S. and Ogbu, J. (1986). African-American students' school success: coping with

   the burden of 'acting white'. Urban Review 18(3):176–206.

Gladwell, M. (2005). *Blink: The power of thinking without thinking*. New York: Little,

   Brown and Co.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and

 double-minority status on the test performance of Latino women. *Personality*

 and Social Psychology Bulletin, 28(5), 659-670.

Good, C., Aronson, J., & Harder, J. (2008). Problems in the pipeline: Stereotype threat

 and women's achievement in high-level mathematics courses. *Journal Of Applied*

 Developmental Psychology, *29*(1), 17-28.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test

 performance: An intervention to reduce the effects of stereotype threat. *Journal*

 Of Applied Developmental Psychology, *24*(6), 645.

Gresky, D. M., Ten Eyck, L. L., Lord, C. G., & McIntyre, R. B. (2005). Effects of salient

 multiple identities on women's performance under mathematics stereotype

 threat. *Sex Roles, 53*(9), 703-716.

Harackiewicz, J. M., Durik, A. M., Barron, K. E., Linnenbrink-Garcia, L., & Tauer, J. M.

 (2008). The role of achievement goals in the development of interest: Reciprocal

 relations between achievement goals, interest, and performance. *Journal of*

 Educational Psychology, 100(1), 105-122.

Hess, T.M., & Hinson, J.T. (2006). Age-related variation in the influences of aging

 stereotypes on memory in adulthood. *Psychology and Aging*, *21*, 621-625.

Hess, T.M., Auman, C., Colcombe, S.J., & Rahhal, T.A. (2003). The impact of stereotype

 threat on age differences in memory performance. *Journals of Gerontology:*

 Series B: Psychological Sciences and Social Sciences, 58B, P3-P11.

Hess, T.M., Hinson, J.T., & Statham, J.A. (2004). Explicit and implicit stereotype activation effects on memory: Do age and awareness moderate the impact of priming? *Psychology and Aging*, *19,* 495-505.

Hidi, S., & Renninger, K. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, 41(2), 111-127.

Hippel, C., Walsh, A. M., & Zouroudis, A. (2011). Identity separation in response to stereotype threat. *Social Psychological and Personality Science*, *2*(3), 317-324.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modeling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6* (1), 53-60.

Hu, L.T. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6* (1), 56-83.

Huguet, P., & Régner, I. (2007). Stereotype threat among school girls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99*, 545-560.

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

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A., & Williams, C. (2008). Gender similarities characterize mathematics performance. *Science, 321,* 494–495.

Hyde, J.S. & Mertz, J. (2009). Gender, Culture, and Mathematics Performance. *Proceedings of the National Academy of Sciences*, *106:* 8801-8807.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females

    are susceptible to experiencing problem-solving deficits in the presence of

    males. *Psychological Science, 11*, 365-371.

James, W. (1890/1968). The self. In C. Gordon & K.J. Gergen (Eds.), *The self in social*

    interaction, Vol. 1 (pp. 41-49). New York: Wiley.

Joreskog, K. & Sorborn, D. (1993) LISREL 8: Structural Equation Modeling with the

    SIMPLIS Command Language. Chicago, IL: Scientific Software International.

Kanungo, R. N. (1979). The concepts of alienation and involvement revisited.

    Psychological Bulletin, 86(1), 119–138.

Keller, J. (2002). Blatant Stereotype Threat and Women's Mathematics Performance:

    Self-Handicapping as a Strategic Means to Cope With Obtrusive Negative

    Performance Expectations. *Sex Roles, 47*, 193-198.

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of

    domain identification, task difficulty and stereotype threat on female students'

    maths performance. *British Journal of Educational Psychology, 77*(2), 323-338.

Kellow, T. J., & Jones, B. D. (2007). The effects of stereotypes on the achievement gap:

    Reexamining the academic performance of African American high school

    students. *Journal of African-American Psychology, 34*, 94-120.

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's

    mathematics performance: How implicit gender-mathematics stereotypes

influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology, 43*(5), 825-832.

Kroger, J. (2007). Identity Development: Adolescence through Adulthood, 2nd ed. Newbury Park, CA: Sage, Inc.

Lawrence, J. S., & Charbonneau, J. (2009). The link between basing self-worth on academics and student performance depends on domain identification and academic setting. *Learning and Individual Differences, 19*(4), 615-620.

Lawrence, J. S., & Crocker, J. (2009). Academic contingencies of self-worth impair positively- and negatively-stereotyped students' performance in performance-goal settings. *Journal Of Research In Personality*, *43*(5), 868-874.

Lawrence, J. S., Marks, B. T., & Jackson, J. S. (2010). Domain identification predicts African-American students' underperformance on moderately-difficult tests. *Motivation & Emotion, 34*(2), 105-109.

Lesko, A. C., & Corpus, J. H. (2006). Discounting the difficult: How high math-identified women respond to stereotype threat. *Sex Roles, 54*(1), 113-125.

Leyens, J., Désert, M., Croizet, J., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat?. *Personality And Social Psychology Bulletin*, *26*(10), 1189-1199.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. Psychological Methods, 7, 19–40.

Major, B., Spencer, S. J., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with

negative stereotypes about intellectual performance: The role of psychological

disengagement. *Personality and Social Psychology Bulletin, 24*(1), 34-50.

Marcia, J. E., (1966), Development and validation of ego identity status. *Journal of

Personality and Social Psychology 3*, pp. 551-558.

Marsh, H. W. (2008). The elusive importance effect: More failure for the Jamesian

perspective on the importance of importance in shaping self-esteem. *Journal of

Personality*, 76(5), 1081-1121.

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat:

The effect of self-affirmation on women's intellectual performance. *Journal of

Experimental Social Psychology, 42*(2), 236-243.

Martiny, S. E., Roth, J., Jelenec, P., Steffens, M. C. and Croizet, J.-C. (2011), When a new

group identity does harm on the spot: Stereotype threat in newly created groups.

European Journal of Social Psychology.

Marx, D., Stapel, D., and Muller, D. (2005).   We can do it: The interplay of construal

orientation and social comparisons under threat. *Journal of Personality of Social

Psychology*, **88,** 432–446

McFarland, L. A., Lev-Arey, D. M., & Ziegert, J. C. (2003). An examination of stereotype

threat in a motivational context. *Human Performance*, *16*(3), 181-205.

McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial

reasoning. *Journal of Applied Developmental Psychology, 27*(5), 486-493.

Meeus, W. (1996). Studies on identity development in adolescence: An overview of

    research and some new data. *Journal of Youth and Adolescence, 25*, 569-598.

Morgan, S. L., & Mehta, J. D. (2004). Beyond the Laboratory: Evaluating the Survey

    Evidence for the Disidentification Explanation of Black-White Differences in

    Achievement. *Sociology Of Education*, *77*(1), 82-101.

Neuville, E., & Croizet, J. (2007). Can salience of gender identity impair mathematics

    performance among 7-8 years old girls? The moderating role of task difficulty.

    European Journal of Psychology of Education, 22, 307-316.

Nguyen, H. D., & Ryan, A. (2008). Does stereotype threat affect test performance of

    minorities and women? A meta-analysis of experimental evidence. *Journal Of*

    Applied Psychology, *93*(6), 1314-1334.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Mathematics = male, me = female,

    therefore mathematics ≠ me. *Journal Of Personality And Social Psychology*, *83*(1),

    44-59.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on

    women's mathematics performance. *Personality and Social Psychology Bulletin*,

    29(6), 782-789.

Osborne, J. W. (1997). Identification with academics and academic success among

    community college students. *Community College Review, 25*(1), 59.

Osborne, J. W. (2007). Linking Stereotype Threat and Anxiety. *Educational Psychology*,

    27(1), 135-154.

Osborne, J. W., & Jones, B. D. (2011). Identification with academics and motivation to achieve in school: How the structure of the self influences academic outcomes. *Educational Psychology Review, 23*(1), 131-158.

Osborne, J. W., & Rausch, J. L. (2001). *Identification with academics and academic outcomes in secondary students.* Presented at AERA 2001, Seattle, WA.

Osborne, J. W., & Walker, C. (2006). Stereotype threat, identification with academics, and withdrawal from school: Why the most successful students of colour might be most likely to withdraw. *Educational Psychology, 26*(4), 563-577.

Pasley, K., Futris, T. G., & Skinner, M. L. (2002). Effects of commitment and psychological centrality on fathering. *Journal of Marriage and Family, 64,* 130–138.

Pittinsky, T. L., Shih, M., & Ambady, N. (1999). Identity Adaptiveness: Affect Across Multiple Identities. *Journal of Social Issues, 55*(3), 503-503.

Pronin, E., Steele, C. M., & Ross, L. (2003). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal Of Experimental Social Psychology*, *40*(2), 152-168.

Rabinowitz, S., & Hall, D. T. (1977). Organizational research on job involvement. *Psychological Bulletin, 84*, 265–288.

Riegle-Crumb, C., Moore, C., & Ramos-Wada, A. (2010). Who wants to have a career in science or math?: Exploring adolescents' future aspirations by gender, race/ethnicity. *Science Education, 95*(3), 458-476.

Ryan, K. E., & Ryan, A. M. (2005). Psychological Processes Underlying Stereotype Threat and Standardized Mathematics Test Performance. *Educational Psychologist*, *40*(1), 53-63.

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On Interpreting Stereotype Threat as Accounting for African American--White Differences on Cognitive Tests. *American Psychologist*, *59*(1), 7-13.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's mathematics performance. *Journal of Experimental Social Psychology, 38*(2), 194.

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

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*(2), 336-356.

Schmader, T., Major, B., & Gramzow, R. (2001). Coping with ethnic stereotypes in the academic domain: Perceived injustice and psychological disengagement. *Journal of Social Issues, 57,* 93-112.

Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal Of Personality And Social Psychology*, *83*(3), 638-647.

Simon, R. & Farkas, G. (2008). Sex, class, and physical science educational attainment: Portions due to achievement versus recruitment*." Journal of Women and Minorities in Science and Engineering (14)3*, 269-300.

Smith, J. L., & Johnson, C. S. (2006). A Stereotype Boost or Choking Under Pressure? Positive Gender Stereotypes and Men Who Are Low in Domain Identification. *Basic & Applied Social Psychology*, *28*(1), 51-63.

Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61*(6), 1040-1057.

Smith, J. L., Morgan, C. L., & White, P. H. (2005). Investigating a measure of computer technology domain identification: A tool for understanding gender differences and stereotypes. *Educational & Psychological Measurement, 65*(2), 336-355.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's mathematics performance. *Journal of Experimental Social Psychology, 35*(1), 4-28

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

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

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and

performance. *American Psychologist, 52*(6), 613-629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test

performance of African-Americans. *Journal of Personality and Social Psychology,

69*, 797-811.

Steiger, J.H. (2007). Understanding the limitations of global fit assessment in structural

equation modeling. *Personality and Individual Differences, 42* (5), 893-898.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on

African-American and white athletic performance. *Journal of Personality and

Social Psychology, 77*, 1213-1227.

Stossel, J. & Kendall, K. (Reporters). (2006, September 15). *20/20* [Television

broadcast]. New York, NY: American Broadcasting Company.

Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers'

ethnicity and gender, and standardized test performance. *Journal of Applied

Social Psychology, 34*(4), 665-693.

Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' mathematics performance

under stereotype threat: The moderating role of mothers' gender stereotypes.

Developmental Psychology, 47(4), 943-949.

U.S. Department of Education, Institute of Education Sciences, National Center for

Education Statistics, National Assessment of Educational Progress (NAEP).

(2011). *Average scale scores for mathematics, grade 8 by year, jurisdiction and*

race/ethnicity used to report trends, school-reported: 2011, 2009, 2007, 2005,

2003, 2000, 1996, 1992, and 1990. Retrieved from http://

nces.ed.gov/nationsreportcard/naepdata/

U.S. Department of Labor. (2012). *The African-American Labor Force in the Recovery.*

Retrieved from

http://www.dol.gov/_sec/media/reports/BlackLaborForce/Black

LaborForce.pdf

Van Hoof, A. (1999). The identity status approach: In need of fundamental revision and

qualitative change. *Developmental Review, 19*, 497-556.

Voelkl, K. E. (1996). Measuring students' identification with school. *Educational and*

Psychological Measurement, 56, 760–770.

Voelkl, K. E. (1997). Identification with school. *American Journal of Education, 105*, 294–

318.

Von Hippel, C., Issa, M., Ma, R., & Stokes, A. (2011). Stereotype threat: Antecedents and

consequences for working women. *European Journal Of Social Psychology*, *41*(2),

151-161.

Von Hippel, C., Walsh, A., & Zouroudis, A. (2011). Identity separation in response to

stereotype threat. *Social Psychological and Personality Science, 2*, 317-324

Xie, Y. & Shauman, K.A. (2003). *Women in science: career processes and outcomes.*

Cambridge, MA:Harvard University Press

Walton, G. M., & Cohen, G. L. (2003). Stereotype Lift. *Journal Of Experimental Social Psychology*, *39*(5), 456-467.

Waterman, A. (1982). Identity development from adolescence to adulthood: An extension of theory and a review of research. *Developmental Psychology, 18*, 341-358.

Williams, K. B. (2006). The effects of stereotype threat on test performance of male and female college students. *College Student Journal, 40*(3), 679-684.

Williams, T., Ferraro, D., Roey, S., Brenwald, S., Kastberg, D., Jocelyn, L., Smith, C., and Stearns, P. (2009). *TIMSS 2007 U.S. Technical Report and User Guide* (NCES 2009–012). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.