

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Psychology

2018

EXPLORING COGNITIVE AND NON-COGNITIVE FACTORS IN STEREOTYPE THREAT EFFECTS: AN INDIVIDUAL DIFFERENCES APPROACH

Holden, LaTasha Renee

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp013n2041793>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

**EXPLORING COGNITIVE AND NON-COGNITIVE FACTORS IN STEREOTYPE
THREAT EFFECTS: AN INDIVIDUAL DIFFERENCES APPROACH**

LaTasha Renee Holden

**A DISSERTATION PRESENTED
TO THE FACULTY OF
PRINCETON UNIVERSITY IN
CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY**

**RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF PSYCHOLOGY**

Advisers: Stacey Sinclair and Andrew Conway

June 2018

© Copyright by LaTasha Renee Holden, 2018. All rights reserved.

Abstract

Stereotype threat effects occur when stigmatized individuals are primed with a negative stereotype and underperform relative to a control group. During such situations the pressure of performing coupled with evaluative anxiety may cause even the most capable students to “choke.” This dissertation examines ways stereotype threat effects may be different from one student to the next. Chapter 1 lays the foundation for exploring stereotype threat in a novel way, considering individual differences from cognitive and non-cognitive perspectives—examining the effect beyond the group level. As a cognitive factor, working memory capacity (WMC) has been implicated. Here, the role of WMC is investigated as a mediator and/or moderator of stereotype threat for gender and ethnicity. Chapters 2-4 reveal a general lack of evidence for the threat effect in the form of a performance decrement compared to control. Chapters 2-3 reveal evidence that trait WMC moderates the effect of stereotype threat for ethnicity such that higher WMC is associated with higher predicted scores on standardized tests under threat. Higher scores on non-cognitive factors of mindset, grit, and/or conscientiousness have been suggested to aid performance during threatening academic situations but effects of these generally were not found. Chapter 3 provides no strong evidence that more malleable mindsets are associated with higher predicted scores on standardized tests. Chapter 4 presents evidence that under threat for gender identity participants performed better, revealing no differential effect of malleable versus fixed mindsets under stereotype threat. General discussion considers replication issues within the stereotype threat literature as well as the implications of the present results for future work in this area.

Acknowledgments

As I round the corner and finish what has taken several years of my life, I realize how true it is that it really does “take a village.” So I’ll take a few moments to personally thank my wonderful village.

To the Psychology Department at Princeton, the administrative staff, and all my committee members from generals to my dissertation defense, thank you so much for all you have done to make this possible for me! Andrew Conway, Stacey Sinclair, Nicole Shelton, Susan Fiske, Matt Botvinick, and Justin Junge thank you for your help and advice during various stages of my doctoral studies. Alin Coman and Ken Norman thank you for working with me on my tutorial papers during my first years at Princeton. These were great experiences for learning more about your respective areas of study and helped expand the way I view my own work. Keisha Craig, RoseMarie Stevenson, Tina McCoy, Beth Porter, Carol Agans, David Carter, Adrian Cupid, Samita Mezger and Jawaan Lowery, thank you all so much for your guidance and assistance throughout the process. It has been wonderful getting to know and work with you all!

Stacey Sinclair, thank you for your diplomacy, patience, acceptance, and gentle nudging! I can’t imagine that taking an adopted student is easy. I am forever grateful for the space you’ve allowed me in the lab to be my quirky self while exploring my interests. Thank you, for teaching me how to convey my ideas more clearly and for exposing me to research that would further spark my curiosity.

Andrew Conway, thank you for taking me under your wing and teaching me how to think more critically and analyze more carefully. Thank you for your candor, diligence, and humor! Thank you for encouraging me and pushing me to take on new challenges—even when I didn’t have the confidence. Being part of Conway lab has shown me that we are a family and the

connections and support provided has been amazing. To Jim Spera, Mike Kane, and everyone in the extended lab family, thank you all so much!

Nicole Shelton, thank you for your continued advice and support over the years! I still remember when you introduced me to Claude Steele during my first year at Princeton!

To Travis Riddle, Lauren Feldman, Jordan Starck, Nate Cheek, Lindsey Eikenburg, Jen Yeaton, Sachiko Donley and all of SSP lab past and present, thank you for the constructive feedback and all the laughs. Stacey once said: “in SSP lab, we are each other’s people” and I am so grateful for the space of support and acceptance that you all provide.

Justin Junge, thank you for countless hours of helpful advice. You are so generous and kind! In the process of showing me how awesome a teacher and mentor you are, you taught me how to believe in myself more. It has been a great honor to learn from you over the years.

To Adam Brockett, Michael Chow, Rachel Connor, Cydney Dupree, Jeremy Borjon, Gary Kane, Kyonne Isaac, Lauren Feldman, DongWon Oh, Aaron Kurosu, Kaite Yang, and Brooke Macnamara—you all are truly remarkable friends and scholars! I have learned so much from each of you. Whether it was important lessons about science or about life. Thank you all for being there on the hardest days of the journey and making them brighter. Thank you for dancing with me at the DBar and Campus Club, for all the dinners, parties at Murray Place, and celebrating birthdays with me. These are wonderful memories that I will cherish forever.

Kerri Goodwin, thank you for your kindness and grace. You are such a wonderfully supportive and sharp advisor! You taught me how to develop my ideas as a budding scholar. I appreciate all that you have done and continue to do in mentoring me. A great deal of my dissertation wouldn’t be the same without you. Thank you for your unyielding support through tough spots and bumps in the road.

Peter Delaney, thank you for support and steady guidance even beyond my years as your honors thesis student. From our first meeting after switching my major to psychology, you've reminded me to trust the process and that as long as I continued working hard things would turn out ok. You are the mentor who entered my life at one of those "critical times" in my academic path. Mentors like you demonstrate how one person *can* shift the trajectory of someone's life. Thank you for being the first in my collegiate career to encourage me and assure me that I belong.

To Rich Daker, Max Drascher, David Cruikshank, Winston Lie, Bear Goldstein, Trent Schneider, John Broyles, Gaby Escalante, Wendy Zhao, Evan Klein, Daniel Jackson, Colin Neff, Zach Engner and the rest of my wonderful team of research assistants—who I lovingly refer to as my "research army"—thank you for data collection, scoring, and entry. It has been great to get to know, teach, and mentor you all. I gained valuable experience through running my own form of a mini-research team. This was a huge undertaking at Princeton and Towson and this project would not be possible without all of you!

Finally, thank you to my awesome family. Mom and Dad, you are my foundation. Thank you for all the love and support and for instilling me with faith—I can do all things through Him! Mom, thank you for teaching me to be strong and for always encouraging me to follow my dreams. Daddy, thank you for showing me how to live more in the moment, and how to retain a sense of wonder. To my brother, Darrion, thank you for keeping me young and hip—you are the kindest and funniest person I know! Connie, thank you for always being there as my best friend and sister for life. Joseph, thank you for being such a loving and encouraging partner (and trusted copy editor!). You're the best!

Table of Contents

Abstract	iii
Acknowledgments	iv
Table of Contents	vii
Chapter 1	
Introduction	1
Chapter 2	
A Moderated Mediation account of Working Memory Capacity during Stereotype Threat.....	15
Chapter 3	
Exploring Mindset, Personality, and Working Memory Capacity during Contexts of Stereotype Threat.....	46
Chapter 4	
A Mindset Intervention to moderate Stereotype Threat Effects on Working Memory Capacity and Standardized Test Performance.....	94
Chapter 5	
General Discussion & Conclusions	113
References	123
Appendices	135
Supplemental Information	140

Chapter 1

Introduction

“After the Egyptian and Indian, the Greek and Roman, the Teuton and Mongolian, the Negro is a sort of seventh son, born with a veil, and gifted with second-sight in this American world, a world which yields him no true self-consciousness, but only lets him see himself through the revelation of the other world. It is a peculiar sensation, this double-consciousness, this sense of always looking at one’s self through the eyes of others, of measuring one’s soul by the tape of a world that looks on in amused contempt and pity. One ever feels his two-ness—an American, a Negro; two souls, two thoughts, two unreconciled strivings; two warring ideals in one dark body, whose dogged strength alone keeps it from being torn asunder.”

-W.E.B. Du Bois on “Double Consciousness” from *The Souls of Black Folk*, 1903.

The quote above comes from William Edward Burghard (W.E.B) Du Bois, a world-renowned sociologist, civil rights activist, and the first African-American to receive a doctoral degree from Harvard University. As Du Bois speaks of a “double consciousness” that many African-Americans of his time experienced, his ideas about two-ness, specifically, “two unreconciled strivings” alert us to what it may feel like for any person of minority status when required to perform and achieve in environments that *are* or *feel* stigmatizing. This notion of two-ness is especially relevant in terms of achievement because in our current society, the ability to achieve plays a big part in education and can impact future life outcomes (c.f., Herrnstein & Murray, 1994; Aronson, Fried, & Good, 2002, Paunesku et al., 2015).

Today, success in life is based on the ability to achieve one’s goals both personally and professionally. As this is the case, it is important to consider factors that might misalign the “playing field” where specific groups of people receive unfair advantages. Although the topic of privilege is inextricably linked to both achievement and success in future life outcomes (see McIntosh, 1988) the perspective taken here will focus on understanding how stigmatized group members manage to *overcome disadvantages and excel in spite of them*. Thus, the question to arise is: how do those who are stigmatized in their environment manage to achieve their goals and garner success?

As an area of great importance, academic achievement can be thought of as the degree to which someone can *demonstrate* their intellectual ability both within and across different educational contexts. In addition, the extent to which someone can demonstrate her intellectual ability could be *influenced by* how she thinks during a specific context. In turn, one can conceive Du Bois' idea of "able-ness" as more or less achievable based on many environmental factors (e.g., resources, access to information, financial and moral support, or even, the presence of unmitigated stress). How then, do people go about resolving the conflict between what they are truly capable of and environmental influences? In the process of resolving this tension, do some arrive at success based on believing in themselves in spite of obstacles and disadvantages? If this were the case then after the removal of environmental hindrances the main thing still standing in the way of success would be mental barriers. Conceivably, mental barriers could be any issues within an individual that keep her from believing that she can achieve something. This may include her thoughts and beliefs about herself and her ability.

Du Bois also alludes to the issue of measuring oneself according to the view of others. While he is elaborating on what he feels it is like to live as a Black person in America, his description enlightens racial, social, cultural, and class issues of yesterday and today. In fact, previous research has explored the notion of double consciousness as a cognitive burden that those with stigmatized social identities face. In an investigation of the role of thought suppression during women's math performance, Logel, Iserman, Davies, Quinn and Spencer (2009) demonstrate that women's awareness of the stereotype about their group performing poorly and subsequent thoughts on this topic zapped useful mental space that could be directed toward performing. Moreover, results revealed that thought-substitution as opposed to suppression may aid performance during these situations. These findings suggest the power of

positive thoughts and attitudes during identity threatening situations. In addition, individual differences in cognitive resources and strategies may moderate the effectiveness of thought-substitution as a potential intervention. Taken together, this highlights the importance of considering the environment/situation, the individual, and the measurement tool used to assess performance. Turning more toward Du Bois' idea of measurement (i.e., "measuring one's soul by the tape of a world that looks on with amused contempt and pity.") implies that for a person of minority status in America, her thoughts about her abilities and self-worth may not be accurate, or even realistic looking through the lens of out-group and majority group members. This perspective highlights the importance of considering social and environment influences on how one measures and evaluates herself and her potential performance. For example, if taxed with the weight of "double consciousness" each time one performs, this additional perspective might overemphasize feeling evaluated. In turn, this can interfere with one's performance during key moments and impact future life outcomes.

Du Bois's stance is grounded in the legacy of racism but applies to sexism, classism, elitism, and beyond. His arguments foreshadow later findings of achievement gaps, stereotype threat effects, and achievement motivation—particularly informative areas when considering how we evaluate ability, intelligence, and general cognitive performance. If how others perceive us can impact how we view ourselves and potentially how we perform, then by extension it becomes necessary to consider *how cognitive abilities are assessed*, in addition to *factors that impact or bias* how these abilities are assessed. In order to explore this first, informative literature on the process of measuring ability will be reviewed. Then, how environmental and situational factors come into play when assessing cognitive ability and performance will be considered; looking specifically at stereotype threat effects. Finally, the scope of the paper will

expand to incorporate both cognitive and non-cognitive factors that may influence and/or drive intra- and inter-individual differences in cognitive performance under conditions of stereotype threat.

Assessment of Cognitive Ability

Measuring the exact nature of ability has proven a challenge. Since the inception of Spearman's *g* factor (1904), individual differences researchers have employed latent variable models to more closely derive the nature of true ability. In his seminal paper, Charles Spearman claimed to have "objectively determined and measured general intelligence." By administering several task batteries, the intuition was that one has more accurately measured certain aspects of some latent ability. Then, aggregating the measurements of several task batteries as factor loadings onto latent constructs, domain specific assessments can be combined to produce better measurements of domain general abilities.

Because intelligence is a complicated construct that consists of several subtests, there are a lot of important sub-measures involved which can predict domain-general ability from different domain-specific abilities. For example, previous work resolves that working memory capacity (WMC) is an especially meaningful sub-measure of intelligence because it involves an executive-control mechanism—which promotes increased vigilance against task interference by employing portions of the prefrontal cortex (Conway, Kane & Engle, 2003). This so-called executive control mechanism is thought to drive the relationship between WMC and Spearman's *g*, or "general intelligence." In addition, contrary to prior notions, researchers examined the relationship between WMC and *g*, ultimately concluding, though highly related, WMC and *g* are not identical, yet, both remain important for understanding the nature of ability.

Moreover, when reviewing the literature on intellectual ability and intelligence measures, Spearman's finding of the *positive manifold*—performance on one task being highly positively correlated with performance on other tasks remains one of the most highly replicated findings in psychology and psychometric literatures (Spearman, 1904; Conway & Kovacs, 2015). This is important to note because performance on intelligence measures have been shown to predict many real world outcomes from everything to better academic performance in school to getting divorced, to the number of offspring and even the risk of early death (see Ritchie, 2015). Although the over emphasis of intelligence in terms of a “be all, end all” measure is neither accurate nor particularly useful, more recent theories of intelligence characterize it as more of a tool— or specifically, a factor that emerges from the interaction of various domain general and domain specific factors (Kovacs & Conway, 2016). As such, depending on differences in ability on these domain specific and domain general factors the intellectual ability that emerges will vary from one person to the next.

As outlined above, when one is aiming to measure and ultimately make inferences about the nature of ability it is important to consider inter and intra individual differences. This will allow for more informed assessments both within and across individuals (Conway & Kovacs, 2013) and contexts. The history and research underlying assessments of ability is especially important because it tells the story of what scores on these measures actually mean. But how does this hold up in terms of the influence of social and environmental factors and the assessment of ability across different groups? As Du Bois argues, those with socially stigmatized identities could have additional environmental or situational factors (i.e., “double consciousness”) influencing their performance on these measures. If this is the case it is

important to consider the extent to which social and/or environmental factors play a role in influencing group differences in performance on tasks of intellectual ability.

Measuring Ability across Different Groups

In terms of explaining differences in ability across groups, the most prevalent argument on the topic comes from *The Bell Curve*, where Herrnstein and Murray (1994) suggest that 70% of differences in intelligence and ability can be explained by genetic inheritance, whereas, the remaining 30% is explained by environmental factors. Moreover, Herrnstein and Murray (1994) suggested that based on these findings it is reasonable to conclude that differences in intelligence and cognitive abilities between racial/ethnic groups might be driven by genetic inheritance. From this argument, Herrnstein and Murray (1994) de-emphasized the consideration for potential biases that may have interfered with how tests of intelligence actually “measure” these constructs.

For example, Wicherts and Millsap (2009) indicate that measurement bias “exists when two individuals who are identical on the construct(s) measured by a test, are from different groups [and] have different probabilities of attaining the same score on the test (p. 281).” In other words, if people from different groups who are matched in terms of their ability based on some construct of interest but the measure of the construct reveals they have different probabilities for getting the same score this violates measurement invariance and suggests the test is biased in measurement of the construct of interest (see Wicherts, Dolan & Hessen, 2005; Wicherts & Millsap, 2009; Wicherts & Dolan, 2010). In this case one group has an advantage over the other in terms of the scores they can obtain. This highlights issues of fairness in assessment of ability and how to ensure that some groups are not being unfairly advantaged on these tests over others. It should not be the case that any measure that is systematically favoring one group over the

other should be used for entrance to academic institutions or for job hiring practices. Instead, we either have to create a new form of assessment that is more fair or we have to change the way we think about how such tests are used and what they tell us. The good news is that companies like the Educational Testing Service, who administer and score many standardized tests of academic assessment and achievement, implement fairness practices from the item level (individual questions on the test) through the test as a whole. In order to ensure fairness they are “committed to producing tests and other products that acknowledge the multicultural nature of society and treat its diverse populations with respect...ensuring that test takers and others who make up our increasingly diverse customer base enjoy equal access to our products (Educational Testing Service, 2018, *Fairness Guidelines*).”

Intelligence: a psychometric factor that can be “socially situated”

In terms of arguments posed regarding whether measures of intelligence and cognitive ability are measurement invariant, there are two main schools of thought: those who subscribe to more fixed and innate views of intelligence (Jensen, 1969; Herrnstein & Murray, 1994), and alternatively, those who challenge fixed views of intelligence and ability based on evidence for measurement bias and/or the influence of social/environmental factors (Binet, 1909/1973; Gould, 1981; Flynn, 1984; Steele & Aronson, 1995; Wicherts, Dolan, & Hessen, 2005; Jaeggi et al., 2008). Although this is a contentious topic, there is a growing body of evidence asserting that focusing on fixed views of intelligence and ability is not only potentially problematic (Dweck, 2006) but may be inaccurate (Wicherts, Dolan, & Hessen, 2005). However, different perspectives on intelligence do not necessarily have to be mutually exclusive. Previous work demonstrates that intelligence measures are valid and predict many real world behaviors, but it could also be useful to view performance on measures of intellectual ability as “socially situated”

(see Walton, 2013, also see Rogoff & Lave, 1984). Based on a socially situated view, intelligence and performance on tasks of intellectual ability can be thought of as emerging from the interaction of the person with their social context/environment. In turn, with this view performance is characterized by the extent to which an individual's potential is either disrupted or facilitated based on aspects of the social situation/environment. As outlined above, measuring the exact nature of ability is complicated, particularly when considering differences in social/group identity in an ever-changing, diverse and multicultural society. For this reason it is important to consider the costs and benefits for different individuals when viewing intelligence and intellectual ability as socially situated.

Stereotyping and Measurement Bias: The costs and benefits of stigmatization

The problem with being assessed is that one might feel the pressure of being evaluated by others, producing anxiety. Previous research exploring the mere presence of others confirms that both physical proximity and the awareness of the physical proximity of others is enough to induce anxiety and shift behavior in both positive and negative direction depending upon the task at hand (see Triplett, 1898; Zajonc, Heingartner & Herman, 1969). Based on different aspects of group identity, this evaluative stress can be even more anxiety provoking.

For many people, the stigma associated with belonging to certain group identities could influence additional variation and induce a form of measurement bias when being assessed. This may come from worries of being stereotyped. Stereotyping can be thought of as the cognitive process of categorizing and overgeneralizing about people based on a few topics that are loosely associated with some aspect of their group identity (Lippmann, 1922). Stereotyping is a phenomenon that has been researched widely and has been observed to be both necessary for human cognitive functioning yet, as a byproduct of the cognitive system can also be maladaptive

(Macrae & Bodenhausen, 2000). Overall, stereotyping provides a mental shortcut for the cognitive system, impacting the processes by which we think, feel, and ultimately, behave. In this sense, stereotyping does provide us with a heuristic that allows us to conserve cognitive resources each time we encounter new information, however, as one might imagine, this can be damaging and detrimental to those who are the targets of stereotyping (i.e., performance decrements in the form of stereotype threat effects).

Stereotype Threat: Definition & Mechanisms

In 1995, Steele and Aronson published their seminal work discussing the effect of stereotype threat on performance. By inducing threat—making some negative aspect of group membership salient—those under threat suffered significantly decreased task performance relative to control groups. There is a tendency to explain the factors driving the threat effect as affective and physiologically derived (Steele & Aronson, 1995; Steele, 1997). The threat effect has been explained as negative affect—specifically, anxiety and worries associated with poor task performance. Cognitive accounts for stereotype threat effects focus on the allocation of cognitive resources. This line of research explains threat through cognitive monitoring mechanisms. The impact of threat is demonstrated when cognitive resources (via WMC) are wasted on monitoring performance rather than actually performing (Schmader & Johns, 2003). In addition, prior research indicates that WMC moderates the relationship of threat effects on task performance (Regner et al., 2010); proposing that higher WMC participants are better equipped to manage the cognitive demands of threat, whereas, low WMC participants show classic threat effects in the form of performance decrements relative to controls.

Another cognitive interpretation is that negative affect associated with the stereotype threat is internalized in that a person is aware of the stereotype and preoccupied from the task

and this “spills over” onto subsequent task performance (Inzlicht & Schmeichel, 2012). Moreover, the logic behind spillover is that prior to the threat induction cognitive resources (potentially via WMC) can be allocated most efficiently toward task performance, however, once threat is induced a portion of these resources is wasted on thoughts, negative affect, and any worries associated with task performance. This in turn puts the individual in a state of resource depletion based on the inability to exert self-control as efficiently.

Taken together, these are useful perspectives for unpacking what may happen cognitively when students are subjected to stereotype threat. However, it is important to acknowledge that previous research finds replication issues within the stereotype threat literature. This is important to mention because if there is a failure to replicate the basic effect of stereotype threat on performance then the cognitive mechanistic accounts for understanding stereotype threat become less informative.

Replication of Stereotype Threat Effects

In a real world replication by researchers at the Educational Testing Service—Stricker and Ward (1998) induced stereotype threat by priming college students with their ethnic/racial identity and sexual/gender identity before taking important computerized course placement exams (or after the exams in the control) across areas of mathematics and reading comprehension. Results revealed no differences in performance across groups based on the racial or sexual group identity primes. However, the students in this study were college students who were already admitted thus, there could be a qualitative difference between the situational pressure to perform on placement tests compared to those students feel when their performance on tests can determine their admittance to a college/university.

In another study, Stricker (1998) investigated the effect of the same type of prime for race and gender/sexual identity on performance on the Calculus Advanced Placement Test. Again, results revealed no significant differences in performance for Black students or women based on being primed for their racial or sexual/gender identity before taking the test (also see Stricker & Ward, 2004).

More recently, in a meta-analysis, Flore and Wicherts (2014) found evidence that the average stereotype threat effect on young girls' math performance was small and negative (which matches the prediction of the effect on math performance) there was also evidence of an overrepresentation of statistically significant results among these samples. Flore and Wicherts also suggest that their findings indicated the presence of publication bias such that more statistically significant effects were present in these samples "as would be expected based on the cumulative power of all the samples." Importantly, the take-away of these findings is that the effect of stereotype threat could be inflated based on the impact of publication bias and the authors suggest that the effect may be much smaller and non-significant in more highly powered studies, however, implicit threat manipulations were found to be more effective than more explicit threat manipulations.

In additional work, Sackett, Hardison, and Cullen (2004) assert the importance of understanding Steele and Aronson's (1995) original findings and not mischaracterizing them—many characterize the threat effect in such a way that when it is removed that the achievement gap difference between White and Black students should also be removed. This is inaccurate. Steele and Aronson originally demonstrated that standardized test performance could be impacted by environmental factors thus, such factors are important to consider when interpreting and using scores on such measures to make decisions related to individual's access or admittance

to institutions of higher education or for various hiring procedures. Based on Steele and Aronson's finding that a threat manipulation could have the power to shift performance on tests which are developed psychometrically to be un-amenable to subtle situational influences, Sackett and colleagues (2004) suggest that research in this area should continue to be conducted but that researchers and laypeople should take caution in their interpretation that removing the stereotype threat will act as a cure-all for gaps in achievement between White and Black students.

Although stereotype threat researchers have settled on both sides of the argument in terms of the magnitude and robustness of the effect, (see Schimmack, 2016; Radiolab Interview, 2017; Inzlicht, 2016) these debates have brought the many complicated issues of educational inequity, stigma, social inequality, as well as the assessment of intellectual ability and academic achievement to the forefront. All in all, stereotype threat research remains unresolved and the best way to address issues of replicability is to continue conducting research in this area while reporting the findings as openly and as honestly as possible (see Open Science Collaboration, 2015; also see Simmons, Nelson, & Simonsohn, 2011). Doing so is important work for the field's understanding of the complexity of the psychological process of and the impact of stereotype threat on the students of yesterday and today.

Goals of the present work

In a review of social psychological interventions, Yeager and Walton (2011, pp. 268) state "Social-psychological interventions hold significant promise for promoting broad and lasting change in education, but they are not silver bullets. They are powerful tools rooted in theory, but they are context dependent and reliant on the nature of the educational environment." Yeager and Walton (2011) provide a necessarily informative way to begin thinking about interventions to reduce negative performance outcomes related to effects of stereotyping. In

essence, there is no “one size fits all;” the contextual and environmental factors should be considered in addition to differences in ability and motivation that individuals bring to these situations. With the revival of interest in non-cognitive factors and their application in education research and policy, properly measuring and exploring these skills (Duckworth & Yeager, 2015; also see Nettles, 2017; Durlak et al., 2011) alongside classic cognitive ability measures will help expand to an individual differences model of stereotype threat effects and performance. Hence, arriving at the aims of the present investigation.

A goal of the current work is to better understand how people are able to achieve and overcome in spite of situational and/or environmental obstacles. Previous research has suggested that differences in cognitive functions and differences in personality/attitudes might explain differences in the levels of protection students experience when combatting stigmatizing or threatening experiences during assessments of cognitive or intellectual ability. Moreover, as detailed above, making inferences about an entire group’s challenges and performance can be both inaccurate and complicated; consequently, any solution must take this into consideration. As the literature on the assessment and measurement of cognitive ability informed, psychometric perspectives are important to consider in the context of stereotype threat. Psychometric models provide us with more information than simple aggregated measures and conclusions. Because the richness and nuance of the data are lost based on drawing inferences from aggregated measures so the same could be argued about over relying on point estimates when examining the impact effects like stereotype threat have on performance. The advantage of investigating stereotype threat from an individual differences perspective lies in the power of considering skills and abilities beyond just the group level. This will provide novel and more nuanced strategies and solutions that are yet to have been considered or revealed.

Here, the goal is uncover the roles that differences in cognitive functioning along with differences in personality/attitudes play when considered together during contexts of stereotype threat. Across several studies this work investigates whether individual differences in cognitive functioning mediate and/or moderate stereotype threat effects on cognitive outcomes such as state WMC and standardized test performance. In addition, these studies also investigate whether differences in attitudinal measures such as mindset, grit and personality may also moderate the effects of stereotype threat on performance. It also remains critical however, to uncover more about what drives stereotype threat effects, mechanistically. Through gaining a better understanding of mechanism at the inter- and intra-individual difference levels this work will aid future intervention research by contributing better-informed solutions for mitigating stereotype threat effects.

Chapter 2

A Moderated Mediation account of Working Memory Capacity during Stereotype Threat

The present experiment considers a moderated mediation model of stereotype threat effects on cognitive performance. Specifically, investigating the roles of both inter- and intra-individual differences in WMC and how and when they may protect students from stereotype threat during testing situations. The experiment aimed to test several hypotheses. First, starting with conceptual replications of previous findings, it is expected that when stereotype threat is induced there will be a performance decrement on all outcome measures relative to the control condition. Here, the threat effect is defined in terms of a performance decrement. Thus, a null effect would indicate that the difference between threat and control group means would approach 0. Likewise, based on previous findings, it was expected that WMC is an important mechanism involved and it will both mediate and moderate the threat effect on performance. Thus, for individuals having higher trait WMC, this should be associated with less of an impact of threat relative to lower trait WMC individuals. In addition, state WMC should explain the performance differences observed between the control and threat conditions. That is, performance differences were expected to be larger when a threat is present and internalized, whereas, the difference was expected to be smaller and approach 0 when threat is mitigated/controlled.

Method

Participants

A total of 469 undergraduate students from Princeton University were recruited to participate in the study. Eleven participants were removed with list-wise deletion due to variables missing completely at random—these were cases such as computer errors or missing portions of the standardized test worksheets. One subject had portions of paper worksheet

missing but this did not interfere with the measure necessary to be included in the final analysis. The remaining total was 458 participants (270 female, 188 male). Subjects were invited to participate if they were at least 18 years of age, self-identified as White (366 students, 202 female) or Black (87 students, 66 female). Five participants identified as multiracial or multiethnic with some combination of different racial identities other than Black and White (2 female, 3 male). Those who identified as bi-racial and as Black and White were categorized as Black because bi-racials who are Black and White reported to identify with Black as their “primary” ethnic identity). All other bi-racial, multiracial/multiethnic participants were included in a separate category from those in the Black and White groups. Due to the low statistical power of this third category these subjects were not included in the main statistical analyses reported below. All subjects were required to be native English speakers. In exchange for their participation, students received either course credit or \$16 cash.

Design

A 2 (Condition: Threat vs. Control) x 2 (WMC Span Task Order: Verbal Capacity & Intelligence vs. Math Capacity & Intelligence) x 2 (Race: White vs. Black) factorial design was administered where participants were randomly assigned to either stereotype threat or control conditions and either Verbal Capacity & Intelligence or Math Capacity & Intelligence task order. It is important to note that half of the participants performed a Verbal task at baseline and were then threatened for math capacity and intelligence while the other half performed a Math task first and were then threatened for verbal capacity and intelligence. Also, participants in the control condition did not receive any directly threatening instructions and were not given any domain specific information about the tasks, instead they were only told they would complete a working memory task (see Figure 2.1 and the Procedure section below).

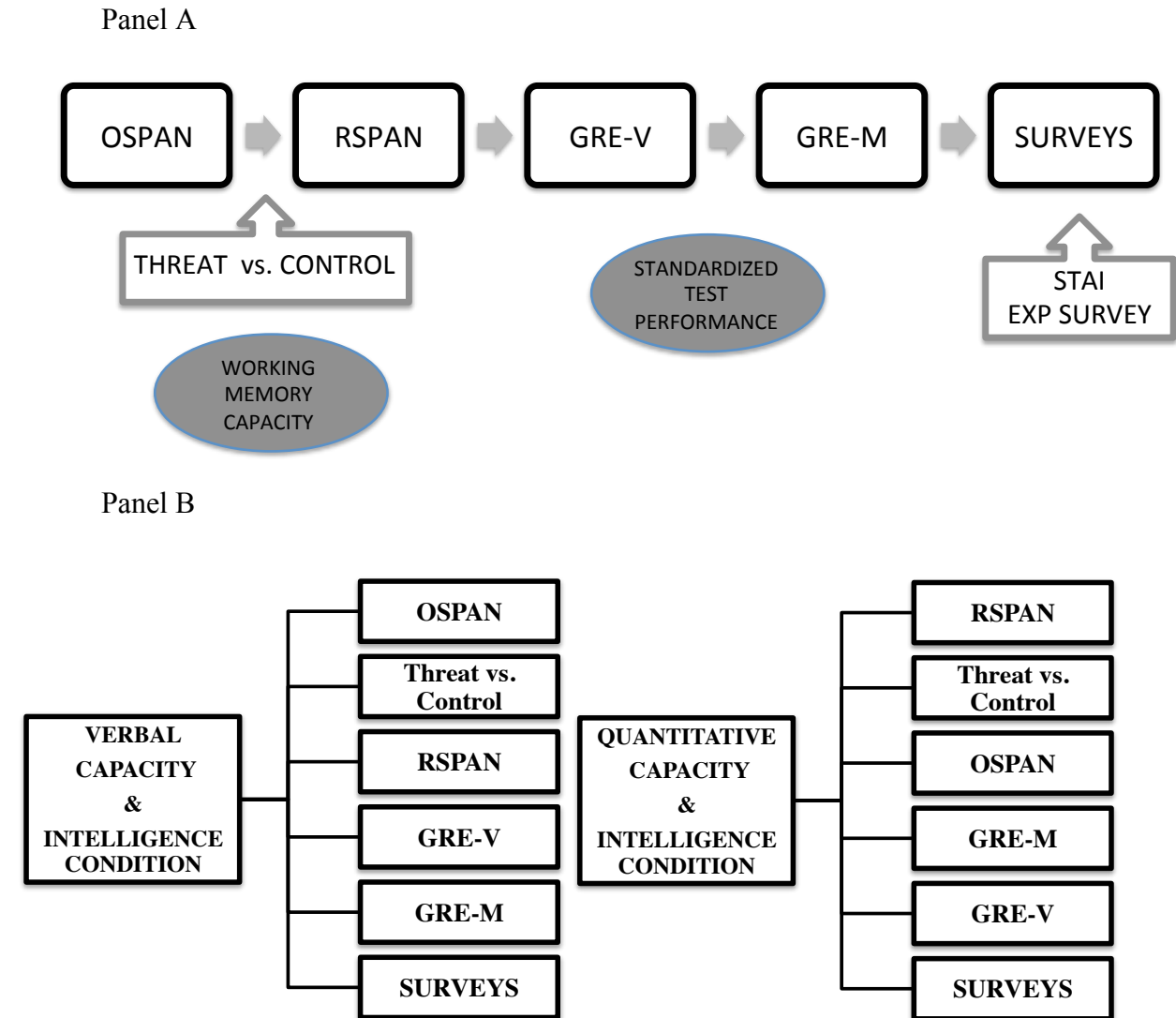


Figure 2.1. Experimental Procedure.

Panel A demonstrates the order of the tasks including the variables characterized as WMC and standardized test performance. Panel B demonstrates the procedures for the verbal capacity & intelligence and quantitative capacity & intelligence conditions.

Procedure

First, participants completed informed consent. Next, participants completed a complex WM span task. Participants in the math capacity and intelligence condition completed the Reading Span task (RSPAN), while participants in the verbal capacity and intelligence condition

completed the Operation Span task (OSPAN). After completing one complex WM span task participants either received threat or control instructions followed by a second WM span task. Participants who performed RSPAN before the threat manipulation completed OSPAN after the threat manipulation and vice-versa. In the threat condition, participants received instructions similar to those of Schmader & Johns (2003) and filled out an “ethnicity survey” which served to prime race and induce stereotype threat. In the control condition, participants received similar instructions, which were modified to exclude the race prime.

Following the second WM span task, all participants completed two sections of the Graduate Record Examination (GRE): one verbal and one quantitative section. After the GRE sections, participants completed the Spielberger State-Trait Anxiety Inventory (STAI) to assess state anxiety (Marteau & Becker, 1992). Then, participants completed a survey containing the ethnicity question (for those in the control condition) in addition to assessing any general thoughts about the study. After completing all tasks and surveys all participants received a debriefing form describing the purpose of the study as investigating whether WMC moderates the effect of stereotype threat on test performance.

Stereotype Threat Manipulation. The threat manipulation procedure was based on Schmader and Johns (2003), with slight alterations—we included two measures of WMC, one before the threat manipulation and one after. In order to make the present study amenable to threat effects for both operation (OSPAN: Turner & Engle, 1989; Unsworth, Heitz, Schrock & Engle, 2005) and reading (RSPAN: Daneman & Carpenter, 1980) spans, we induced a race prime before having subjects complete the working memory tasks. For OSPAN and RSPAN subjects were informed that the task was indicative of quantitative or verbal capacity, respectively and that the task was highly related to measures of intelligence. Moreover, they

were informed that some aspect of performance could be attributed to group membership. Following this information, they were instructed to complete the ethnicity survey, which consisted of a single question asking participants to indicate their race.

Order Manipulation. Following Schmader and Johns (2003), an assessment of WMC was included in order to replicate the effect of threat on WMC and to test the hypothesis that WMC mediates the effect of threat on standardized tests. To detect these effects, WMC must be measured after the threat manipulation. However, we also wanted to test the hypothesis that WMC moderates the effect of threat on standardized testing. To test this hypothesis WMC must be measured before the threat manipulation. Thus, half the participants performed RSPAN (reading) before the threat manipulation and OSPAN (math) after the threat manipulation and vice-versa. The math threat and intelligence group is analogous to Schmader and Johns (2003), where OSPAN was administered after the threat manipulation (see Figure 2.1).

Tasks

Operation Span. The automated OSPAN task requires completing a series of arithmetic problems while remembering a list of letters. For example, an individual trial in a list would take the following form: $[5 \times 9] + 5 = 45$? Yes or No? P. Participants were instructed to solve the equation, answer yes or no, and then remember the letter for later recall. At the end of a list of such trials, participants were asked to recall the letters in serial order (total score was calculated using the partial unit method, see Conway et al., 2005). Participants received three to seven letters per trial and three sets of each trial length, totaling 15 trials total—yielding a maximum score of 75.

Reading Span. The automated RSPAN requires making veridical judgments for sentences while remembering a list of letters (e.g., making a veridical judgment for the sentence

“when at last his eyes opened, there was no gleam of triumph, no shade of anger” Yes or No? R). Participants were instructed to read the sentence, answer yes or no, and then remember the letter for later recall. At the end of a list of such trials, participants were asked to recall the letters in serial order (total scored using the partial unit method, see Conway et al., 2005). Participants received three to seven letters per trial and three sets of each trial length, totaling 15 trials total—yielding a maximum score of 75.

GRE-Math. The GRE mathematics subsection consisted of 25 multiple-choice or short answer questions, each requiring mathematical reasoning and quantitative comparison skills (e.g., Solve the equation $5 = 4x + 3$ for x . Is x greater than, less than or equal to 1?) Participants were given 20 minutes to complete the subsection. The section was taken from free online practice materials provided by the Educational Testing Service. The final score was the proportion of questions correct, i.e., total correct out of 25 possible questions.

GRE-Verbal. The GRE verbal subsection consisted of 25 multiple-choice questions. Verbal questions required the abilities to analyze, evaluate and synthesize written material, in addition to recognizing relationships among words and concepts (e.g., Read the passage and answer the following: “The passage addresses which of the following issues related to the author’s use of popular elements in his classical compositions?”). Participants were given 20 minutes to complete the subsection. The section was taken from free online practice materials provided by the Educational Testing Service. Calculating the proportion correct—we took the total correct out of 25 possible questions scored the verbal section.

STAI. Participants completed the Spielberger State-Trait Anxiety Inventory in order to observe state anxiety. The inventory consisted of 6 questions where participants were asked to indicate the most appropriate answer to each statement based on their present feelings. The

inventory contained a 4-point Likert scale (e.g., “I feel calm?” on a scale of 1 = not at all and 4 = very much). Questions associated with low levels of anxiety (e.g., “I am relaxed.”) were reverse scored. For scoring purposes the Likert values for each question were summed—the scores ranged from 6 to 24.

Experiment Survey. All participants completed a final experiment survey. Its purpose was twofold: to mimic the ethnicity question for those in the control group and to gain a general assessment of participants’ experiences (e.g., “Do you know anyone else in this study?” or “Did you feel the study was too long?”).

Results

Data Preparation & Analytic Approach

Upon looking at the distributions of the data, 6 additional subjects were removed on the basis of being identified as univariate outliers falling outside the 3 SD range of the mean. Thus, the following analyses were conducted with a final sample of 447 participants who were White (360 total, 198 female) or Black (87 total, 66 female).

Because there are several hypotheses under investigation, these require different approaches to the data. In an attempt to simplify the presentation of results we present two “sets” of results, one for each task order condition. Because we manipulated stereotype threat for either “verbal” or “quantitative” capacity in the threat condition we separate the threat effects based on these task orders (refer to Figure 2.1). In addition, for each of the sets of analyses below, homogeneity of variance assumptions were tested with *Levene’s* test. Where the test was statistically significant, *Bonferonni* corrections were used and reported below. In cases where *Levene’s* tests were non-significant these results were not reported and normal t-tests and tests of analysis of variance are reported below.

It is also important to note that in each regression analysis below, levels of condition were dummy coded with control = 0, threat = 1. Because the threat effect was expected to manifest as a performance decrement for black subjects, only their data were analyzed for moderation effects.

Summary Statistics and Correlations.

Looking at the general trends in the data, descriptive statistics revealed the measures of WMC (i.e., ospan and rspan) were strongly positively correlated with each other as well as moderately positively correlated with measures of standardized test performance (i.e., math and verbal GREs). It was also revealed that higher scores on the anxiety measure were negatively associated with all of the performance measures (presented in Table 2.1 below).

Table 2.1. *Summary Statistics and Correlations of Variables in Experiment 1*

	1	2	3	4	5	6	7
1. ospan	--						
2. rspan	.590***	--					
3. gremProp	.260***	.270***	--				
4. grevProp	.220***	.270***	.465***	--			
5. gremAttProp	-0.007	.080	.601***	.186***	--		
6. grevAttProp	.021	.113*	.246***	.405***	.425***	--	
7. STAI	-0.056	-.103*	-.209***	-.092*	-.175***	-0.023	--
<i>M</i>	63.090	58.650	0.56	0.60	0.82	0.94	12.66
<i>SD</i>	10.32	11.75	0.19	0.20	0.15	0.10	3.61

Note. ospan = Operation Span, rspan = Reading Span, gremProp = Proportion Correct on Math GRE, grevProp = Proportion Correct on Verbal GRE, gremAttProp = Proportion Attempted on Math GRE, grevAttProp = Proportion Attempted on Math GRE, STAI = Speilberger State-Trait Anxiety Inventory. *** $p < .001$, ** $p < .01$, * $p < .05$.

Verbal Capacity & Intelligence Condition.

Baseline Effects.

The first analysis was an independent t -test to detect differences in baseline WMC. Thus, race was treated as the (quasi) independent variable and OSPAN as the dependent variable. Results indicated a significant difference in baseline OSPAN based on race, $t(246) = -2.5$, $p < .05$, Cohen's $d = .42$, with White subjects significantly outperforming Black subjects by almost a half standard deviation difference at baseline (see Figure 2.2 below).

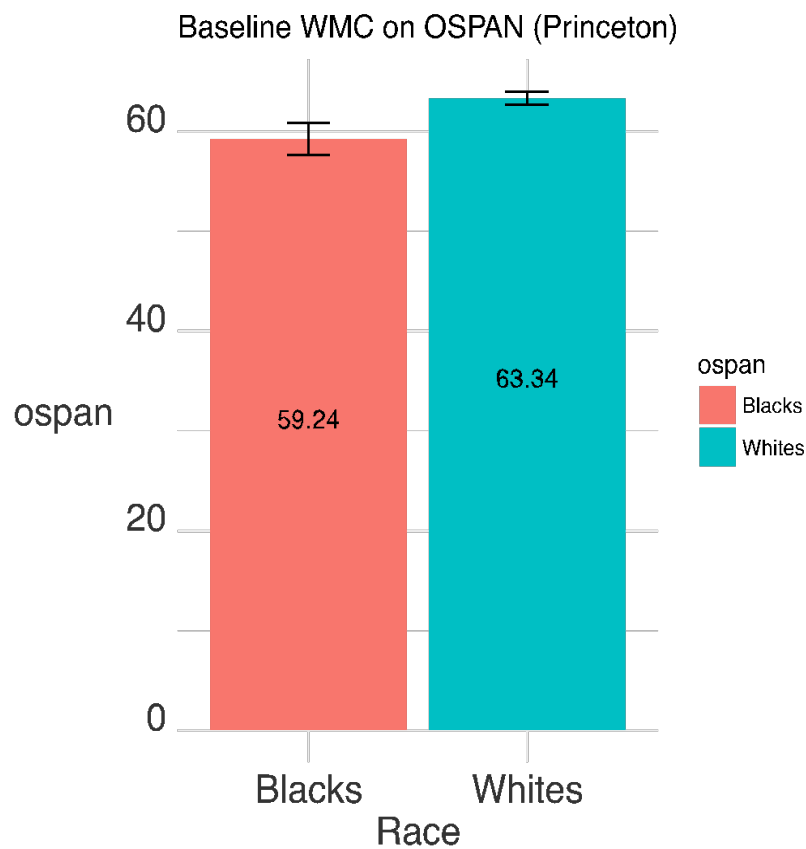


Figure 2.2. Trait/Baseline WMC on OSPAN by Race.

Threat Effects.

In the first 2 (race) x 2(condition) ANOVA, looking at the effect of stereotype threat for “verbal capacity,” RSPAN was the dependent measure. Results indicated a main effect for race, $F(1, 244) = 10.2, p < .01, \eta_{\text{partial}}^2 = .041$ such that white subjects outperformed ($M = 59.3, SD = 11.9$) Black subjects ($M = 52.8, SD = 12.1$) on the RSPAN. The effects of condition, $F(1, 244) = .17, p > .05, \eta_{\text{partial}}^2 = .00070$ and the interaction, $F(1, 244) = .11, p > .05, \eta_{\text{partial}}^2 = .00045$ were non-significant. These results are demonstrated in Figure 2.3 below.

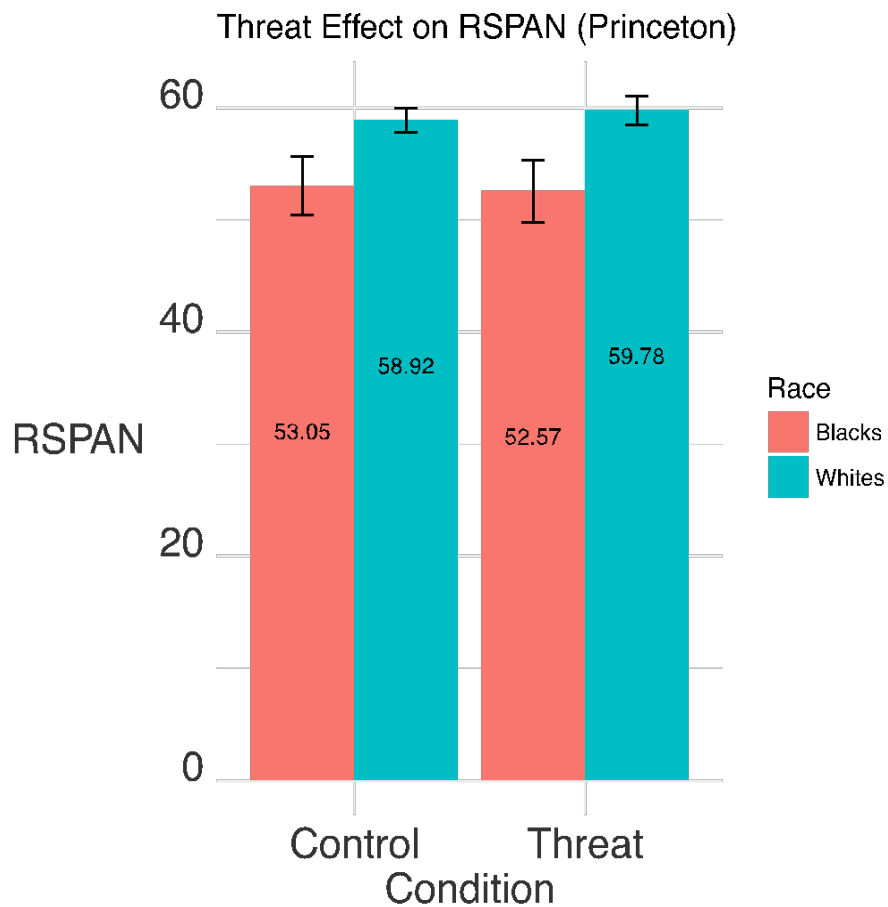


Figure 2.3. Effect of Stereotype Threat on State WMC.

In the second 2 (race) x 2 (condition) ANOVA, looking at the effect of stereotype threat for “verbal capacity,” GRE-V was the dependent measure, *Levene’s* test revealed a significant

homogeneity of variance violation, $F(3, 244) = 3.4, p < .05$. Thus, more conservative criteria for significance was used in the analysis of variance model below—based on making six pairwise comparisons a *Bonferroni* correction was used to compute the new significance criteria of .0083. Results indicated only a significant main effect of race, $F(1, 244) = 13.9, p < .001, \eta_{\text{partial}}^2 = .053$ such that White subjects in the threat ($M = .63, SD = .21$) significantly outperformed Black subjects ($M = .50, SD = .17$) on the verbal GRE. The effects of condition, $F(1, 244) = .10, p > .05, \eta_{\text{partial}}^2 = .00041$ and the interaction, $F(1, 244) = 1.2, p > .05, \eta_{\text{partial}}^2 = .00049$ were non-significant. These findings are demonstrated in Figure 2.4 below

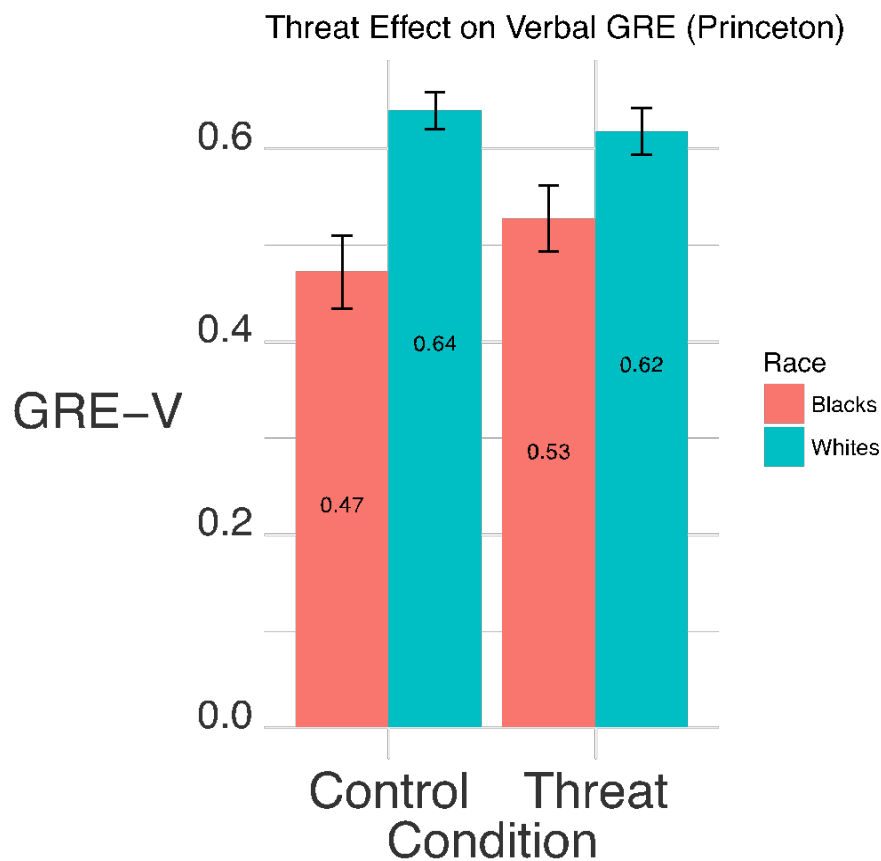


Figure 2.4. Effect of Stereotype Threat on GRE-V.

In the third 2 (race) x 2 (condition) ANOVA, looking at the effect of stereotype threat for “verbal capacity,” GRE-M was the dependent measure¹. Results indicated the main effect of race was significant, $F(1, 243) = 13.9, p < .001, \eta_{\text{partial}}^2 = .067$ such that White subjects ($M = .60, SD = .18$) significantly outperformed Black subjects ($M = .47, SD = .20$). The effects of condition, $F(1, 243) = 1.2, p > .05, \eta_{\text{partial}}^2 = .0046$ and the interaction, $F(1, 243) = .053, p > .05, \eta_{\text{partial}}^2 = .00022$ were non-significant. These findings are demonstrated in Figure 2.5 below.

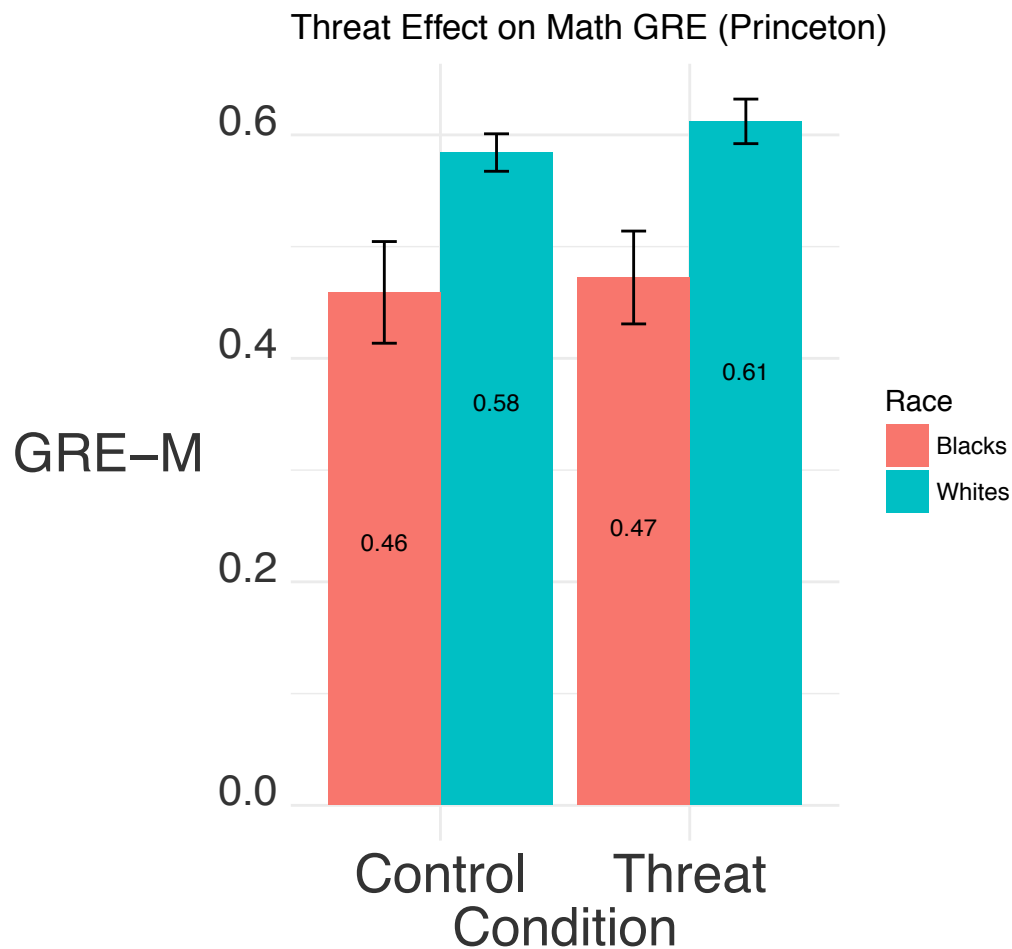


Figure 2.5. Effect of Stereotype Threat on GRE-M.

¹ One subject's data was missing on the math GRE in this condition.

Mediation.

A mediation analysis for Black subjects with condition as the independent variable, RSPAN, the mediator variable, and GRE-V as the outcome variable was conducted. The direct effect of threat on GRE-V was non-significant and close to 0, $t(40) = 1.08$, $b = .055$, $p > .05$ indicating that the threat condition had a predicted .055 increase on the GRE-V for black students. The effect of threat on RSPAN was also non-significant, $t(40) = -.126$, $b = -.48$, $p > .05$, indicating that the threat effect predicted a -.48 decrease on RSPAN. The indirect effect of Threat on GRE-V when RSPAN was added to the model was non-significant and did not appear to change much based on including RSPAN in the model, $t(39) = .057$, $p > .05$. In the indirect path, the effect of RSPAN on GRE-V was trending toward significance, $t(39) = 1.8$, $b = .0037$, $p = .08$. Moreover, *Sobel's* test supported the lack of evidence for statistical mediation of RSPAN in the effect of threat on GRE-V, $Z = -.12$, $p = .26$ (see Figure 2.6).

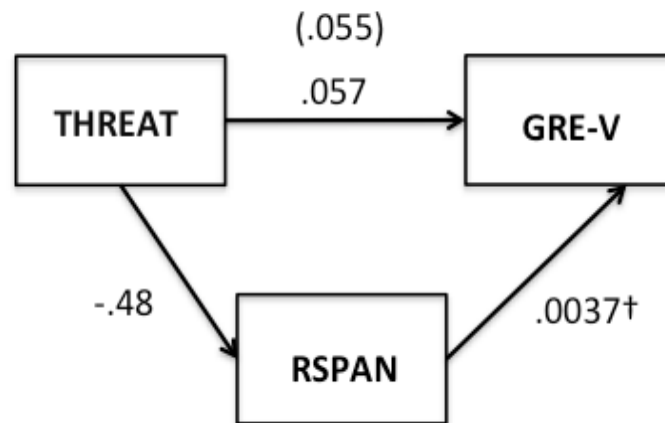


Figure 2.6. Model of RSPAN Mediating Threat Effects on GRE-V.

Using nonparametric bootstrapping methods to resample 1000 simulations, results indicated non-significant effects for the average causal mediation effect, parameter estimate = -

.0018, CI [-.038, .032], $p = .92$, as well as the average direct effect of threat on GRE-V, parameter estimate = .057, CI [-.043, .16], $p = .25$. Taken together, these findings reveal a lack of evidence that Black subjects' WMC (as assessed via the RSPAN) mediated the effect of threat on GRE-V (see Figure 2.7).

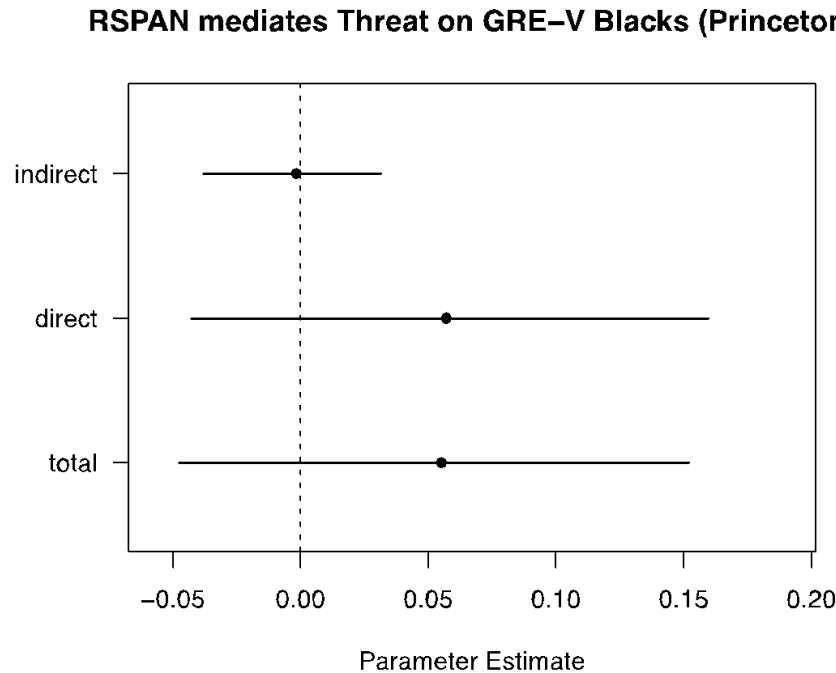


Figure 2.7. RSPAN Mediation Model for Threat on GRE-V using Bootstrapping.

A mediation analysis for Black subjects with condition as the independent variable, RSPAN, the mediator variable, and GRE-M as the outcome variable was conducted. The direct effect of threat on GRE-M was non-significant and close to 0, $t(40) = .22$, $b = .013$, $p > .05$ indicating that the threat condition had a predicted .013 increase on the GRE-M for black students. The effect of threat on RSPAN was also non-significant, $t(40) = -.13$, $b = -.48$, $p > .05$, indicating that the threat effect predicted a -.48 decrease on RSPAN. In the indirect path the effect of threat on GRE-M when RSPAN was added to the model was non-significant and did

not appear to change much based on including RSPAN in the model, $t(39) = .22$, $b = .014$, $p > .05$. In the indirect path, the effect of RSPAN on GRE-M was also not significant, $t(39) = .44$, $b = .0012$, $p > .05$. Moreover, *Sobel's* test supported the lack of evidence for statistical mediation of RSPAN in the effect of threat on GRE-V, $Z = -.12$, $p = .82$ (see Figure 2.8).

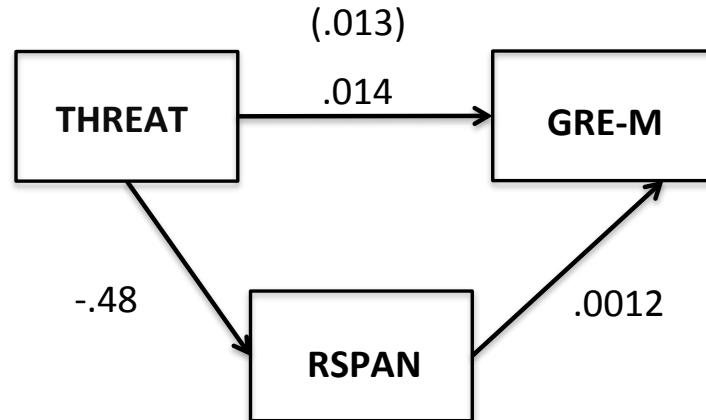


Figure 2.8. Model of RSPAN Mediating Threat Effect on GRE-M.

Using nonparametric bootstrapping methods to resample 1000 simulations, results indicated non-significant effects for the average causal mediation effect, parameter estimate = $-.00059$, CI $[-.020, .014]$, $p = .91$, as well as the average direct effect of threat on GRE-M, parameter estimate = $.014$, CI $[-.11, .13]$, $p = .87$. Taken together, these findings further support a lack of evidence that Black subjects' WMC (as assessed via the RSPAN) mediated the effect of threat on GRE-M. And moreover, resampling results also indicated that the threat manipulation had a non-significant effect on Black subjects' performance on the math GRE (see Figure 2.9).

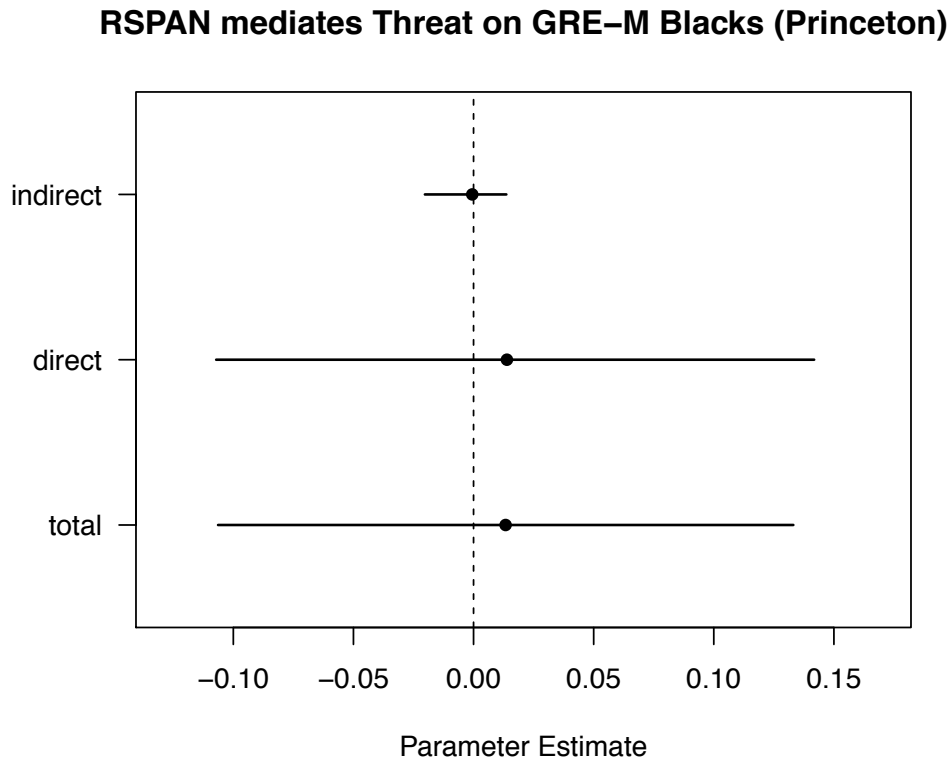


Figure 2.9. RSPAN Mediation Model for Threat on GRE-M using Bootstrapping.

Moderation.

To test whether trait WMC moderates the effect of stereotype threat on standardized testing, hierarchical regressions were conducted for Black subjects with the outcomes of verbal and math GREs. For both GREs, in the first step, the main effects of OSPAN (mean centered) and condition were investigated. In the second step, the product of OSPAN (mean centered) and condition (dummy coded with control = 0, threat = 1) was added in order to test for moderation.

For the verbal GRE, in step one, the effects of OSPAN, $b = .0032$, $t(39) = 1.31$, $p > .05$, condition, $b = .059$, $t(39) = 1.16$, $p > .05$, and the model were non-significant, $F(2, 39) = 1.45$, Multiple $R^2 = .069$, $p > .05$.

In step two, the effect of OSPAN was trending toward significance, $b = .0061$, $t(38) = 1.93$, $p = .06$ indicating that there was only a marginally significant effect of trait WMC such that higher OSPAN scores were associated with higher predicted scores on the verbal GRE. The effects of condition, $b = .031$, $t(38) = .582$, $p > .05$, the interaction, $b = -.0070$, $t(38) = -1.42$, $p > .05$, and the model, $F(3, 38) = 1.67$, Multiple $R^2 = .116$ were not significant (see. Figure 2.10 below).

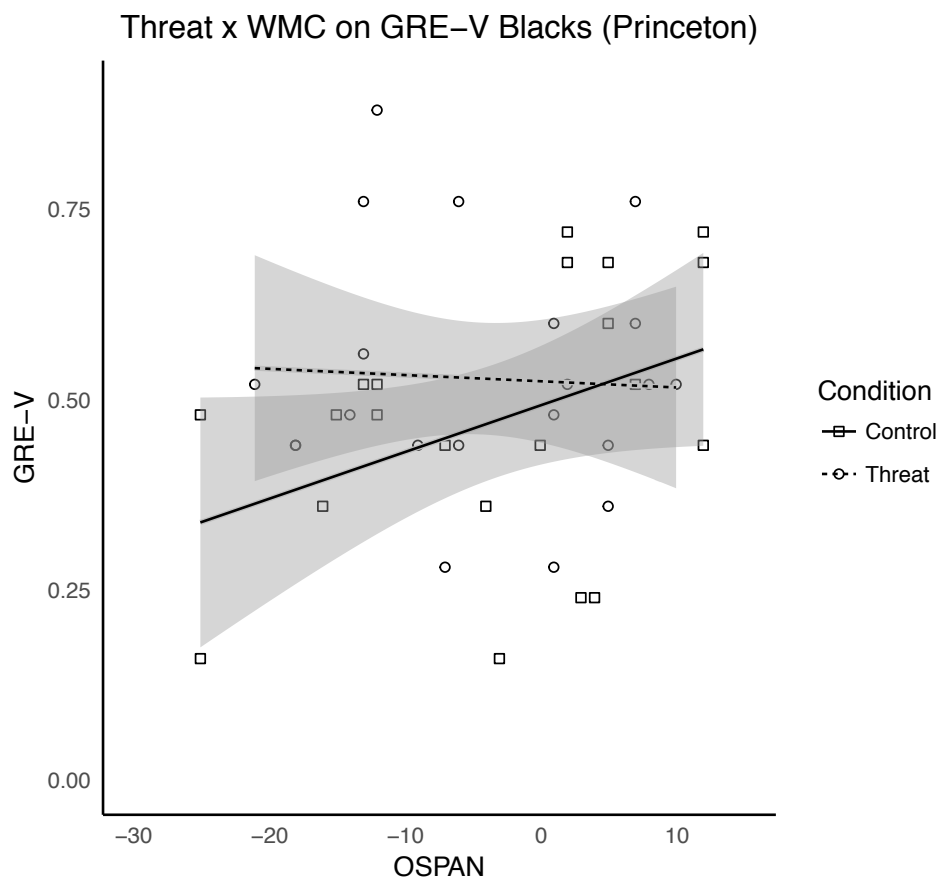


Figure 2.10. OSPAN Moderation Model for Threat on GRE-V

Looking at performance on the math GRE, in step one, the effects of OSPAN, $b = .0032$, $t(39) = 1.08$, $p > .05$, condition, $b = .017$, $t(39) = .27$, $p > .05$, and the model were non-significant, $F(2, 39) = .61$, Multiple $R^2 = .03$, $p > .05$.

In step two, neither the effects of OSPAN, $b = .0059$, $t(38) = 1.5$, $p > .05$, condition, $b = -.009$, $t(38) = -.13$, $p > .05$, their interaction, $b = -.006$, $t(38) = -1.08$, $p > .05$, nor the model, $F(3, 38) = .80$, Multiple $R^2 = .059$ were significant (see Figure 2.11).

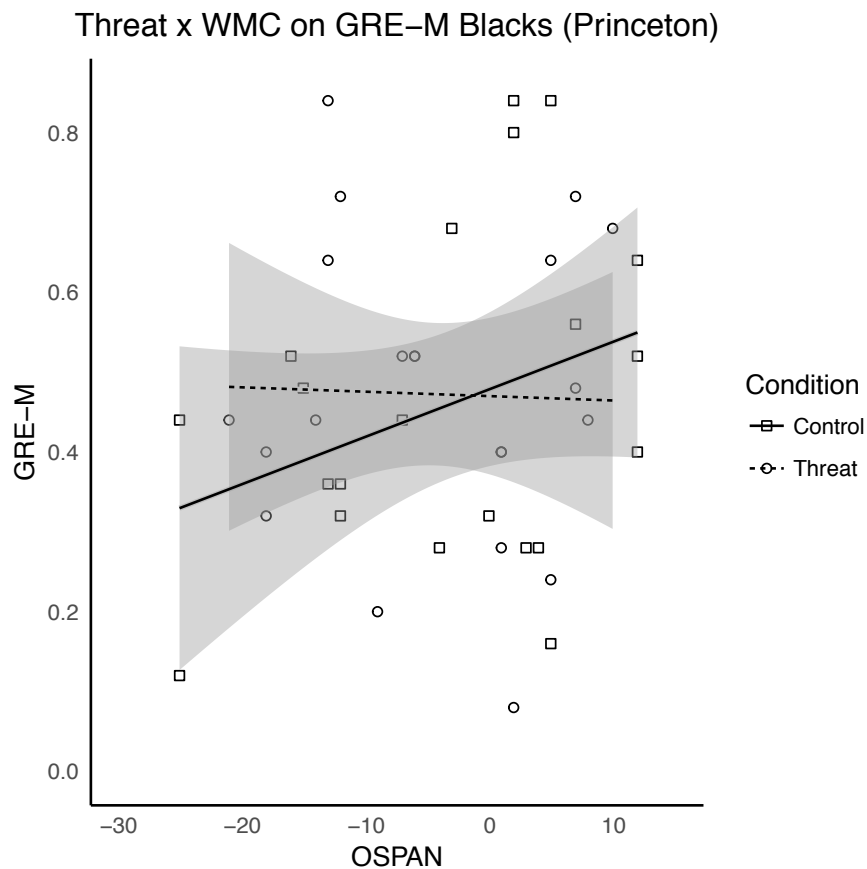


Figure 2.11. OSPAN Moderation Model for Threat on GRE-M

Quantitative Capacity & Intelligence Condition.

Baseline Effects.

Here, the initial t-test was conducted to detect differences in baseline WMC, with race as the independent variable and RSPAN as the dependent variable. Results indicated a significant difference in baseline RSPAN, $t(197) = -2.03$, $p < .05$, Cohen's $d = .34$, revealing that Black subjects performance was significantly lower on RSPAN at baseline relative to Whites (see Figure 2.12).

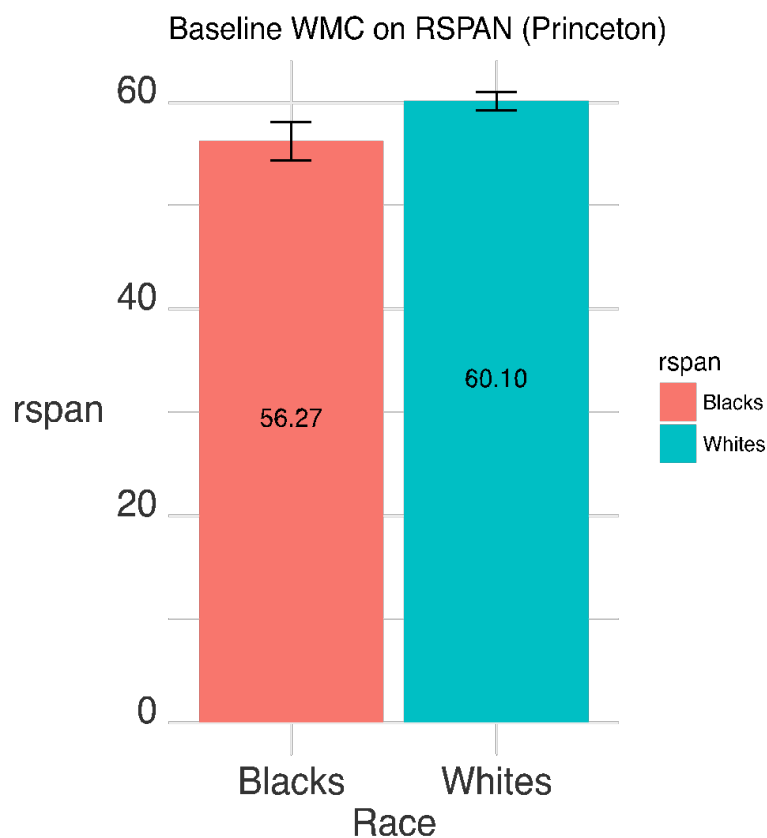


Figure 2.12. Trait/Baseline WMC for Race in the Math Threat Condition Order.

Threat Effects.

In the first 2 (race) x 2 (condition) ANOVA, looking at the effect of stereotype threat for “quantitative capacity,” OSPAN was the dependent measure. Results revealed non-significant effects of condition, $F(1, 195) = .66, p > .05, \eta_{\text{partial}}^2 = .0034$, race, $F(1, 195) = 2.3, p > .05, \eta_{\text{partial}}^2 = .012$ and their interaction, $F(1, 195) = .028, p > .05, \eta_{\text{partial}}^2 = .00015$ (see Figure 2.13).

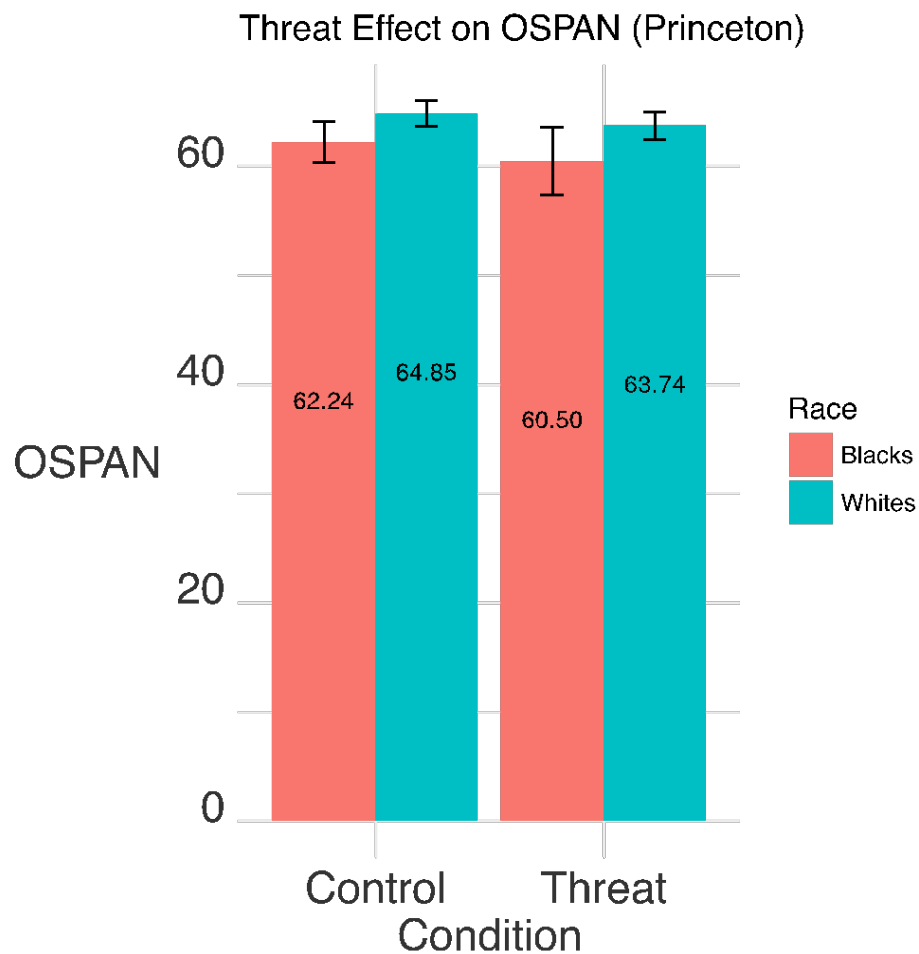


Figure 2.13. Effect of Stereotype Threat on OSPAN.

In the second 2 (race) x 2 (condition) ANOVA, looking at the effect of stereotype threat for “quantitative capacity,” GRE-M was the dependent measure. Results indicated a main effect

of race, $F(1, 195) = 14.2, p < .001, \eta_{\text{partial}}^2 = .066$, indicating that Whites ($M = .57, SD = .19$) significantly outperformed Blacks ($M = .45, SD = .17$) on the math GRE. There were no significant condition, $F(1, 195) = 1.09, p > .05, \eta_{\text{partial}}^2 = .0055$ or interaction effects, $F(1, 195) = 1.67, p > .05, \eta_{\text{partial}}^2 = .0085$, indicating there was no evidence of a threat effect. Instead, Black subjects appeared to receive a small boost in the threat condition compared to the control (see Figure 2.14).

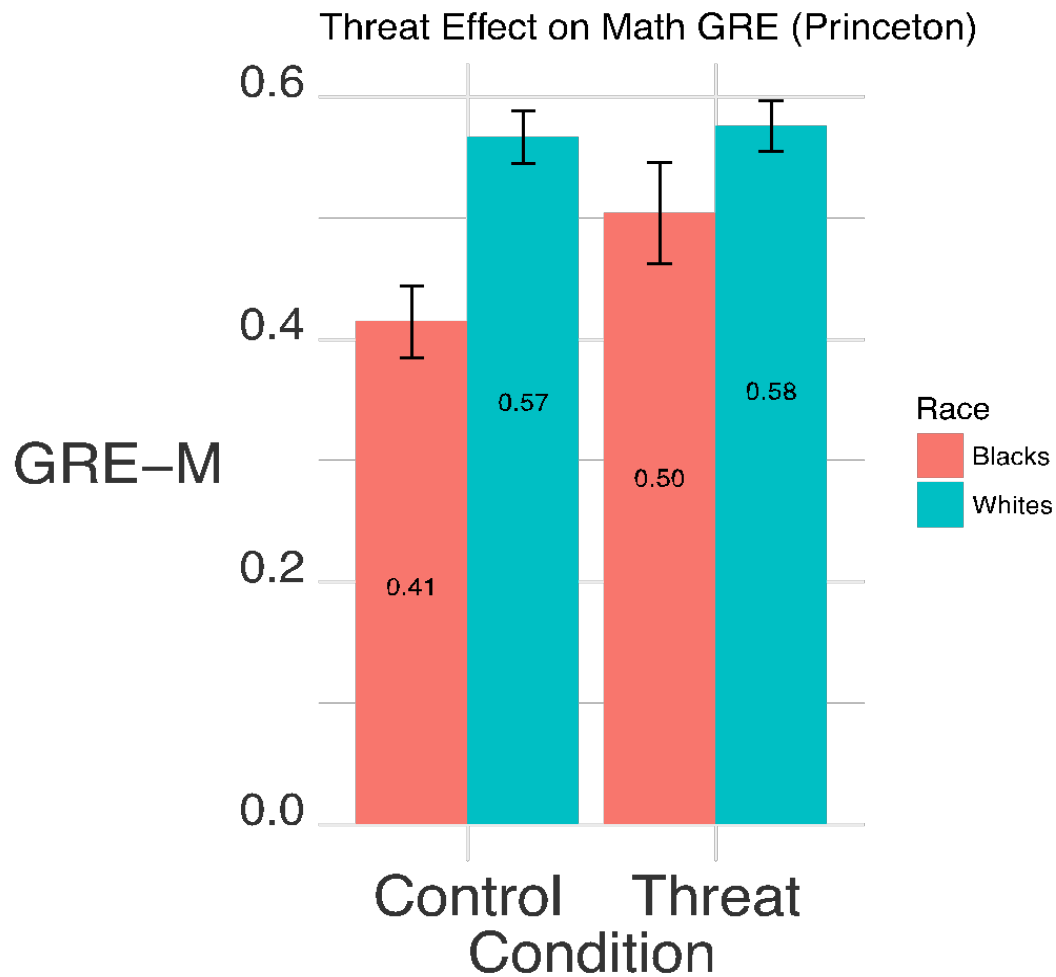


Figure 2.14. Effect of Stereotype Threat on GRE-M.

In the third 2 (race) x 2 (condition) ANOVA, looking at the effect of stereotype threat, GRE-V was the dependent measure. Results indicated a main effect of race, $F(1, 195) = 11.6,$

$p < .001$, $\eta_{\text{partial}}^2 = .055$, indicating that Whites ($M = .62$, $SD = .20$) significantly outperformed Blacks ($M = .51$, $SD = .17$) on the math GRE. There were no significant condition, $F(1, 195) = .12$, $p > .05$, $\eta_{\text{partial}}^2 = .00060$ or interaction effects, $F(1, 195) = 1.14$, $p > .05$, $\eta_{\text{partial}}^2 = .0058$, indicating there was no evidence of a threat effect. Instead, Black subjects appeared to receive a small boost in the threat condition compared to the control (see Figure 2.15).

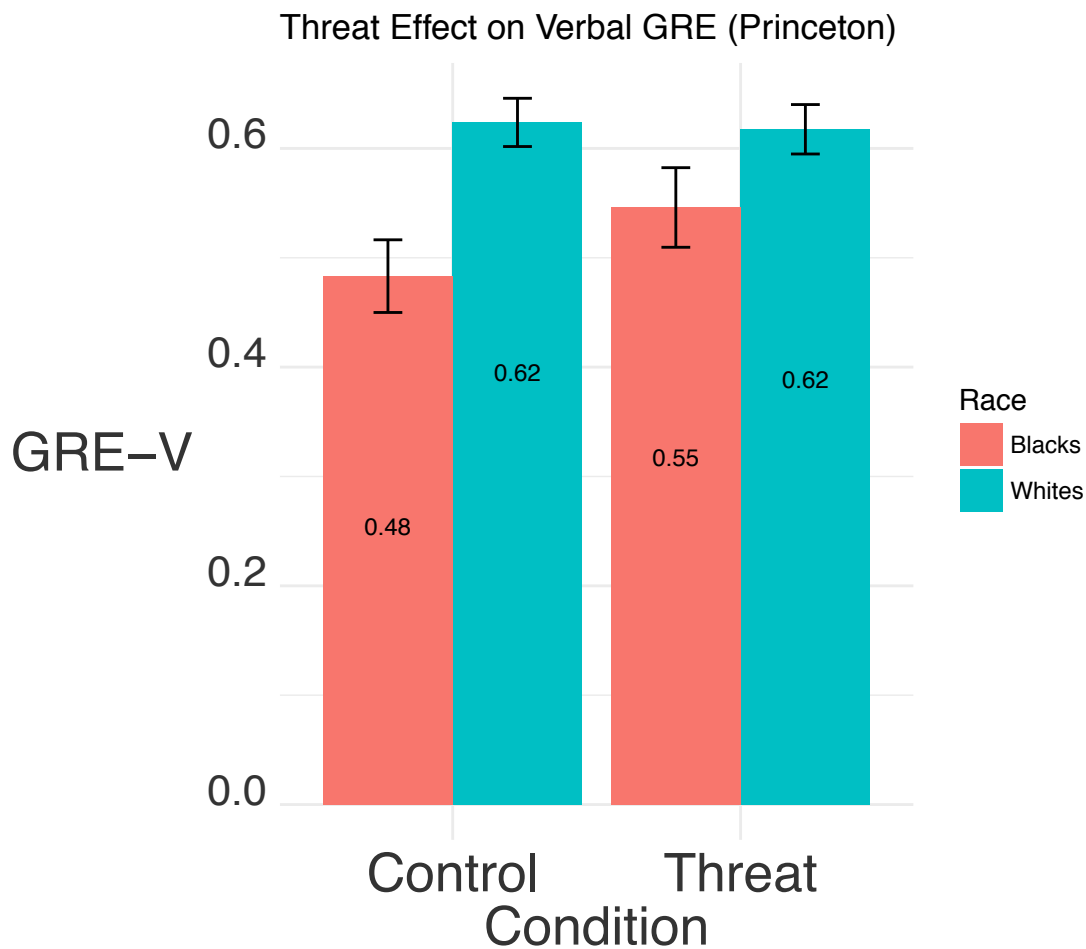


Figure 2.15. Effect of Stereotype Threat on GRE-V.

Mediation.

For only Black subjects, a mediation analysis with condition as the independent variable, OSPAN, the mediator variable, and GRE-M as the outcome variable was conducted. The direct effect of threat on GRE-M was trending toward significance, $t(43) = 1.8$, $b = .089$, $p = .08$ indicating that the threat condition had a predicted .09 unit increase on the GRE-M for black students. The effect of threat on OSPAN was non-significant, $t(43) = -.50$, $b = -1.7$, $p > .05$, indicating that the threat effect predicted a -1.7 unit decrease on OPSAN. The indirect effect of Threat on GRE-M when OSPAN was added to the model was significant (and increased from .089 to .099), $t(42) = 2.16$, $b = .0998$, $p < .05$, indicating the threat manipulation predicted a .099 increase on GRE-M. In the indirect path, the effect of OSPAN on GRE-M was also significant, $t(42) = 2.9$, $b = .0059$, $p < .01$. Moreover, Sobel's test revealed evidence for statistically significant mediation of OSPAN in the effect of threat on GRE-M, $Z = -.49$, $p = .04$ (see Figure 2.16 below).

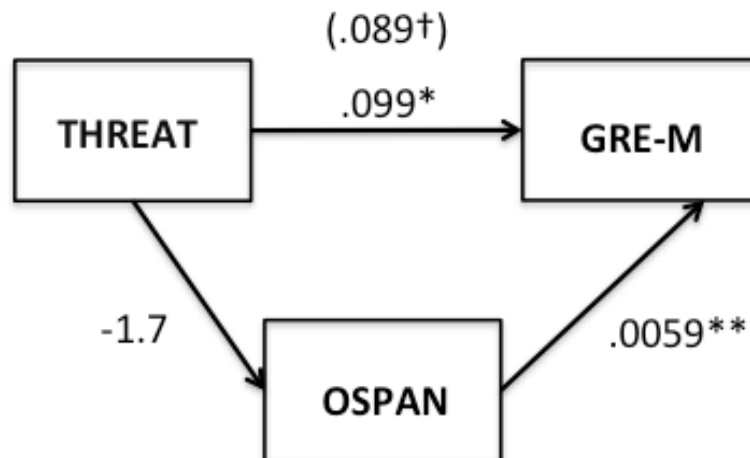


Figure 2.16. Model of OSPAN Mediating Threat Effect on GRE-M.

Using nonparametric bootstrapping methods to resample 1000 simulations, results indicated non-significant effects for the average causal mediation effect, parameter estimate = $-.0102$, CI $[-.062, .023]$, $p = .60$, however, the average direct effect of threat on GRE-M was significant, parameter estimate = $.099$, CI $[.014, .19]$, $p = .03$. Taken together, these findings reveal a lack of evidence that Black subjects' WMC (as assessed via the OSPAN) mediated the effect of threat on GRE-M (see Figure 2.17 below).

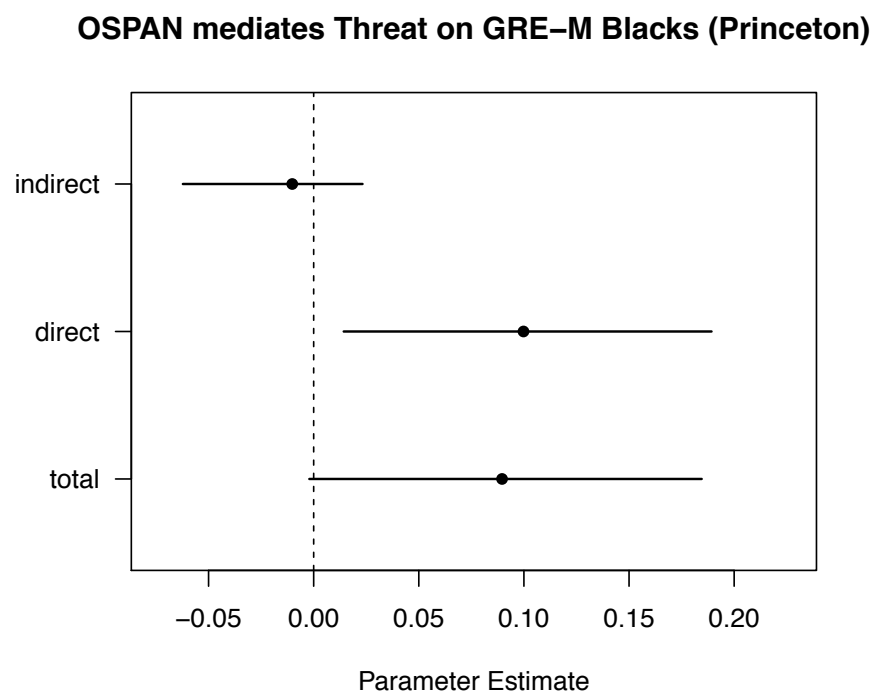


Figure 2.17. WMC Mediation Model for Threat on GRE-M using Bootstrapping.

Next, a mediation analysis with condition as the independent variable, OSPAN, the mediator variable, and GRE-V as the outcome variable was conducted. The direct effect of threat on GRE-V was not significant, $t(43) = 1.3$, $b = .063$, $p > .05$ indicating that the threat condition had a predicted .06 unit increase on the GRE-V for black students. The effect of threat on OSPAN was non-significant, $t(43) = -.50$, $b = -.17$, $p > .05$, indicating that the threat effect

predicted a -1.7 unit decrease on OPSAN. The indirect effect of threat on GRE-V when OPSAN was added to the model was not significant (and increased from .063 to .069), $t(42) = 1.5$, $b = .0696$, $p > .05$, indicating the threat manipulation predicted a .07 increase on GRE-V. In the indirect path, the effect of OPSAN on GRE-V was marginally significant, $t(42) = 1.9$, $b = .0039$, $p = .07$. Sobel's test confirmed a lack of strong evidence for statistically significant mediation of OPSAN in the effect of threat on GRE-V, $Z = -.48$, $p = .16$ (see Figure 2.18 below).

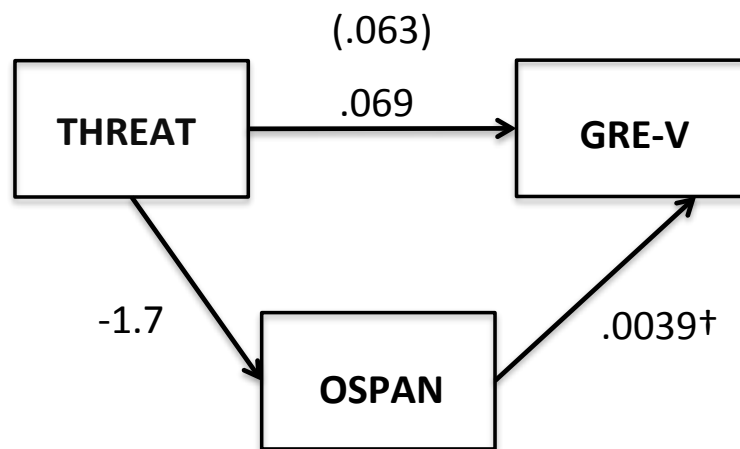


Figure 2.18. Model of OPSAN Mediating Threat Effect on GRE-V.

Using nonparametric bootstrapping methods to resample 1000 simulations, results indicated non-significant effects for the average causal mediation effect, parameter estimate = -.0068, CI [-.037, .019], $p = .60$. The average direct effect of threat on GRE-V was also not significant, parameter estimate = .0696, CI [-.035, .165], $p = .17$. Taken together, these findings reveal a lack of evidence that Black subjects' WMC (as assessed via the OPSAN) mediated the effect of threat on GRE-V (see Figure 2.19 below).

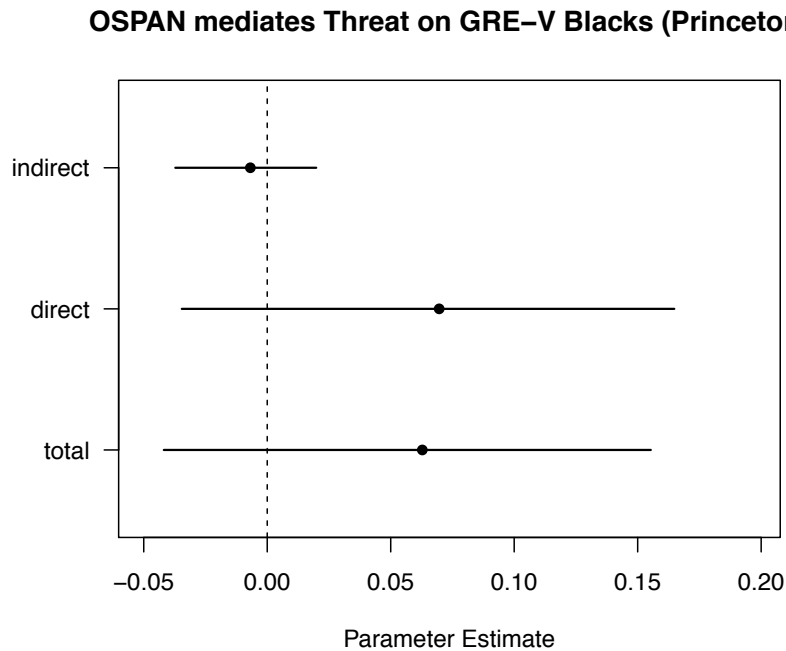


Figure 2.19. WMC Mediation Model for Threat on GRE-V using Bootstrapping.

Moderation.

To test whether trait WMC moderates the effect of stereotype threat on standardized testing in the quantitative capacity and intelligence threat condition, hierarchical regressions were conducted with the outcome measures of math and verbal GRE. In the first step of the analysis, RSPAN (mean centered) and condition were entered. In the second step, the product of RSPAN (mean centered) and condition (dummy coded with control = 0, threat = 1) was added in order to test for moderation.

Looking at the math GRE, in step one, the effect of RSPAN was marginally significant, $b = .0036$, $t(42) = 1.77$, $p = .08$, indicating that higher scores on RSPAN predicted higher scores on the math GRE. The effect of condition, $b = .073$, $t(42) = 1.48$, $p > .05$ was not significant but the model was significant, $F(2, 42) = 3.26$, Multiple $R^2 = .135$, $p < .05$.

In step two, the effect of RSPAN, $b = .00045$, $t(41) = .18$, $p > .05$ was not significant. The effect of condition, $b = .086$, $t(41) = 1.8$, $p = .08$ was marginally significant, indicating that students under stereotype threat had higher predicted scores on the math GRE. The interaction, $b = .0075$, $t(41) = 1.89$, $p = .07$, *approached significance*. Simple slopes tests revealed that when participants are under threat higher-WMC participants had higher predicted scores on the math GRE compared to those with lower WMC, $b = .0080$, $p < .05$; whereas, there was no significant change in GRE based on higher or lower WMC in the control condition, $b = .0005$, $p > .05$ (see Figure 2.20). The model was also significant $F(3, 41) = 3.51$, Multiple $R^2 = .20$, $p > .05$.

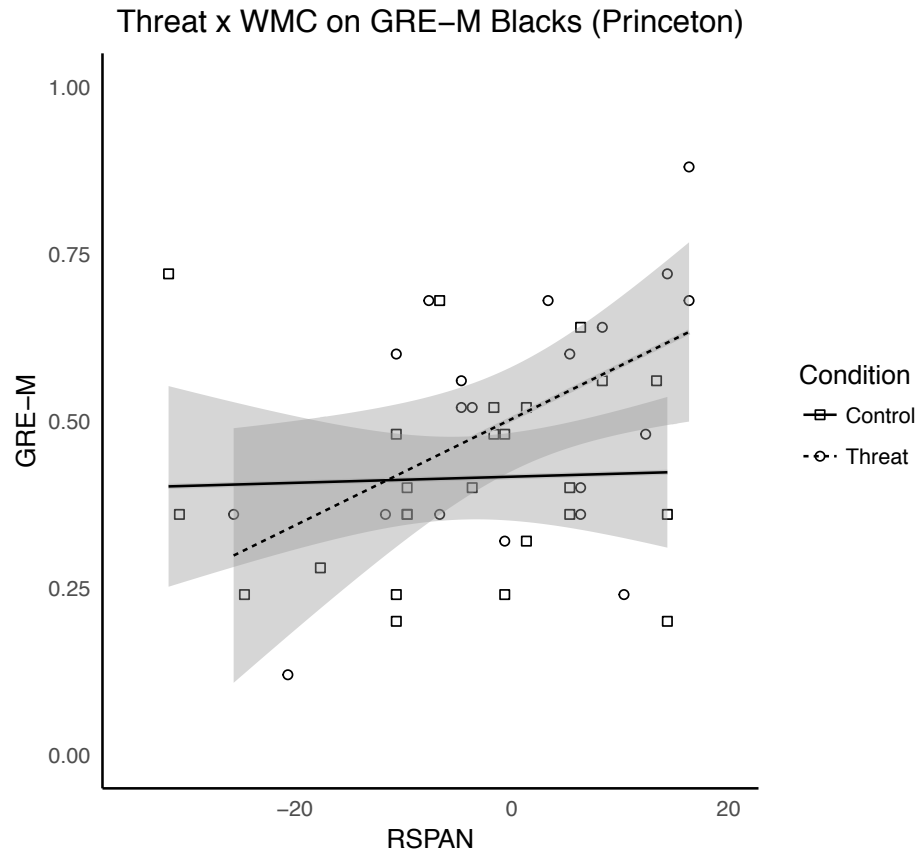


Figure 2.20. WMC Moderation Model for Threat on GRE-M

For verbal GRE, in step one, neither the effect of RSPAN, $b = .0018$, $t(42) = .90$, $p > .05$, the effect of condition, $b = .054$, $t(42) = 1.08$, $p > .05$, nor the model were significant, $F(2, 42) = 1.21$, Multiple $R^2 = .054$, $p > .05$.

In step two, the effects of RSPAN, $b = -.0013$, $t(41) = -.52$, $p > .05$, and condition, $b = .068$, $t(41) = 1.4$, $p > .05$ were non-significant. However, their interaction, $b = .0077$, $t(41) = 1.91$, $p = .06$ *approached significance*. Simple slopes tests indicated that under threat, higher WMC participants had higher predicted scores on the verbal GRE compared to those with lower WMC, $b = .0063$, $p < .05$. There was no significant change in predicted GRE scores in the control condition, $b = -.0013$, $p > .05$. (see Figure 2.21). The model also approached significance, $F(3, 41) = 2.06$, Multiple $R^2 = .13$, $p = .12$.

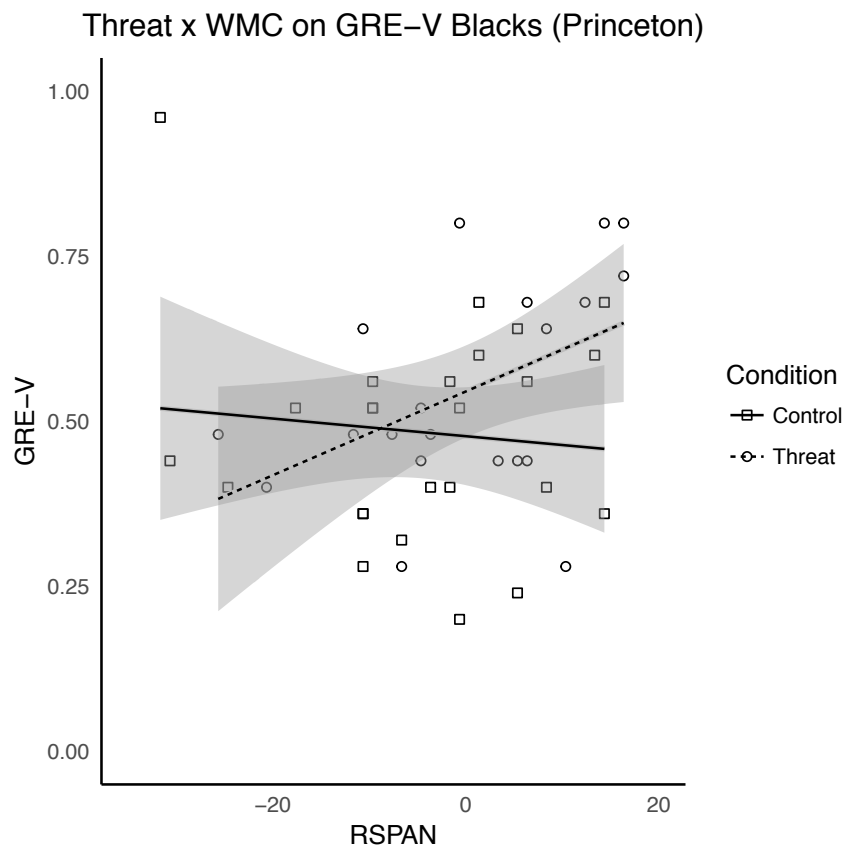


Figure 2.21. WMC Moderation Model for Threat on GRE-V

Experiment 1 Summary & Discussion

On the contrary to what was expected, these data did not show evidence of stereotype threat effects on performance. In addition, the present experiment did not show statistically significant evidence that state WMC mediates the effect of stereotype threat on standardized test performance.

There was some evidence that higher trait WMC moderated the effect of stereotype threat on standardized test performance, but only when baseline WMC was assessed on the RSPAN. These data suggest that higher trait WMC may protect Black students from the effect of stereotype threat for ethnicity but specifically threat in the quantitative domain. In the quantitative capacity and intelligence threat condition higher trait WMC predicted higher scores on the math and verbal GREs relative to lower WMC. However, when trait WMC was assessed via OSPAN there was no evidence for WMC moderating threat on either verbal or math standardized test performance. It was expected that higher WMC would help students combat stereotype threat across all conditions but instead the data revealed higher WMC did not protect students from threat in the verbal capacity and intelligence condition. Students who had higher or lower WMC scores on the OSPAN performed about the same in terms of predicted scores on the math and verbal GREs. This is an interesting finding because it implies that whether WMC moderates stereotype threat in high achieving students could depend on the kind of threat and the kind of task domain as to whether WMC is protective against stereotype threat effects. For example, this could be because of the task domain or it could be because students have different views about performance and ability in the mathematics domain compared to the verbal domain. At present, these data are unable to untangle whether differences in students' beliefs about their ability impacted the effect that stereotype threat might have on performance.

Overall, our participants appeared to be resilient to the effects of stereotype threat. This might've been the case because participants are highly motivated and high achieving students who were recruited to an ivy-league university based on their academic and standardized test performance. In addition, it may be possible that this sample could have higher trait WMC compared to samples of students at other universities—which would allow these students to remain resilient in the face of stereotype threat. However, it is also possible that this sample is also highly practiced with difficult academic tests. For this reason we would expect that most students would perform well on the working memory assessment and the standardized test measures.

The evidence in the data appeared to support a moderation account more than a mediation account. Here, we saw that there was some weak evidence that students' cognitive performance changed as a function of the level of their trait WMC, however the evidence in some cases also seemed to go in the opposite direction of the prediction such that in some cases higher WMC students performed the same under threat as lower WMC participants.

Based on the findings of Experiment 1 it is possible that coming from an ivy-league institution, this sample may consist of especially highly motivated and high achieving students. For these reasons it may be possible that these students are relatively better equipped to excel in the face of stigmatizing situations. Taken together, the aforementioned findings are presently unclear, however they provide some potential explanations for why and how people succumb to the effects of stereotype threat from a cognitive mechanism view. Because stereotype threat can be viewed as social-affective environmental phenomenon that disrupts cognitive performance these data highlight the potential benefit of having more cognitive resources at baseline in order to combat this phenomenon. In addition, with having such a potentially highly motivated sample

of students it might have been the case that these students were better equipped both cognitively and in their attitudes toward taking on academic challenges compared to other students. It may have also been the case that the ability to override negative beliefs about oneself and one's ability might have been present in this sample on a level that may not be the case in other groups. In order to assess this, the current work will consider beliefs about one's ability and other potential non-cognitive factors in addition to cognitive ones within contexts of stereotype threat.

Chapter 3

Exploring Mindset, Personality, and Working Memory Capacity during Contexts of Stereotype Threat

Cognitive Performance while Considering Non-Cognitive Factors

Recently, *Education Week* urged scholars to “note the flaws in current methods of assessing non-cognitive skills like *mindset* and *grit*.” Specifically, based on reports from the Brookings Institution and Dr. Evan Heit (division director of the NSF), researchers have been advised to design “more observational protocols that would help teachers and researchers understand what grit or a growth mindset looks like in day-to-day practice (Education Week, Vol. 35, Issue 28, Page 6).” This alludes to the point that the use of non-cognitive measures and methods for designing interventions to support academic achievement will only work to the extent that we understand these measures and methods. This does and should give educators and administrators pause when using these interventions for accountability purposes—if we do not have a good understanding of how the theories underlying these measures work then it is clear that more research is needed. In order to understand how these interventions work in practice, one should begin with taking a closer look at self-report measures in addition to post-intervention behavior within and across individuals.

Understanding the issues with measuring non-cognitive skills and how these skills might impact behavior is one approach for unveiling biases in cognitive performance that are driven by something other than cognitive ability alone. In addition, honing in on the mechanisms at play during environments when stigmatized individuals are asked to perform is another approach for combating biases or issues related to underachievement of minority group members. Taken

together, the implementation of these two approaches will allow for better methods for resolving these problems. In considering this, the current work will aim to focus on how in spite of potential stigma minority group members achieve success during difficult social and environmental situations. This work will be approached based on a new theoretical model of stereotype threat and performance, considering the role of non-cognitive factors of implicit beliefs about ability and individual differences in personality traits like conscientiousness and grit. To build up to this new theoretical model, first, relevant literature on non-cognitive measures of ability and cognitive performance will be discussed. Next, findings from previous research and the current data will be used to motivate this new theoretical model for investigating stereotype threat effects. The ultimate aim of this research is considering cognitive and non-cognitive measures in order to develop more comprehensive models for mitigating stereotype threat effects.

Stereotype threat appears to be about how you think about your ability and how you can demonstrate this in an environment with real or imagined people (who may view your perceived ability negatively). The key is about reconciling the difference between what you are actually capable of and what you think you are capable of in the “face” of others. The subsequent performance decrement that emerges is the stereotype threat effect. Because there is variation in actual ability as well as variation in beliefs about ability, both of these should be at play when someone is put under stereotype threat. If one feels anxiety when put in a situationally pressured scenario then one must consider where that feeling comes from. Moreover, if the situation is about demonstrating ability then how could the way one thinks about ability have no influence? If this is true then it should be possible to assess the threat effect in several ways. Generally, it can be measured in terms of the resources it takes up for a person to complete a task when in a

stigmatizing environment. Furthermore, as evidenced above, stereotype threat involves cognitions about performance based on one's own views and the views of real or imagined others. As this is the case, it seems relevant to consider how one's views about ability might influence the efficacy with which the threat effect impacts performance. As such, literature on the topic of beliefs about ability will be discussed.

Mindset: Implicit Beliefs about Ability

Although there are several strategies for approaching questions surrounding cognitive performance, only recently has there been greater attention paid to how people's *beliefs* about their ability might impact their *actual* performance. The most prominent theoretical perspectives are the fixed vs. malleable framework of intelligence (Bandura & Dweck, 1985; Dweck & Leggett, 1988), and the topic of motivation and the self (Dweck, 1999; Heckhausen & Dweck, 1998; Elliott & Dweck, 2005). Specifically, there is growing interest in the notion of *mindset*, which refers to the idea that people maintain implicit theories about their own ability (Bandura & Dweck, 1985; Dweck & Leggett, 1988). Indeed, certain academic domains have been found to promote strong biases toward endorsing a fixed rather than a malleable view of ability. Some examples include areas of: intelligence, domain general and domain specific mathematics, and verbal ability (Dweck & Leggett, 1988; Steele & Aronson, 1995; Cury et al., 2008; Onwuegbuzie & Wilson, 2003).

Previous research revealed that those who hold beliefs that their successes are the result of some innate, fixed, ability (i.e., "entity theorists") and in turn, do not relish challenges (Dweck, 2006). Often, entity theorists' confidence is shaken more than others when confronted with difficulties. Based on previous research students who subscribe to more malleable theories of intelligence (i.e., "incremental theorists") have been shown to be better equipped to take on

new challenges and are able to remain diligent when confronted with difficulties. Today, the fixed vs. malleable framework of intelligence has transformed into what is considered the motivational mindsets of achievement.

The crux of mindset involves beliefs about control: entity theorists tend to believe that people have less control over their abilities, whereas, incremental theorists believe that people have the power to enhance their abilities, if desired (Dweck, 2011). In addition, previous research on the psychological function of implicit theories has revealed that people are receptive to mindset interventions because these theories shape understanding of how people understand and exert control in the world (see Plaks & Stecher, 2007).

From a student-centered perspective of general cognitive abilities, on average, 40% of subscribe to entity theory, 40% to incremental theory, and 20% fail to consistently promote either theory. Based on prior research, students who subscribe to entity theory have not been found to approach difficulties gracefully; instead, they tend to question their abilities (Licht & Dweck, 1984a; Leggett, 1985). The students who have demonstrated themselves to be the most capable suffer the most gravely. Additionally, in reviewing this literature, Yeager and Dweck (2012) concluded that prior successes do not foster resilience for entity theorists, but rather those achievements make these students dissociate with tasks that promote future challenges (Dweck, 1999). In a sense, this might indicate that holding an entity or fixed view about ability *may only be beneficial for those who have had and continue to have present and future successes*.

In addition, contrary to how we think about praise, results indicated that when children were praised for intelligence they focused more on performance goals rather than learning, were less persistent, and exhibited poorer performance relative to children praised for effort (Mueller & Dweck, 1998). These findings are useful in the context of stereotype threat because as these

mindsets shape beliefs about ability and achievement, the empirical question remains whether such beliefs might help alleviate the threat effect on cognitive performance, and, if so, then for which students and under what circumstances?

Cury et al. (2008) investigated whether implicit beliefs of ability are related to and/or potentially can influence cognitive performance on intelligence tasks. Using correlational and experimental methods, these researchers tested whether fixed beliefs of ability impair test performance and whether worry and time practicing mediate the relationship between implicit beliefs and test performance. Results indicated support for the hypothesis that fixed beliefs of ability lead to decreased performance even when initial failure was not encountered. Moreover, entity theorists have more pressure and worry about performing because there is more to lose during failure—a finding similar to literature on having a prevention focus (for review the reader is directed to Higgins et al., 2001). Their attribution was based on innate ability, thus entity theorists should employ strategies to protect them from internal attributions of failure. The conclusion was that based on a continuum of fixed to malleable, as fixed beliefs increased IQ performance was decreased. However, these results only address the relationship between implicit beliefs and cognitive performance on an intelligence outcome measure. Next, literature on the impact of non-cognitive measures will be considered in contexts of stereotype threat.

Theoretical Motivation: Stereotype Threat and Non-Cognitive Factors

In the 2014 New York Times article *Who Gets to Graduate?*, Paul Tough described how research on graduation rates found that the problem is not about students simply making it to college. Instead, core issues emerge after students are already there. For example, the article states that 40% of students have not earned their degree after 6 years. Additional research indicates that whether they graduate or not depends on how much money their parents make. In

essence, rich kids graduate and poor working class kids do not. The article supplies that the factors that students struggle with most during college are centered on their sense of belonging and their beliefs about their ability.

Furthermore, during times of intellectual challenges how students think about themselves, their status and position in the college and their abilities are of great importance. If students begin to question their belonging and their abilities a self-fulfilling prophesy of learned helplessness could develop where they persuade themselves that they are incapable. Before long, these students may start to believe that they do not belong in college. In an effort to combat these core issues, the article states that researchers launched messages that appealed to university norms and that allowed students a sense of autonomy. The goal of the messages involved self-persuasion—it was necessary for students to write an essay to persuade others, but in this process they were trying to persuade themselves. The implications were that the process of reflecting on a hopeful message and later aiming to convince other students of the message allowed students to digest and recognize the same possibilities for themselves. This self-persuasion tactic functions as a means for protecting oneself from over-interpreting failures and challenges in the future and has been shown to work in previous research on stereotype threat and belonging (see Walton & Cohen, 2007; Walton & Cohen, 2011).

The process of a student changing the meaning attributed to academic challenges makes it possible to shift their perspective and allow them to approach future challenges with a different belief system (also see Yeager & Walton, 2011 for review). Moreover, the key to these belief altering tactics is that they are properly timed and implemented in a way that keeps students from feeling targeted as “low achieving.” The question remains whether there are individual differences in the way non-cognitive skills and abilities impact cognitive performance during

moments of academic challenges and specifically contexts that activate stereotype threat. And, if we can observe differences in cognitive performance based on cognitive and non-cognitive abilities during situations of stereotype threat, how exactly do such abilities impact performance within groups and across different individuals?

Statement of the Problem

How do implicit beliefs impact threat? Previous work shows some evidence of implicit beliefs reducing threat effects, but threat was never directly manipulated in these studies (see Aronson, 1999; Aronson, Fried & Good, 2002). Because stereotype threat effects are grounded in one's beliefs about ability and overcoming stigmatizing environments in order to perform well, this would imply that beliefs about ability are both useful and informative for thinking about stereotype threat.

The results of Experiment 1 and those from a conceptual replication through a priming induction (c.f. Daker & Conway, unpublished manuscript) may provide evidence that a form of subconscious belief about ability was activated across different aspects of group identities in high achieving, highly motivated samples of ivy-league students. Moreover, Daker and Conway argue that the extent to which stereotype threat impacts behavior depends on the individual level of identification people have with their social and group identity. In this sense, level of identification with the social/group identity is of great importance. In fact, it has been shown to have a moderating effect on performance when individuals are under stereotype threat (see Schmader, 2002).

In summary, the results from Daker and Conway and the current work in Experiment 1 suggest that stereotype threat effects are more complex than merely performance decrements for those with stigmatized identities. For example, factors like WMC in terms of the mental space

that an individual has to devote to the task at hand is an important consideration for thinking about the extent to which stereotype threat may disrupt one's performance. However, there is even more to consider in terms of the individual differences in how the threat impacts one person to the next—such as important personality and attitudinal variables that may moderate differences in the impact of stereotype threat. Thus, in subsequent experiments, the aim is to address the role of implicit beliefs and other non-cognitive factors across different samples of students (ivy-league and non-ivy-leaguers') in contexts of both race and gender threat. As observed in previous threat literature if the effects are generalizable they should be consistent across different kinds of threat inductions (see Wheeler & Petty, 2001 for review). Looking at different kinds of stereotype threat effects across different samples of students allows the potential for greater generalizability and a more comprehensive story about the role of such non-cognitive factors during contexts of stereotype threat within and across different groups.

The goals of the current experiment are twofold: first, to better understand the cognitive mechanisms underlying performance when targets of stereotyping are exposed to stigmatizing situations (explored in Experiment 1) and second, to use results of the aforementioned research to propose a re-conceptualization of stereotype threat which includes the potential theoretical contribution of both cognitive and non-cognitive factors and abilities as potential paths for alleviating the performance decrements (i.e., stereotype threat effects) we observe in targets exposed to environments which are stigmatizing.

Current Hypotheses

Previous research on mindset describes a malleable mindset as a belief that ability can be changed, whereas, a fixed mindset as a belief that ability is somehow fixed and is unlikely to change (e.g., intelligence is viewed as an innate ability where large amounts of training may not

necessarily be beneficial). The first aim is to uncover whether there is a relationship between implicit beliefs and the effect of stereotype threat on performance. This question will be explored in Experiment 2 from an individual differences perspective in order to uncover whether ivy-league compared to students at a selective non-ivy hold specific types of mindsets and whether those are correlated with better or worse performance across varying domains of threat activation. Furthermore, previous research on ego depletion has demonstrated that the impact of one's beliefs can, in fact, dictate future behavior (see Job, Dweck, & Walton, 2010). This idea will be further explored in Experiment 2.

Experiment 2: Exploring The Relationship between Mindset and Threat

The purpose of this experiment is to explore the relationship between implicit beliefs and stereotype threat effects. To do this, mindset in addition to other non-cognitive measures of personality and grit were measured. It is expected that during contexts of stereotype threat activation, when mindset is more malleable this will be associated with an attenuated threat effect on performance. In addition, we will explore relationships between individual differences in personality and grit measures and threat effects—exploring whether higher grittiness and/or conscientiousness predicts better scores on performance measures under stereotype threat.

Method

Participants

A total of 166 undergraduate students at Towson University were recruited for the study. Subjects were invited to participate if they were at least 18 years of age and self-identified as White (98 students, 66 female) or Black (68 students, 55 female). All subjects were required to be native English speakers and received credit toward a course requirement.

Design

While exploring the relationship of mindset and other non-cognitive factors to stereotype threat effects on performance, another aim of Experiment 2 is to investigate the relationship of WMC and stereotype threat in a new sample of students from those in Experiment 1. Thus, the same experimental design as Experiment 1 was employed—a 2 (Condition: Threat vs. Control) x 2 (WMC Span Task Order: Verbal Capacity & Intelligence vs. Math Capacity & Intelligence) x 2 (Race: White vs. Black) factorial design. Similar to Experiment 1, the role of WMC will be tested through the series of ANOVAs, mediation and moderation analyses for Black subjects based on math or verbal stereotype threat. In addition, the relationship of non-cognitive factors of mindset, grit and conscientiousness will be explored (see Appendices A and B below). Because the non-cognitive factors are continuous, the relationship of these variables with stereotype threat will be explored by conducting a series of regression analyses in order to assess the direction and magnitude of their predictive power when looking at differences in standardized test performance under stereotype threat.

Procedure

First, participants entered the lab and completed a baseline (trait) WMC measure, followed by the same threat manipulation or control from Experiment 1. Next, they completed a state WMC measure along with the quantitative and verbal sub-sections of the GRE. At the end of the experimental session, participants completed experiment surveys assessing comfort, fatigue, and anxiety. Finally, after the experiment ended participants also completed a survey battery of non-cognitive measures including: mindset, personality, and perseverance/grit measures.

Tasks

In addition to the tasks described in Experiment 1, participants completed several non-cognitive measures (see Appendix A).

Results

Data Preparation & Analytic Approach

As in Experiment 1, in an attempt to simplify the presentation of results here two “sets” of results are presented, one for each task order condition—for stereotype threat for either “verbal capacity and intelligence” or “quantitative capacity and intelligence.” In addition, for each of the sets of analyses below, the homogeneity of variance assumptions were tested with *Levene’s* test and where the test was statistically significant, *Bonferroni* corrections were used and reported below. In cases where *Levene’s* tests were non-significant these results were not reported and normal t-tests and tests of analysis of variance are reported below.

Also, in each regression analysis below, levels of condition were dummy coded with control = 0, threat = 1. Because the threat effect was expected to manifest as a performance decrement for Black subjects, only their data were analyzed for WMC moderation effects on the standardized tests.

Although the threat effect is not expected to decrease performance for White subjects and the effect is most theoretically relevant for Black subjects, due to high attrition rates for the non-cognitive factors, moderation analyses for these outcome measures were tested with White and Black subjects in order to retain statistical power in these tests. Additionally, moderation results for the non-cognitive factor of mindset are reported in the text below because of their relevance to the subsequent experiment; results for other non-cognitive factors of grit and personality are reported in the supplementary information.

Summary Statistics and Correlations

Looking at the general trends in the data, descriptive statistics revealed the measures of WMC (i.e., ospan and rspan) were strongly positively correlated with each other as well as moderately positively correlated with measures of standardized test performance (i.e., math and verbal GREs). These trends in the data are reported in Table 3.1 below.

Table 3.1. *Summary Statistics and Correlations of Variables in Experiment 2*

	1	2	3	4	5	6	7
1. ospan	--						
2. rspan	.725***	--					
3. gremProp	.174*	.189*	--				
4. grevProp	.202**	.214**	.435***	--			
5. STAI	-0.046	-.155*	0.008	-0.018	--		
6. gremAttProp	-.199**	-.193**	0.112	-0.035	-0.032	--	
7. grevAttProp	-.151*	-.218***	0.134	0.077	-0.113	0.584***	--
<i>M</i>	55.21	50.79	0.27	0.26	12.8	0.87	0.91
<i>SD</i>	14.9	14.89	0.12	0.15	3.86	0.16	0.14

Note. ospan = Operation Span, rspan = Reading Span, gremProp = Proportion Correct on Math GRE, grevProp = Proportion Correct on Verbal GRE, gremAttProp = Proportion Attempted on Math GRE, grevAttProp = Proportion Attempted on Math GRE, STAI = Spielberger State-Trait Anxiety Inventory. *** $p < .001$, ** $p < .01$, * $p < .05$.

Verbal Capacity & Intelligence Condition.

Baseline Effects.

An independent t -test was conducted to detect differences in baseline WMC, race was treated as the (quasi) independent variable and OSPAN as the dependent variable. There was no significant difference in baseline OSPAN based on race, $t(83) = -1.38$, $p > .05$, Cohen's $d = .31$, indicating that although White subjects performed about 4 points higher on average this difference was not statistically significant (see Figure 3.1).

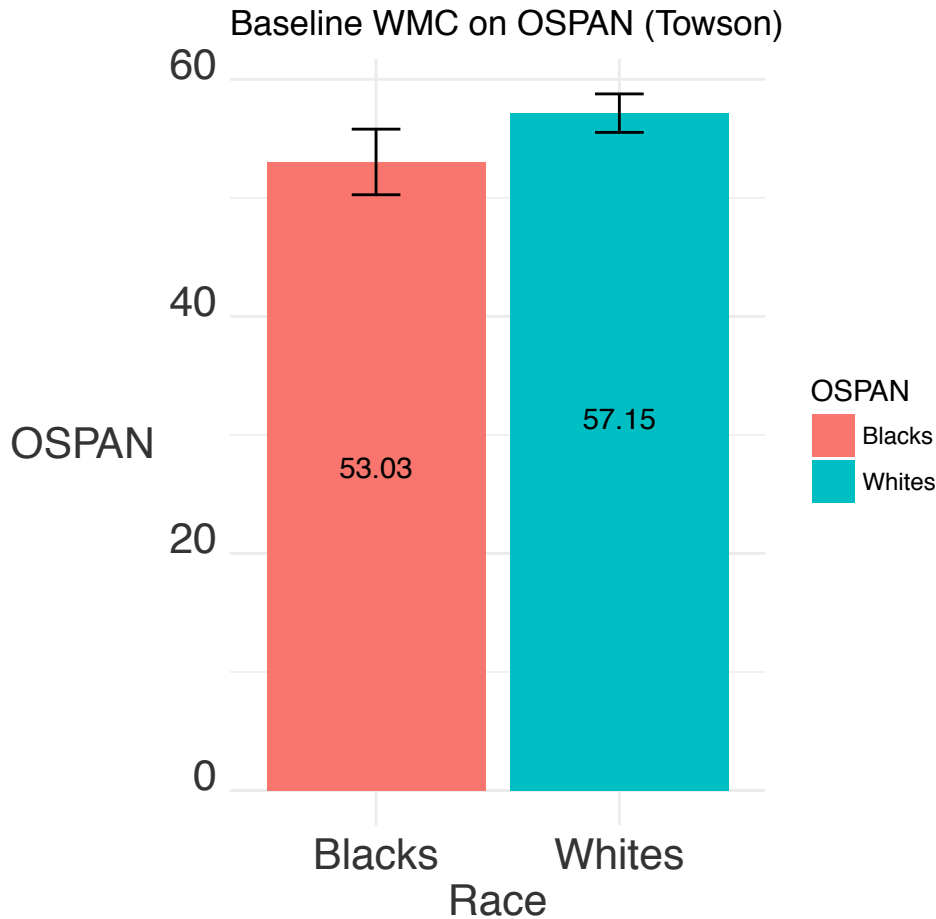


Figure 3.1. Baseline WMC on OSPAN (Towson)

Threat Effects.

Looking at the effect of stereotype threat for “verbal capacity”, I ran a 2 (race) x 2(condition) ANOVA with RSPAN as the dependent measure. . The effects of condition, $F(1, 81) = .61, p > .05, \eta_{\text{partial}}^2 = .024$ and race, $F(1, 81) = 2.3, p > .05, \eta_{\text{partial}}^2 = .0074$ were non-significant. However, results indicated a significant interaction of race and condition, $F(1, 81) = 3.98, p < .05, \eta_{\text{partial}}^2 = .047$. *Tukey's* test of honest significant differences revealed that White subjects in the threat condition did not differ from White subjects in the control condition ($M \text{ diff} = 2.17, p = .95$). In contrast, Black subjects under threat tended to experience a performance

decrease relative to Blacks in the control condition, however this finding was not significant ($M_{diff} = -11.4, p = .17$). These results are demonstrated in Figure 3.2 below.

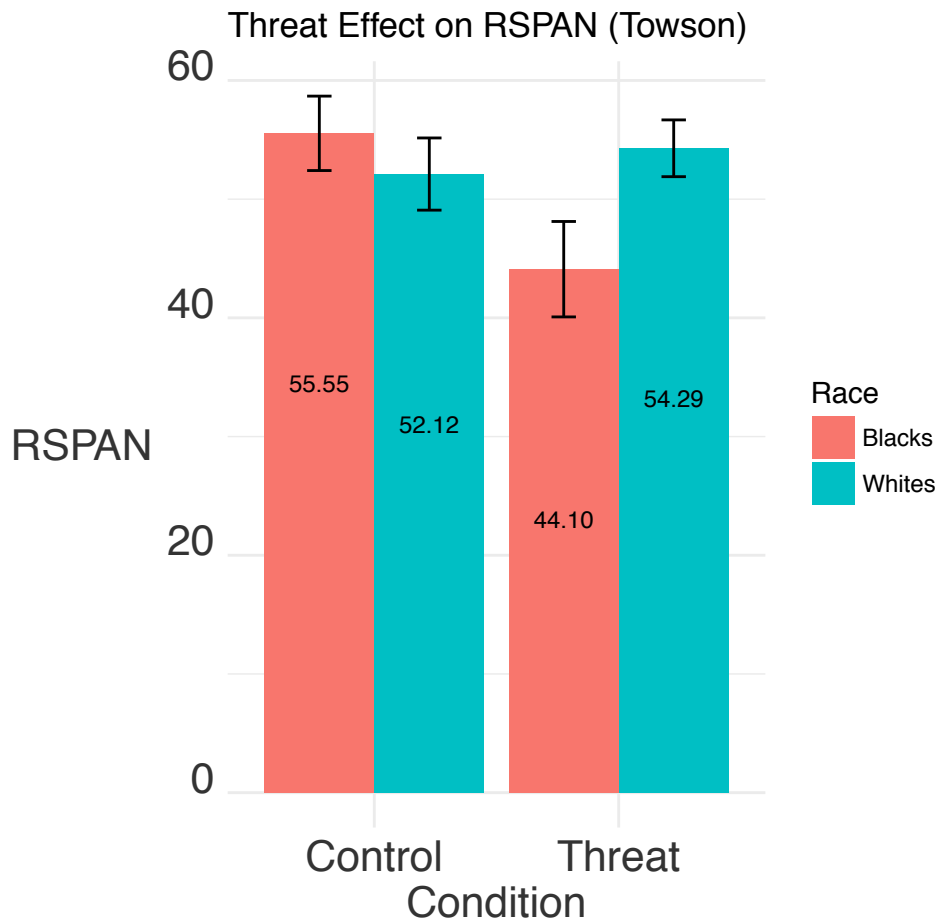


Figure 3.2. Threat Effect on RSPAN (Towson)

In the second 2 (race) x 2 (condition) ANOVA looking at stereotype threat for “verbal capacity”, GRE-V was the dependent measure. *Levene’s* test revealed a significant homogeneity of variance violation, $F(3, 81) = 3.1, p < .05$. Thus, a *Bonferroni* correction was used to compute the new significance criteria of .0083. Results indicated only a main effect of race, $F(1, 81) = 5.3, p = .02, \eta_{\text{partial}}^2 = .065$ —but based on the *Bonferroni* criteria this effect was non-significant; according to the pattern of the means, White subjects ($M = .28, SD = .18$) performed

higher than Black subjects ($M = .20$, $SD = .10$) on the verbal GRE. The effects of condition, $F(1, 81) = .66$, $p > .05$, $\eta_{\text{partial}}^2 = .0082$ and the interaction, $F(1, 81) = .022$, $p > .05$, $\eta_{\text{partial}}^2 = .00026$ were non-significant. These findings are demonstrated in Figure 3.3 below.

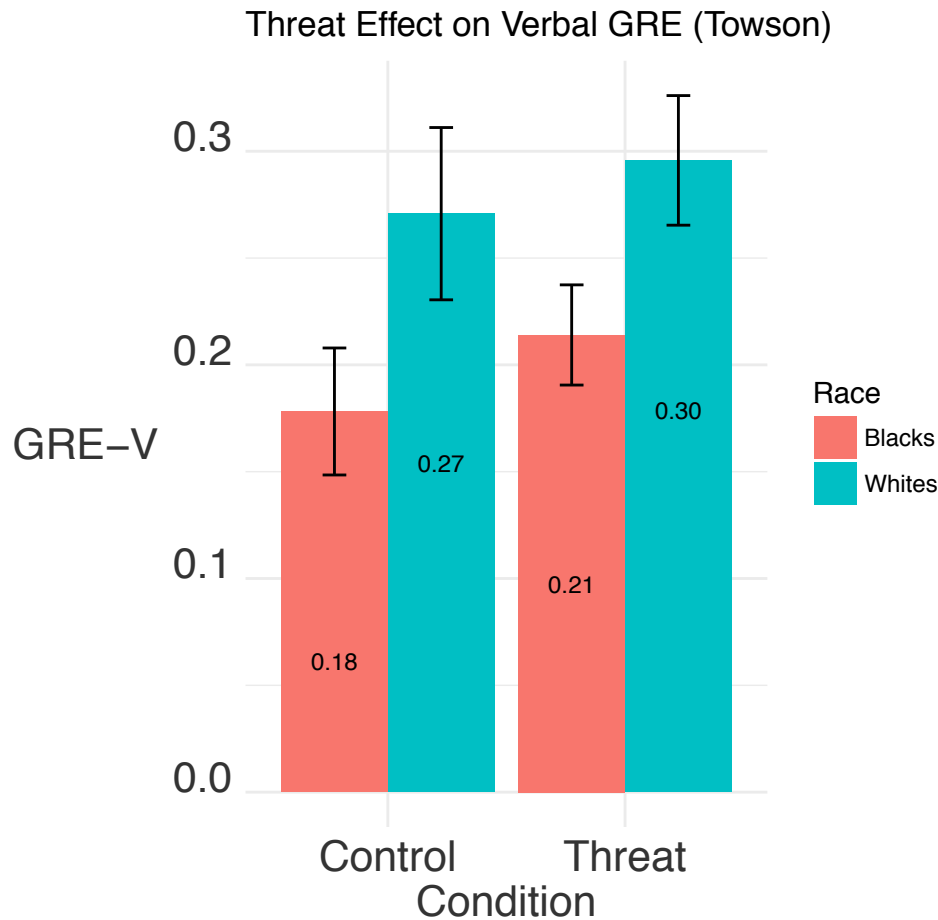


Figure 3.3. Threat Effect on GRE-V (Towson)

In the third 2 (race) x 2 (condition) ANOVA looking at threat effects, GRE-M was the dependent measure. Results indicated only a trending main effect of race, $F(1, 80) = 2.5$, $p = .12$, $\eta_{\text{partial}}^2 = .035$ indicating that White subjects performed better ($M = .29$, $SD = .14$) than Black subjects ($M = .25$, $SD = .10$) on the math GRE. The effects of condition, $F(1, 80) = 1.06$, $p >$

.05, $\eta_{\text{partial}}^2 = .013$ and the interaction, $F(1, 80) = .31, p > .05, \eta_{\text{partial}}^2 = .0038$ were non-significant. These findings are demonstrated in Figure 3.4 below.

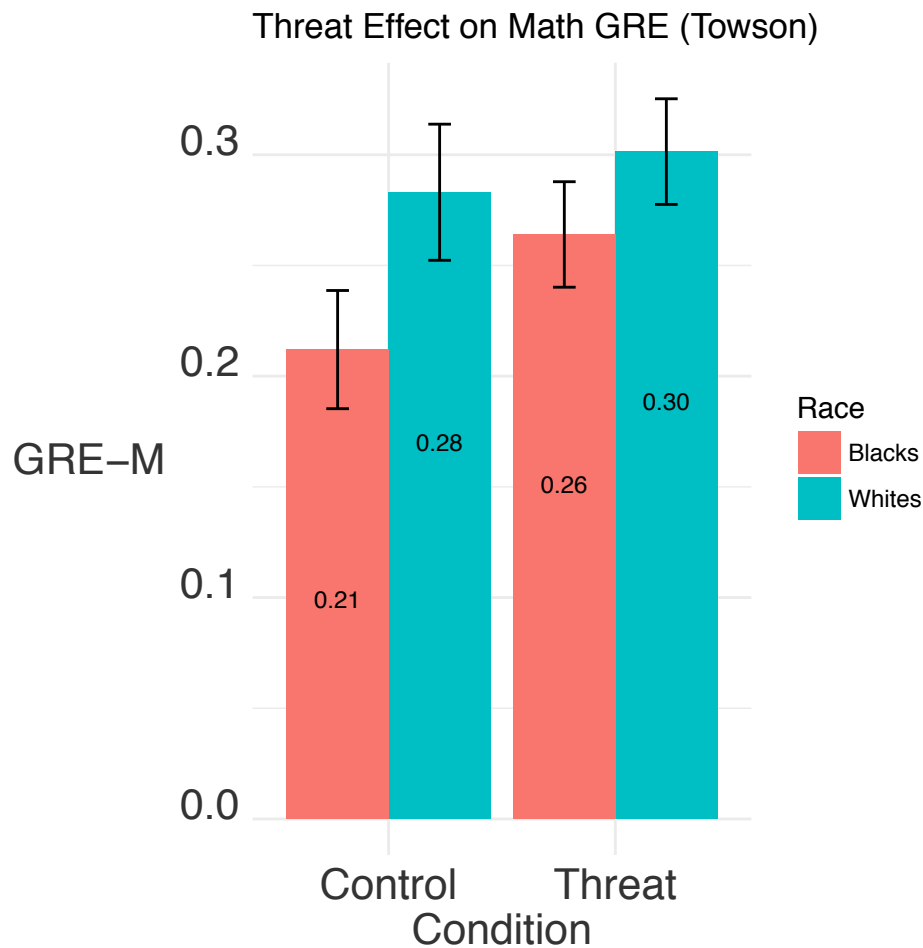


Figure 3.4. Threat Effect on GRE-M (Towson)

Mediation.

A mediation analysis for only Black subjects with condition as the independent variable, RSPAN, the mediator variable, and GRE-V as the outcome variable was conducted. The direct effect of threat on GRE-V was non-significant, $t(29) = .93, b = .036, p > .05$ indicating that the threat condition had a predicted .036 increase on the GRE-V for black students. The effect of threat on RSPAN was trending toward significant, $t(29) = -1.93, b = -11.5, p = .06$, indicating

that the threat effect predicted an -11.5 unit decrease on RPSAN. The indirect effect of Threat on GRE-V when RSPAN was added to the model was non-significant, $t(28) = .36$, $b = .015$, $p > .05$. In spite of the non-significant effect, a decrease in the coefficient for threat (from $b = .036$, to $b = .015$) was observed, indicating that the effect of stereotype threat on verbal GRE was at least partially reduced as a result of including RSPAN in the model. In the indirect path, the effect of RSPAN on GRE-V was also non-significant $t(28) = -1.6$, $b = -.0019$, $p > .05$. *Sobel's* test further supported the lack of strong evidence for statistical mediation of RSPAN in the effect of threat on GRE-V, $Z = 1.12$, $p = .71$ (see Figure 3.5).

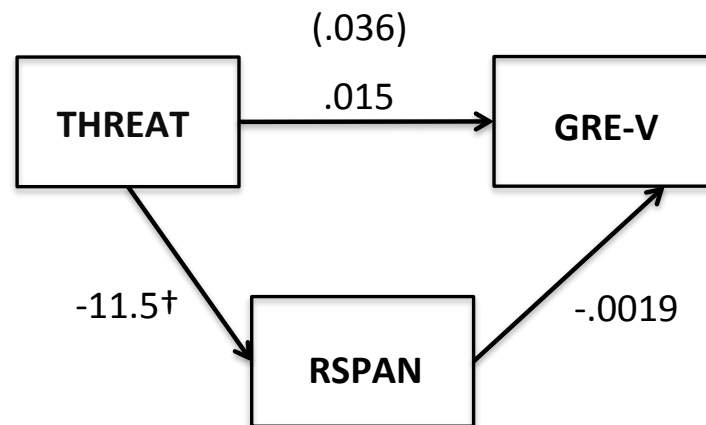


Figure 3.5. Model of WMC Mediating Threat Effect on GRE-V (Towson).

Using nonparametric bootstrapping methods to resample 1000 simulations, results indicated non-significant effects for the average causal mediation effect, parameter estimate = .021, $CI [-.0098, .067]$, $p = .21$, as well as the average direct effect of threat on GRE-V, parameter estimate = .015, $CI [-.055, .095]$, $p = .74$. Taken together, these findings reveal a lack

of evidence that Black subjects' WMC (as assessed via the RSPAN) mediated the effect of threat on GRE-V. (see Figure 3.6 below).

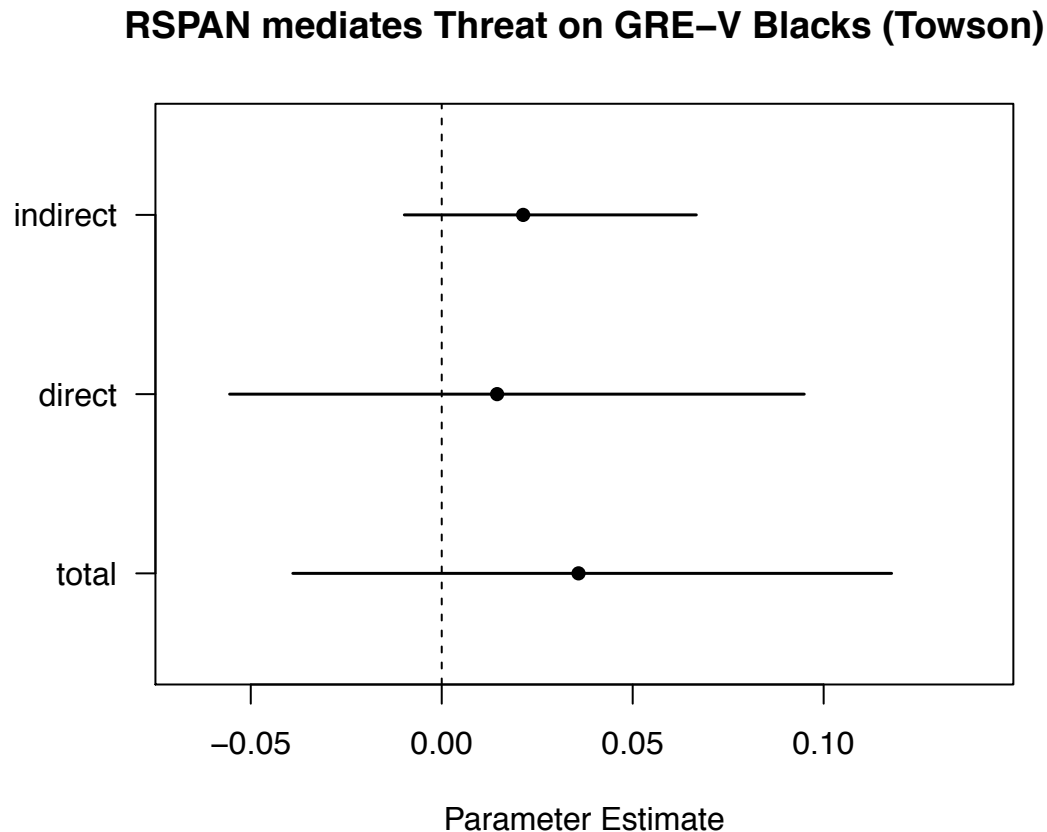


Figure 3.6. WMC Mediation Model for Threat on GRE-V using Bootstrapping (Towson).

A mediation analysis for only Black subjects with condition as the independent variable, RSPAN, the mediator variable, and GRE-M as the outcome variable was conducted. The direct effect of threat on GRE-M was non-significant, $t(28) = 1.34$, $b = .052$, $p > .05$ indicating that the threat had a predicted .052 increase on the GRE-M for black students. The effect of threat on RSPAN was trending toward significant, $t(28) = -1.62$, $b = -9.9$, $p = .12$, indicating that the threat effect predicted a -9.9 unit decrease on RPSAN. The indirect effect of Threat on GRE-M when

RSPAN was added to the model was marginally significant, $t(27) = 1.97, b = .075, p = .06$ indicating an increase in the coefficient for threat (from $b = .052$, to $b = .075$) was observed. In the indirect path, the effect of RSPAN on GRE-M was significant $t(27) = 2.1, b = .0024, p < .05$. Moreover, *Sobel's* test further supported the lack of strong evidence for statistical mediation of RSPAN in the effect of threat on GRE-V, $Z = -1.28, p = .06$ (see Figure 3.7).

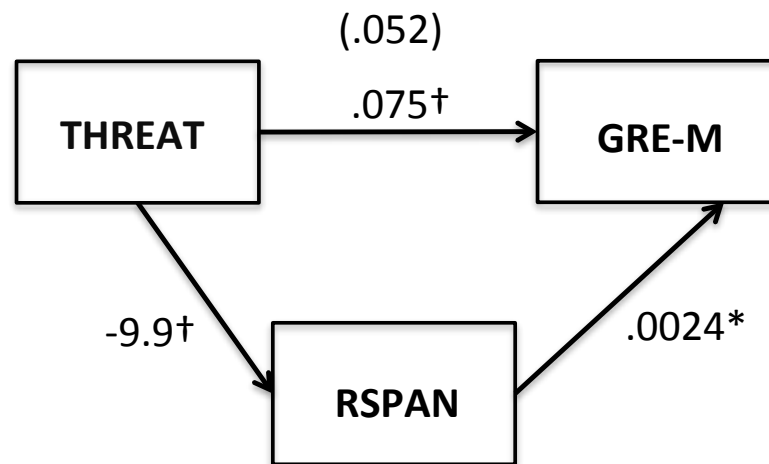


Figure 3.7. Model of WMC Mediating Threat Effect on GRE-M (Towson).

Using nonparametric bootstrapping methods to resample 1000 simulations, results indicated a marginally significant effect for the average causal mediation effect, parameter estimate = $-.023, CI [-.064, .0031], p = .10$. The average direct effect of threat on GRE-M was significant, parameter estimate = $.075, CI [-.0051, .145], p = .04$, suggesting that the effect of threat on GRE-M here predicted a small increase (not a performance decrement). Taken together, these findings reveal a lack of clear evidence that Black subjects' WMC (as assessed via the RSPAN) mediated the effect of threat on GRE-M. Moreover, if anything the effect of threat appeared to go in the opposite direction of the prediction (see Figure 3.8 below).

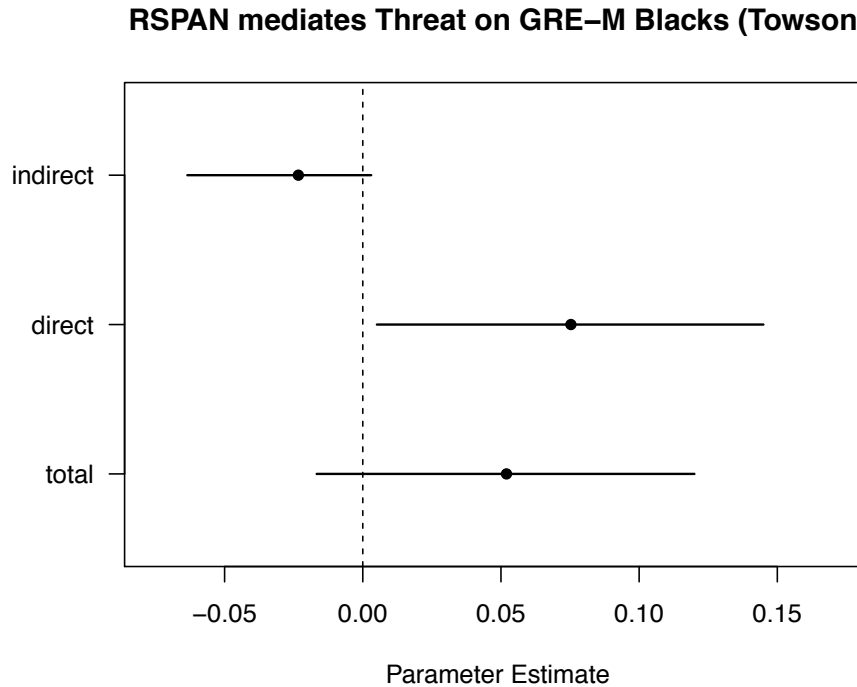


Figure 3.8. WMC Mediation Model for Threat on GRE-M using Bootstrapping (Towson).

Moderation.

To test whether WMC moderates the effect of stereotype threat on standardized test performance, a hierarchical regression was conducted for Black subjects with the outcomes of math and verbal GRE. In the first step, the main effects of OSPAN (mean centered) and condition were investigated. In the second step, the product of OSPAN (mean centered) and condition (dummy coded with control = 0, threat = 1) was added in order to test for moderation.

In step one, looking at the verbal GRE the effects of condition, $b = .034$, $t(27) = .82$, $p > .05$ and OSPAN, $b = -.00011$, $t(27) = -.087$, $p > .05$, and the model, $F(2, 27) = .34$, Multiple $R^2 = .024$, $p > .05$ were non-significant.

In step two predicting verbal GRE, the effect of OSPAN, $b = -.0015$, $t(26) = -.55$, $p > .05$ condition, $b = .035$, $t(26) = .84$, $p > .05$, the interaction, $b = .0018$, $t(26) = .57$, $p > .05$, and the model, $F(3, 26) = .33$, Multiple $R^2 = .037$, $p > .05$ were not significant. Taken together, these results reveal a lack of support for trait WMC on the OSPAN moderating the effect of threat on verbal GRE (see Figure 3.9).

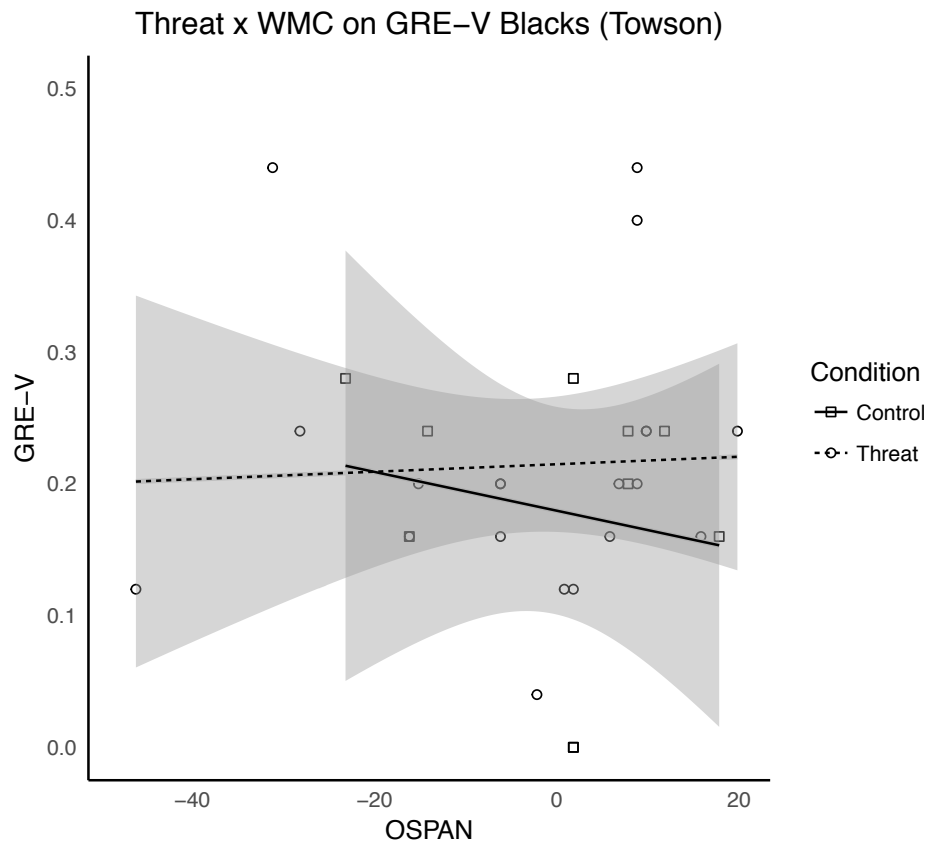


Figure 3.9. OSPAN Moderation Model for Threat on GRE-V

In step one looking at the effects on the math GRE, the effect of condition, $b = .059$, $t(27) = 1.6$, $p > .05$ was non-significant, the effect of OSPAN, $b = .0024$, $t(27) = 2.18$, $p < .05$ and the model, $F(2, 27) = 3.4$, Multiple $R^2 = .201$, $p < .05$ were significant. These results indicate that higher scores on the OSPAN predicted higher scores on the math GRE.

In step two predicting math GRE, the effect of OSPAN, $b = -.0018$, $t(26) = -.79$, $p > .05$ was non-significant. The effect of condition, $b = .064$, $t(26) = 1.9$, $p = .08$ approached significance, indicating that the threat effect predicted higher scores on the math GRE. The interaction, $b = .0054$, $t(26) = 2.14$, $p < .05$, and the model, $F(3, 26) = 4.10$, Multiple $R^2 = .32$, $p < .05$ were significant². Unpacking the interaction, simple slopes tests indicated that under threat, higher WMC participants had higher predicted scores on the verbal GRE compared to those with lower WMC, $b = .0036$, $p < .01$. There was no significant change in predicted GRE scores in the control condition, $b = -.0018$, $p > .05$ (see Figure 3.10).

Based on the significant interaction term an ANOVA was conducted in order to determine whether there was a statistically significant difference in the variance accounted for between these models. Results revealed a statistically significant difference in the variance, $F(1, 26) = 4.6$, $F(1, 26) = 4.62$, $p < .05$ providing additional support for a moderation.

² For these results one subject scored perfectly on the math GRE so this subject was removed and the analyses were re-run.

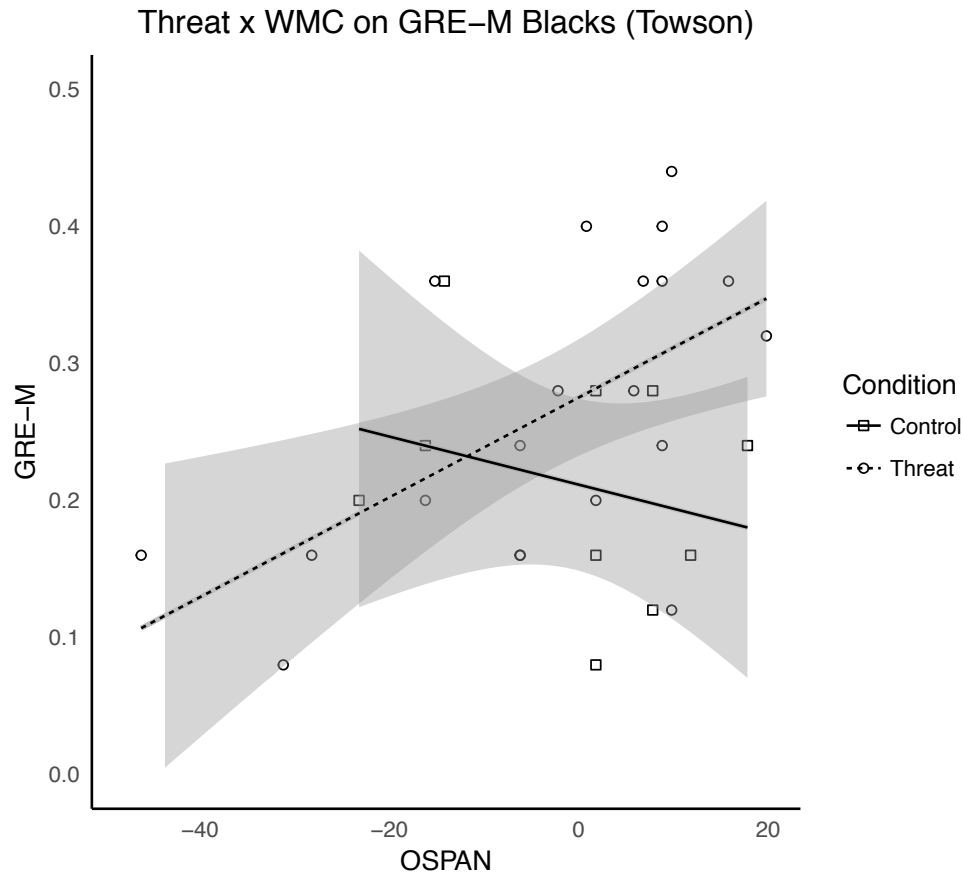


Figure 3.10. OSPAN Moderation Model for Threat on GRE-M

Quantitative Capacity & Intelligence Condition.

Baseline Effects.

Here, the initial t-test was conducted to detect differences in baseline WMC, with race as the (quasi) independent variable and RSPAN as the dependent variable. Baseline WMC on the RSPAN was non-significantly different based on race, $t(79) = -1.17, p > .05$, Cohen's $d = .26$, indicating that although Whites scored about 4 points higher on average than Black subjects this was not a significant difference (see Figure 3.11).

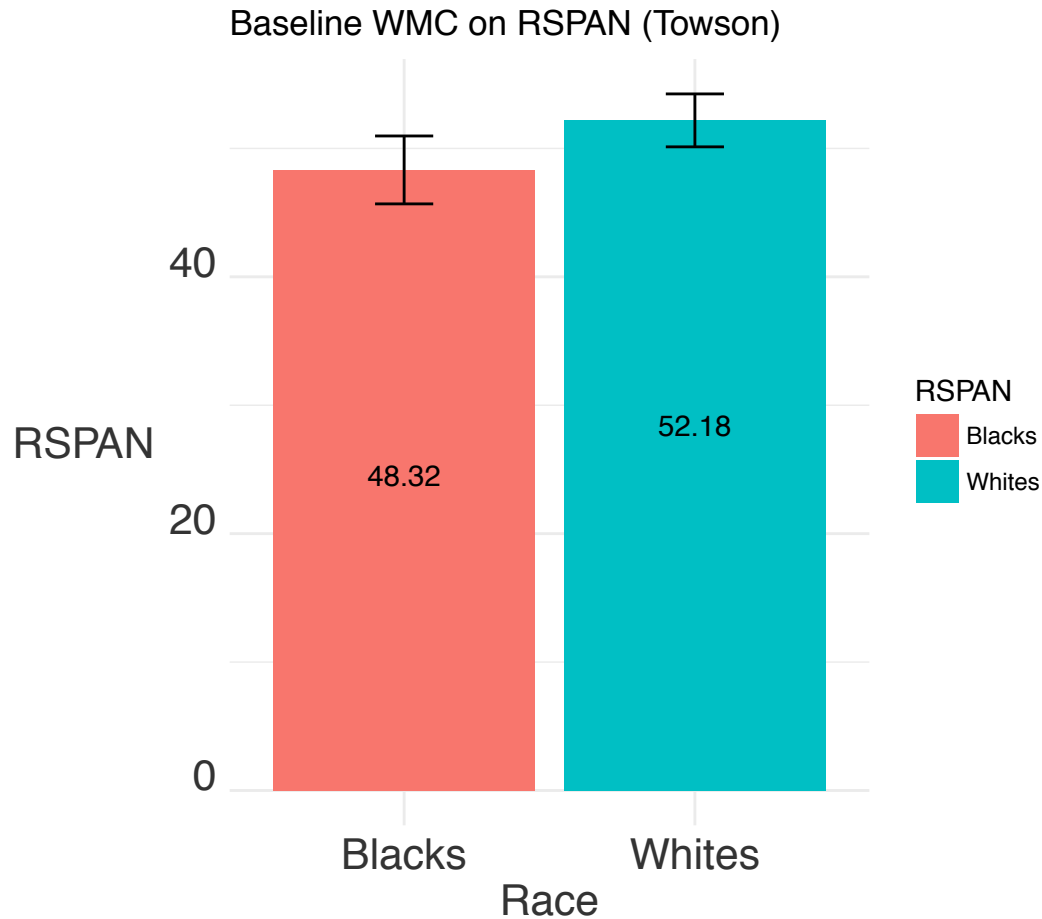


Figure 3.11. Baseline WMC on RSPAN (Towson).

Threat Effects.

Looking at the effect of stereotype threat, a 2 (race) x 2 (condition) ANOVA with OSPAN as the dependent measure revealed non-significant effects of condition, $F(1, 77) = .37$, $p > .05$, $\eta_{\text{partial}}^2 = .0048$, race, $F(1, 77) = 1.03$, $p > .05$, $\eta_{\text{partial}}^2 = .011$ and their interaction, $F(1, 77) = 1.44$, $p > .05$, $\eta_{\text{partial}}^2 = .018$ (see Figure 3.12).

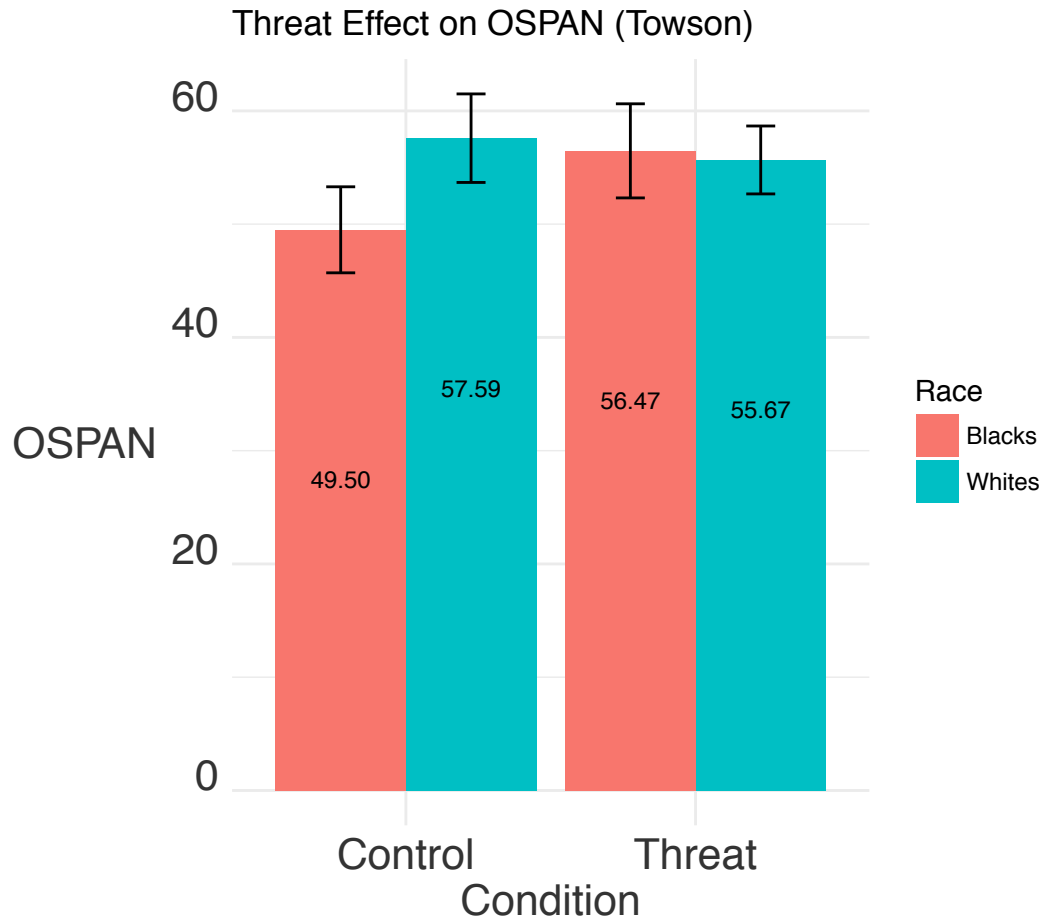


Figure 3.12. Threat Effect on OSPAN (Towson)

In the second 2 (race) x 2 (condition) ANOVA, GRE-M was the dependent measure. Results revealed a main effect of race, $F(1, 77) = 19.4, p < .001, \eta_{\text{partial}}^2 = .193$, indicating that White subjects ($M = .31, SD = .13$) significantly outperformed Blacks ($M = .21, SD = .08$) on the math GRE. There were no significant condition, $F(1, 77) = .11, p > .05, \eta_{\text{partial}}^2 = .0015$ or interaction effects, $F(1, 77) = 2.009, p > .05, \eta_{\text{partial}}^2 = .025$ indicating that there was no evidence of a threat effect (see Figure 3.13).

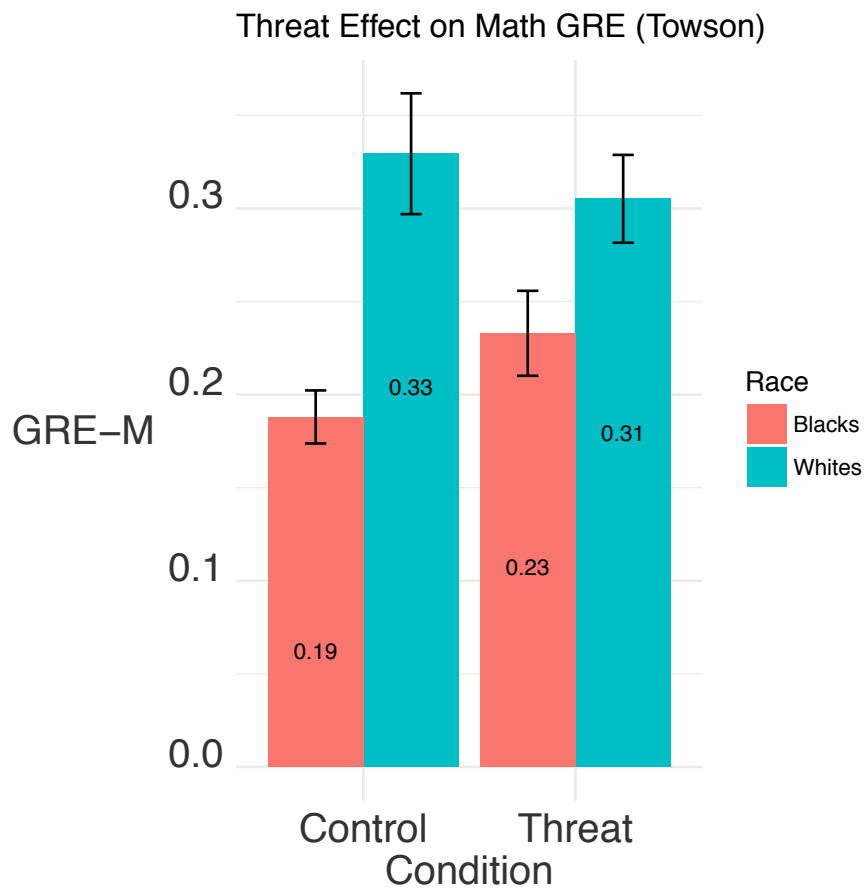


Figure 3.13. Threat Effect on GRE-M (Towson)

In the third 2 (race) x 2 (condition) ANOVA, GRE-V was the dependent measure. Results indicated a significant effect of race such that Whites performed significantly higher ($M = .30, SD = .17$) on the verbal GRE compared to Blacks ($M = .22, SD = .10$), $F(1, 77) = 7.1, p < .01, \eta_{\text{partial}}^2 = .098$. The effect of condition was marginally significant, $F(1, 77) = 2.90, p = .09, \eta_{\text{partial}}^2 = .036$ indicating a trend such that performance was lower in the threat condition ($M = .25, SD = .14$) compared to control ($M = .29, SD = .15$). The interaction was marginally significant, $F(1, 77) = 2.51, p = .11, \eta_{\text{partial}}^2 = .032$. *Tukey's* pairwise comparisons indicated that Blacks' performance was not significantly different under threat compared to the control (M diff

= -.00047, $p = .99$). Whites' performance tended to be lower under threat compared to control (M diff = -.098, $p = .10$; see Figure 3.14).

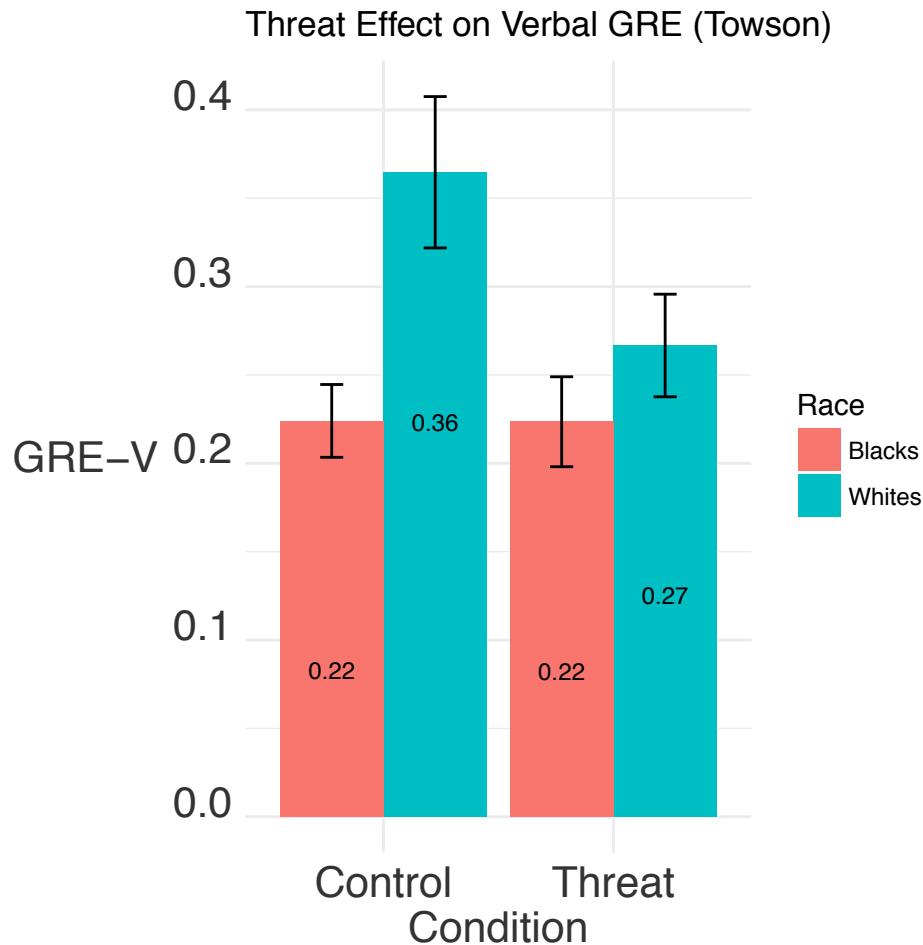


Figure 3.14. Threat Effect on GRE-V (Towson)

Mediation.

A mediation analysis for only Black subjects with condition as the independent variable, OSPAN, the mediator variable, and GRE-M as the outcome variable was conducted. The direct effect of threat on GRE-M was non-significant, $t(35) = 1.7$, $b = .045$, $p = .09$ indicating that the threat condition had a predicted .045 increase on the GRE-M for Black students. The effect of

threat on OSPAN was not significant, $t(35) = 1.2$, $b = 6.97$, $p > .05$, indicating that the threat effect predicted a 6.97 increase on OPSAN. The indirect effect of Threat on GRE-M when OSPAN was added to the model was trending toward significant, $t(34) = 1.61$, $b = .044$, $p = .12$. In the indirect path, the effect of OSPAN on GRE-M was also non-significant $t(34) = .26$, $b = .00021$, $p > .05$. *Sobel's* test further supported the lack of evidence for statistical mediation of OSPAN in the effect of threat on GRE-M, $Z = .254$, $p = .12$ (see Figure 3.15).

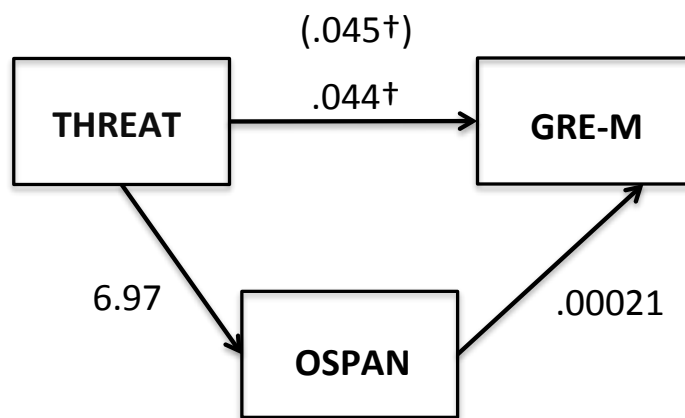


Figure 3.15. Model of WMC Mediating Threat Effect on GRE-M (Towson).

Bootstrapping methods resampled 1000 simulations to further test for mediation. Results indicated non-significant effects for the average causal mediation effect, parameter estimate = .0014, $CI [-.0148, .0215]$, $p = .81$, as well as the average direct effect of threat on GRE-M, parameter estimate = .044, $CI [-.010, .102]$, $p = .11$. Taken together, these findings reveal a lack of evidence that Black subjects' WMC (as assessed via the OSPAN) mediated the effect of threat on GRE-M. (see Figure 3.16).

OSPAN mediates Threat on GRE-M Blacks (Towson)

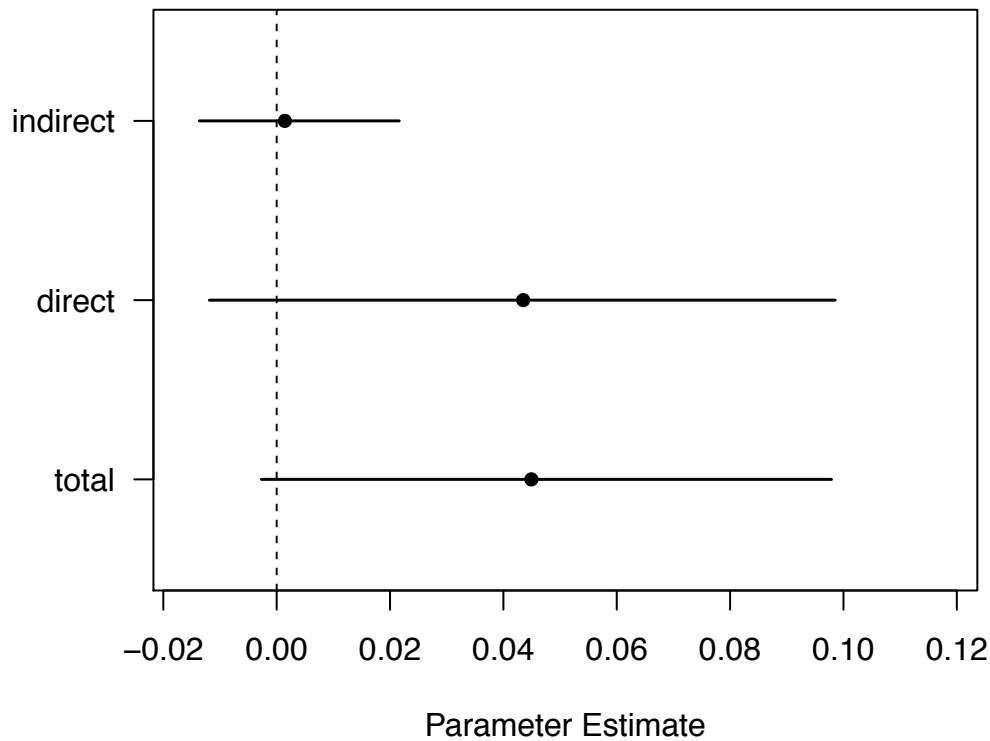


Figure 3.16. WMC Mediation Model for Threat on GRE-M using Bootstrapping (Towson).

A mediation analysis for only Black subjects with condition as the independent variable, OSPAN, the mediator variable, and GRE-V as the outcome variable was conducted. The direct effect of threat on GRE-V was non-significant, $t(35) = -.015$, $b = -.00047$, $p > .05$. The effect of threat on OSPAN was not significant, $t(35) = 6.97$, $b = 1.24$, $p > .05$. The indirect effect of Threat on GRE-V when OSPAN was added to the model was not significant, $t(34) = -.06$, $b = -.0023$, $p > .05$. In the indirect path, the effect of OSPAN on GRE-V was also not significant $t(34) = .27$, $b = .00026$, $p > .05$. *Sobel's* test further supported the lack of evidence for statistical mediation of OSPAN in the effect of threat on GRE-V, $Z = .26$, $p = .95$ (see Figure 3.17).

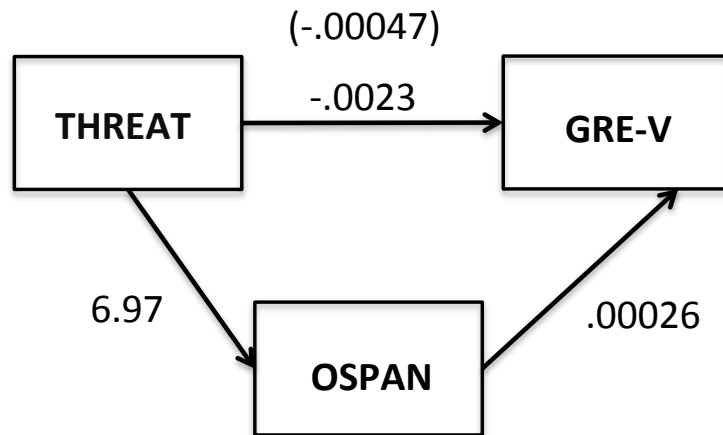


Figure 3.17. Model of WMC Mediating Threat Effect on GRE-V (Towson).

Bootstrapping methods resampled 1000 simulations to further test for mediation. Results indicated non-significant effects for the average causal mediation effect, parameter estimate = $-.0068$, $CI [-.037, .023]$, $p = .63$, as well as the average direct effect of threat on GRE-V, parameter estimate = $.0696$, $CI [-.030, .163]$, $p = .15$. Taken together, these findings reveal a lack of evidence that Black subjects' WMC (as assessed via the OSPAN) mediated the effect of threat on GRE-V (see Figure 3.18)

OSPAN mediates Threat on GRE-V Blacks (Princeton)

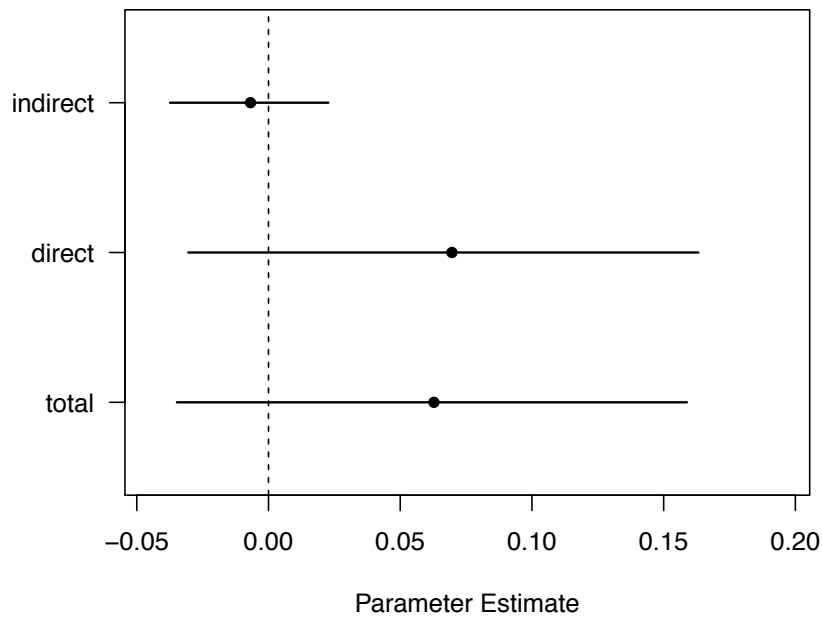


Figure 3.18. WMC Mediation Model for Threat on GRE-V using Bootstrapping (Towson).

Moderation.

To test whether WMC moderates the effect of stereotype threat on standardized test performance, a hierarchical regression was conducted for Black subjects with the outcomes math and verbal GRE. In both models the first step, the main effects of RSPAN (mean centered) and condition were investigated. In the second step, the product of RSPAN (mean centered) and condition (dummy coded with control = 0, threat = 1) was added in order to test for moderation.

Looking at math GRE in step one, the effects of RSPAN, $b = .000071$, $t(34) = -.085$, $p > .05$ and the model, $F(2, 34) = 1.45$, Multiple $R^2 = .078$, $p > .05$ were non-significant. The effect of condition was marginally significant, $b = .045$, $t(34) = 1.7$, $p = .09$.

In step two, the effects of RSPAN, $b = .0034$, $t(33) = .29$, $p > .05$, condition, $b = .043$, $t(33) = 1.59$, $p > .05$, the interaction, $b = -.00086$, $t(33) = -.51$, $p > .05$ and the model, $F(3, 33) =$

1.03, Multiple $R^2 = .086$, $p > .05$ were non-significant. These findings indicate that under threat higher trait WMC did not provide a boost compared to lower trait WMC on the math GRE (see Figure 3.19).

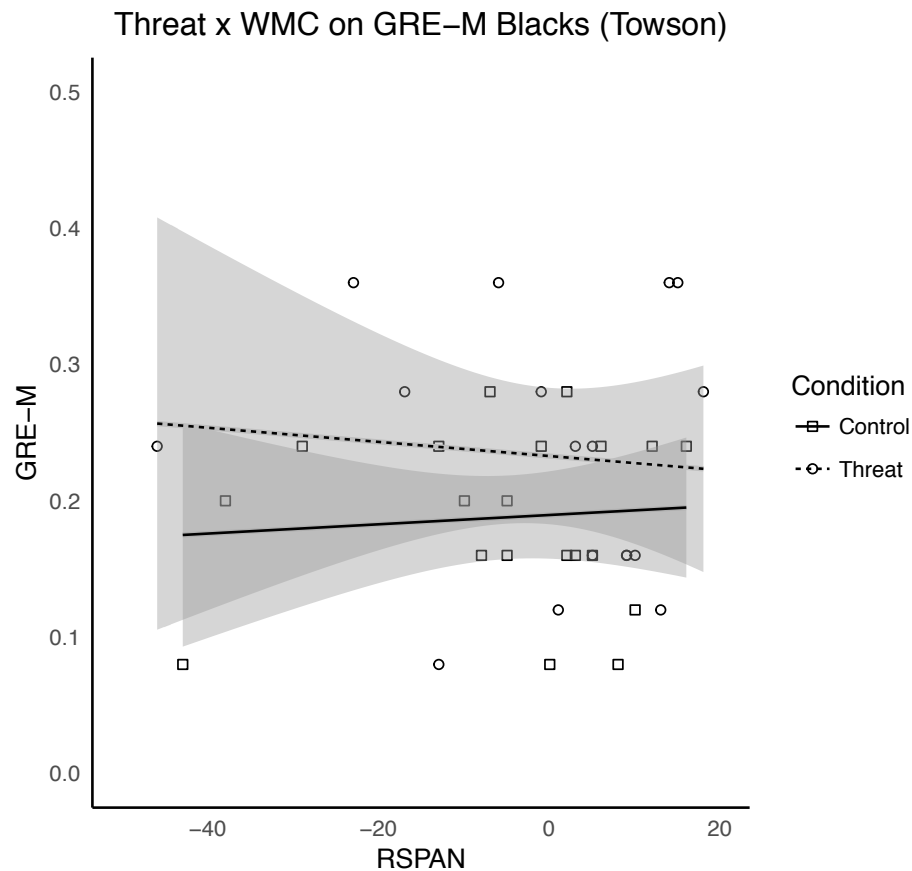


Figure 3.19. RSPAN Moderation Model for Threat on GRE-M

In the model for verbal GRE in step one, the effects of RSPAN, $b = .00083$, $t(34) = .81$, $p > .05$, condition, $b = -.0042$, $t(34) = -.13$, $p > .05$ and the model, $F(2, 34) = .326$, Multiple $R^2 = .0188$, $p > .05$ were non-significant.

In step two, the effects of RSPAN, $b = -.0011$, $t(33) = -.83$, $p > .05$, condition, $b = .0052$, $t(33) = .165$, $p > .05$, and the model, $F(3, 33) = 1.68$, Multiple $R^2 = .132$, $p > .05$ were non-significant. The interaction of RSPAN and condition however, was significant, $b = .0041$, $t(33) =$

2.07, $p < .05$. Simple slopes tests indicated that under threat, higher WMC participants had higher predicted scores on the verbal GRE compared to those with lower WMC, $b = .0030$, $p < .05$. There was no significant change in predicted GRE scores in the control condition, $b = -.0011$, $p > .05$.

Next, based on the significant interaction the models in steps one and two were tested for a significant change in their variances. Analysis of change in model variances indicated there was a significant difference, $F(1, 33) = 4.3$, $p < .05$, providing further evidence for a moderation effect of trait WMC on threat effects for Black students' verbal GRE scores (see Figure 3.20).

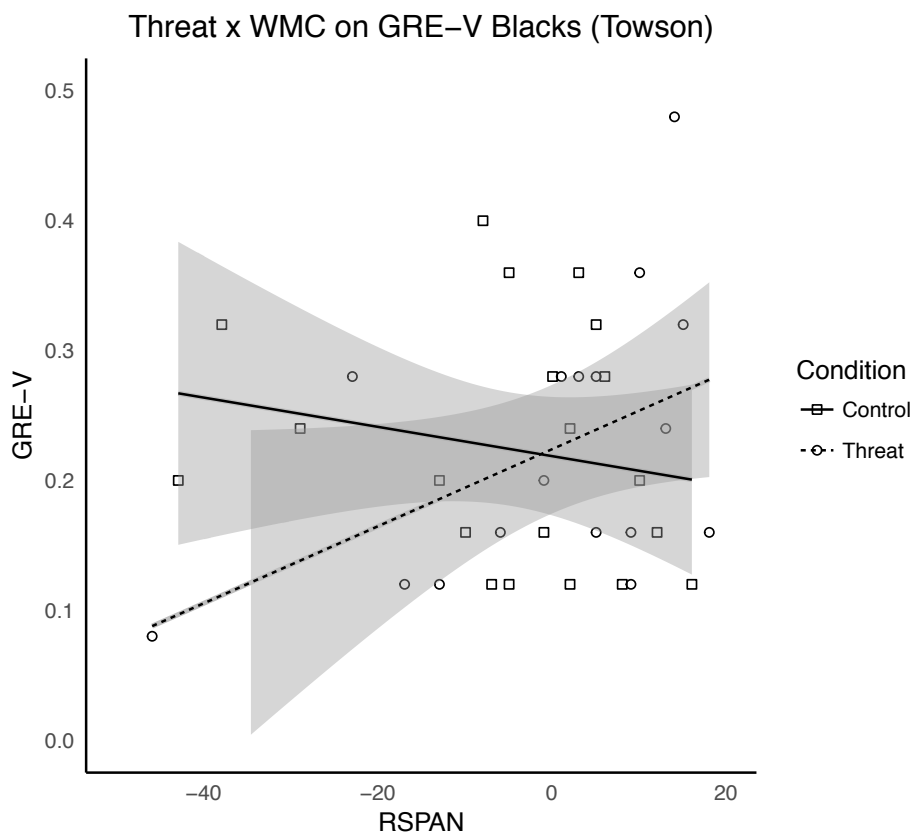


Figure 3.20. RSPAN Moderation Model for Threat on GRE-V

Threat Effects & Non-Cognitive Factors.

Data Preparation & Analytic Approach.

Next, the role of mindset was explored as a possible important non-cognitive factor during contexts of stereotype threat. Mindset measures are typically split based on categories of more fixed or malleable mindsets, but due to participant attrition, the mindset measure was kept as a continuous scale in an effort to preserve statistical power. Other non-cognitive measures such as grit, school belongingness and conscientiousness were also measured. The full report of summary statistics for all measures are presented in Table 3.2³. For the full analyses with the grit and conscientiousness measures, the reader should consult the supplemental information section.

Summary Statistics & Correlations.

There were several interesting trends in the non-cognitive factors data. Focusing on the variables of greatest importance, general trends revealed higher scores on the mindset measure (which corresponds to more malleable mindsets) were strongly positively related to trait conscientiousness—indicating that more malleable mindsets are directly related to higher conscientiousness. Mindset was moderately positively related to grit scores suggesting that having a more malleable mindset was moderately related to higher grit scores. In addition, higher conscientiousness was moderately positively related to grit scores revealing that the more conscientious participants were higher on trait grittiness as well (see Table 3.2).

³ Due to participants failing to complete the non-cognitive surveys which were linked on Qualtrics, the maximum sample size for complete cases of survey data was $n = 65$. All subsequent analyses with the non-cognitive measures will employ this subset of the Towson data.

Table 3.2. *Summary Statistics and Correlations of Non-Cognitive Factors in Experiment 2*

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. ExtraversionScale	--												
2. AgreeablenessScale	0.114	--											
3. ConscientiousnessScale	-0.087	.397***	--										
4. NeuroticismScale	0.008	-.319**	-0.239	--									
5. OpennessScale	.398***	0.239	.402***	-0.074	--								
6. MindsetScore	0.223	.431***	.829***	-0.210	.850***	--							
7. VoiceHeard	-0.070	0.052	0.127	-.349***	0.044	0.115	--						
8. ComfortExpressView	0.084	0.082	0.132	-.335**	0.126	0.176	.513***	--					
9. MyOpinionsMatter	-0.193	0.026	0.210	-0.173	0.017	0.151	.486***	.514***	--				
10. IdentifyRelateTowson	0.118	0.204	0.040	-0.100	-0.060	-0.010	.362***	0.215	.410***	--			
11. IdentifyRelateTUSudents	0.039	0.133	0.000	0.108	0.040	0.020	.252*	.265*	.417***	.650***	--		
12. MyOpinionsMatterUD	-0.107	0.068	.240*	-0.051	0.160	.240*	.297*	.258*	.517***	.270*	.330**	--	
13. GritScore	0.115	.420***	.570***	-0.218	.520***	.650***	.462***	.569***	.604***	.550***	.600***	.660***	--
<i>M</i>	3.18	3.99	3.49	3.21	3.4	3.45	3.77	4.28	4.8	4.43	4.48	3.77	3.89
<i>SD</i>	0.73	0.66	0.68	0.85	0.71	0.58	1.05	1.21	1.21	1.26	1.25	1.47	0.61

Note. ExtraversionScale = Trait Extravertedness, AgreeablenessScale = Trait Agreeableness, ConscientiousnessScale = Trait Conscientiousness, NeuroticismScale = Trait Neuroticism, OpennessScale = Trait Openness, MindsetScore = Malleable Mindset Score, VoiceHeard = Perceived Score Voice Heard at Towson University, ComfortExpressViews = Perceived Comfort Expressing Views at Towson University, MyOpinionsMatter = Perceived Score My Opinions Matter at Towson University, IdentifyRelateTowson = Perceived Score Identify and Relate with Towson University, IdentifyRelateTUSudents = Perceived Score Identify and Relate with Towson University Students, MyOpinionsMatterUD = Perceived Score My Opinions Matter in University Decisions at Towson University, GritScore = Trait Grittiness Score.

*** $p < .001$, ** $p < .01$, * $p < .05$.

Mindset Regression Analyses.

A series of regression analyses were conducted to assess whether individual differences in mindset moderate student performance under conditions of stereotype threat. Stereotype threat effects were expected to only negatively impact Black student performance. However due to the low statistical power in these data both White and Black subjects were included in the moderation analyses with non-cognitive factors. This change in analytic approach as compared to the approach in Experiment 1 was implemented in order to run more precise statistical analyses. As such, this change in approach by including White subjects in the moderation analyses was less theoretically motivated and more so driven by practical reasons.

Thus, in each of the regression analyses below the hypothesis that as a trait non-cognitive factor, mindset moderates the effect of stereotype threat on the outcomes of WMC and

standardized test performance. The outcome measures were regressed on race (“White”: 1, “Black ”: –1) condition (“Control”: 1, “Threat ”: –1), trait mindset (mean-centered), and their interactions. Next, the moderation results of the trait mindset scores are presented across the conditions of verbal capacity and intelligence and quantitative capacity and intelligence.

Verbal Capacity & Intelligence Condition.

Next, a series of hierarchical regression analyses were conducted in order to test whether mindset moderated stereotype threat on WMC (via RSPAN). Only the main effect of mindset was trending toward significant in the first model such that subjects with more malleable mindsets tended to have higher predicted scores on the RSPAN. None of the remaining main effects or their interactions were significant (see table 3.3 and Figure 3.21 below). The moderation model was also not significant, $F(7, 26) = .796$, Multiple $R^2 = .17$, $p > .05$.

Table 3.3. *Mindset Moderating Stereotype Threat on RSPAN*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	-1.434	2.802	-0.512	-0.092	0.613
condition	-1.347	2.480	-0.543	-0.097	0.591
mindset	13.955	7.041	1.982†	0.352	0.057
Model 2					
race	-3.738	3.615	-1.034	-0.240	0.311
condition	0.937	3.615	0.259	0.067	0.798
mindset	-2.542	26.098	-0.097	-0.064	0.923
race x condition	-4.262	3.615	-1.179	-0.308	0.249
condition x mindset	-11.125	26.098	-0.426	-0.286	0.673
race x mindset	19.060	26.098	0.730	0.470	0.472
race x condition x mindset	17.273	26.098	0.662	0.448	0.514

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

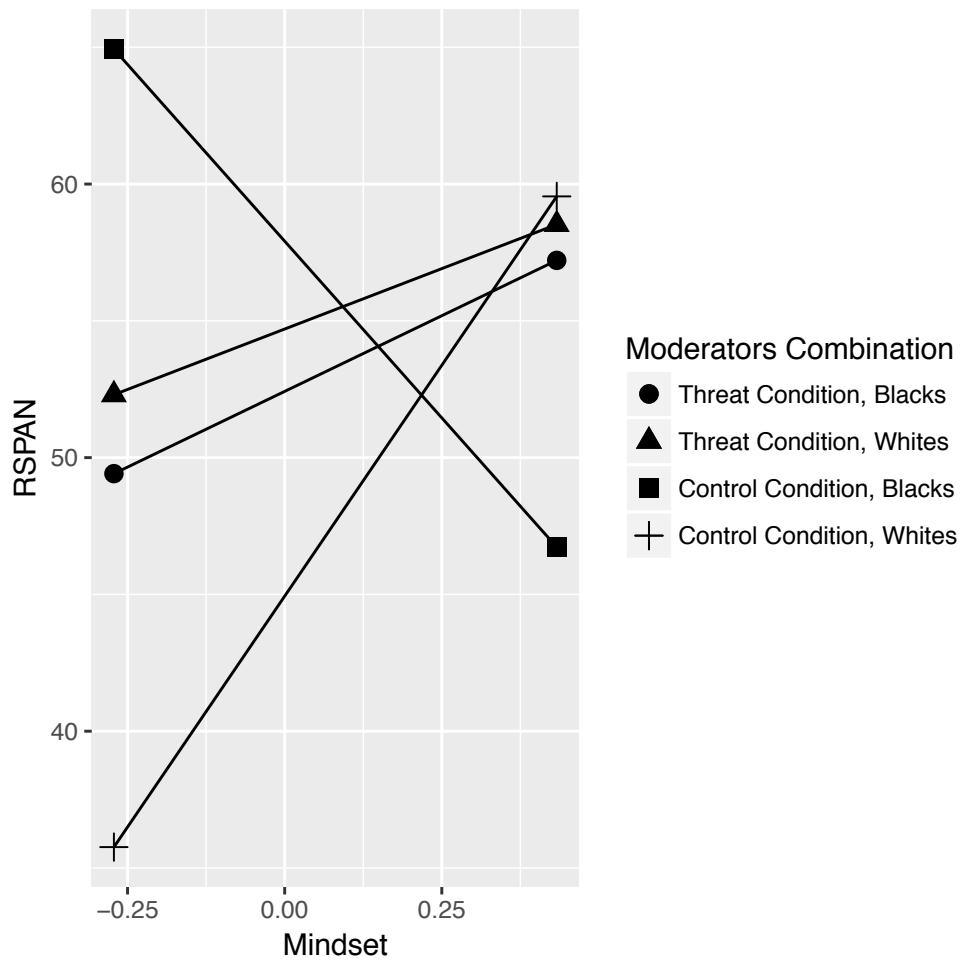


Figure 3.21. Simple Slopes of Mindset moderating Stereotype Threat on RSPAN.

In the second analysis, the effects of mindset, race and condition along with their interactions were tested for their potential moderating effect on verbal standardized test performance. Results revealed no significant main effects or interactions. (see Table 3.4 and Figure 3.22 below). The moderation model was also not significant, $F(7, 26) = .559$, Multiple $R^2 = .13$, $p > .05$.

Table 3.4. *Mindset Moderating Stereotype Threat on GRE-V*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.057	0.039	1.471	0.271	0.152
condition	-0.007	0.034	-0.192	-0.035	0.849
mindset	0.029	0.097	0.296	0.054	0.769
Model 2					
race	0.052	0.050	1.032	0.246	0.311
condition	0.002	0.050	0.041	0.011	0.968
mindset	-0.309	0.362	-0.854	-0.578	0.401
race x condition	0.002	0.050	0.044	0.012	0.965
condition x mindset	-0.271	0.362	-0.749	-0.516	0.461
race x mindset	0.367	0.362	1.013	0.670	0.321
race x condition x mindset	0.253	0.362	0.700	0.486	0.490

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

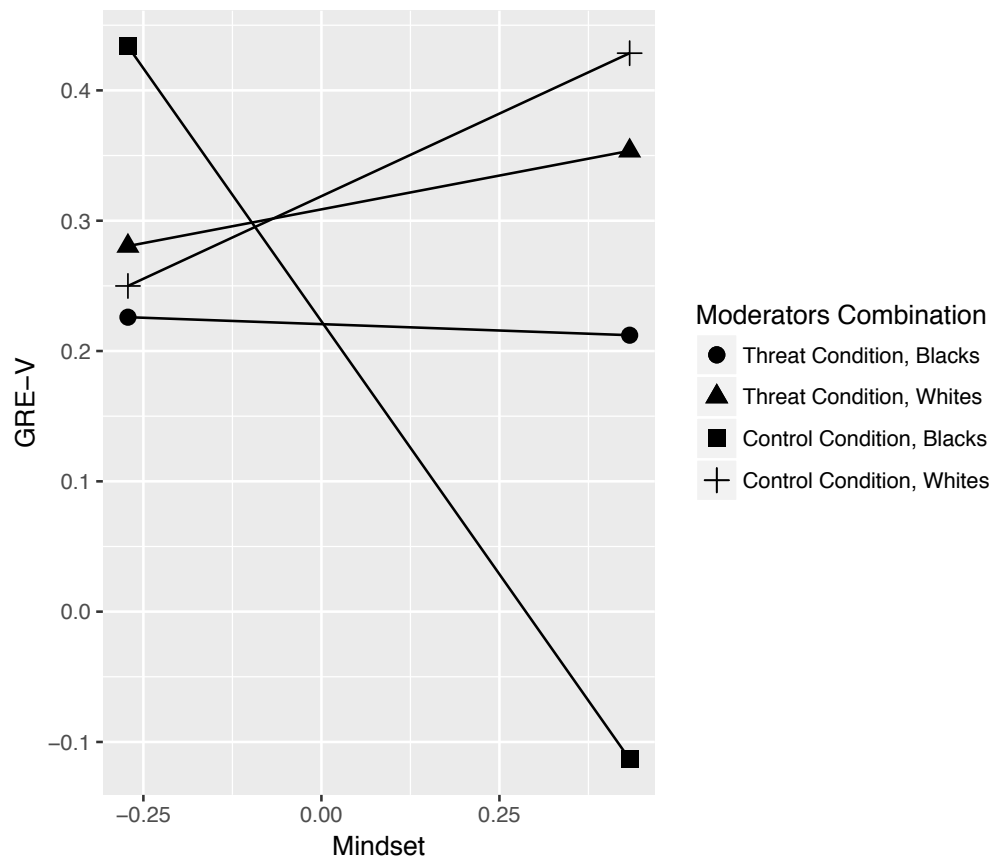


Figure 3.22. Simple Slopes of Mindset moderating Stereotype Threat on GRE-V.

In the next analysis, the effects of mindset, race and condition along with their interactions were tested for their potential moderating effect on quantitative standardized test performance. Results revealed no significant main effects or interactions. (see Table 3.5 and Figure 3.23 below). The moderation model was also not significant, $F(7, 26) = .613$, Multiple $R^2 = .14$, $p > .05$.

Table 3.5. *Mindset Moderating Stereotype Threat on GRE-M*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.019	0.028	0.658	0.124	0.516
condition	0.015	0.025	0.602	0.112	0.552
mindset	-0.030	0.071	-0.424	-0.079	0.675
Model 2					
race	0.046	0.035	1.291	0.306	0.208
condition	-0.002	0.035	-0.070	-0.019	0.945
mindset	0.101	0.256	0.395	0.266	0.696
race x condition	0.044	0.035	1.244	0.332	0.224
condition x mindset	0.119	0.256	0.464	0.318	0.647
race x mindset	-0.151	0.256	-0.590	-0.387	0.561
race x condition x mindset	-0.229	0.256	-0.896	-0.619	0.379

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

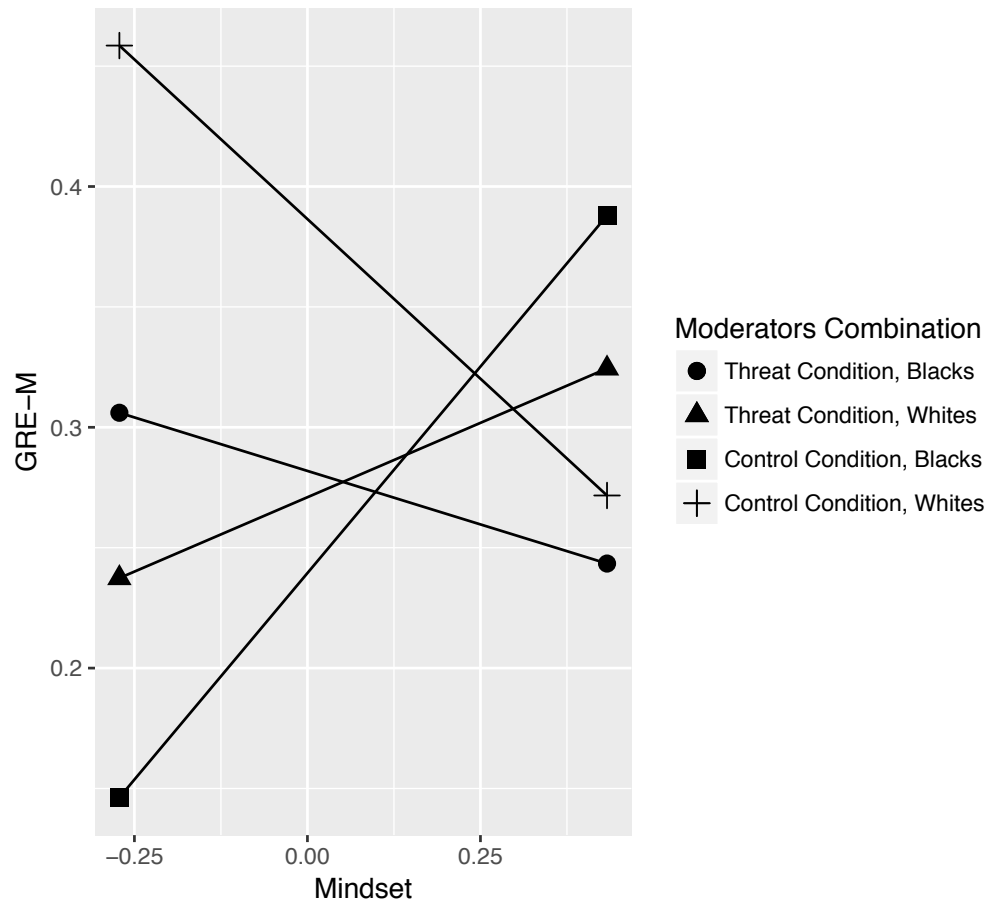


Figure 3.23. Simple Slopes of Mindset moderating Stereotype Threat on GRE-M.

Quantitative Capacity & Intelligence Condition.

Looking at the mindset moderation model predicting WMC on the OSPAN, neither the main effects of race, condition, mindset, nor their interactions were significant (see Table 3.6 and Figure 3.24). The moderation model was also not significant, $F(7, 23) = .918$, Multiple $R^2 = .21$, $p > .05$.

Table 3.6. *Mindset, Threat and Race Effects on OSPAN Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.455	3.561	0.128	0.025	0.899
condition	-2.030	3.637	-0.558	-0.112	0.581
mindset	4.190	4.878	0.859	0.171	0.398
Model 2					
race	2.003	3.682	0.544	0.109	0.592
condition	-1.800	3.682	-0.489	-0.099	0.630
mindset	5.929	8.096	0.732	0.242	0.471
race x condition	3.990	3.682	1.084	0.214	0.290
condition x mindset	7.816	8.096	0.965	0.306	0.344
race x mindset	9.476	8.096	1.170	0.384	0.254
race x condition x mindset	4.329	8.096	0.535	0.178	0.598

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

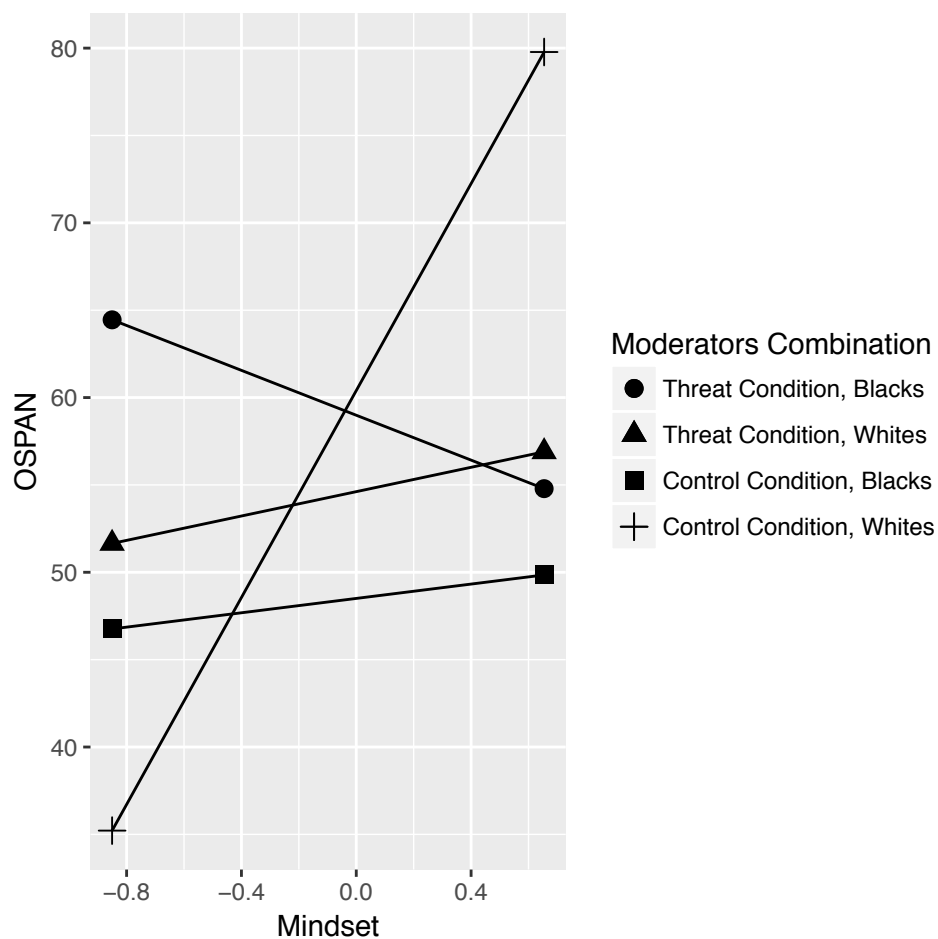


Figure 3.24. Simple Slopes of Mindset moderating Stereotype Threat on OSPAN.

Next, a series of hierarchical regression analyses were conducted in order to test whether mindset moderated stereotype threat on math standardized test performance. Only the main effect of race was significant or trending toward significant in the models below such that White subjects had higher predicted scores on the math GRE (see Table 3.7). None of the remaining main effects or their interactions were significant. (see Figure 3.25). The moderation model was also not significant, $F(7, 23) = 1.62$, Multiple $R^2 = .33$, $p > .05$.

Table 3.7. *Mindset, Threat and Race Effects on GRE-M Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.053	0.024	2.223*	0.402	0.035
condition	-0.002	0.024	-0.092	-0.017	0.928
mindset	0.008	0.033	0.247	0.046	0.807
Model 2					
race	0.047	0.024	1.904†	0.353	0.069
condition	0.002	0.024	0.063	0.012	0.950
mindset	0.062	0.054	1.156	0.353	0.259
race x condition	0.033	0.024	1.370	0.251	0.184
condition x mindset	-0.063	0.054	-1.166	-0.342	0.256
race x mindset	0.065	0.054	1.202	0.365	0.242
race x condition x mindset	-0.044	0.054	-0.827	-0.255	0.417

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

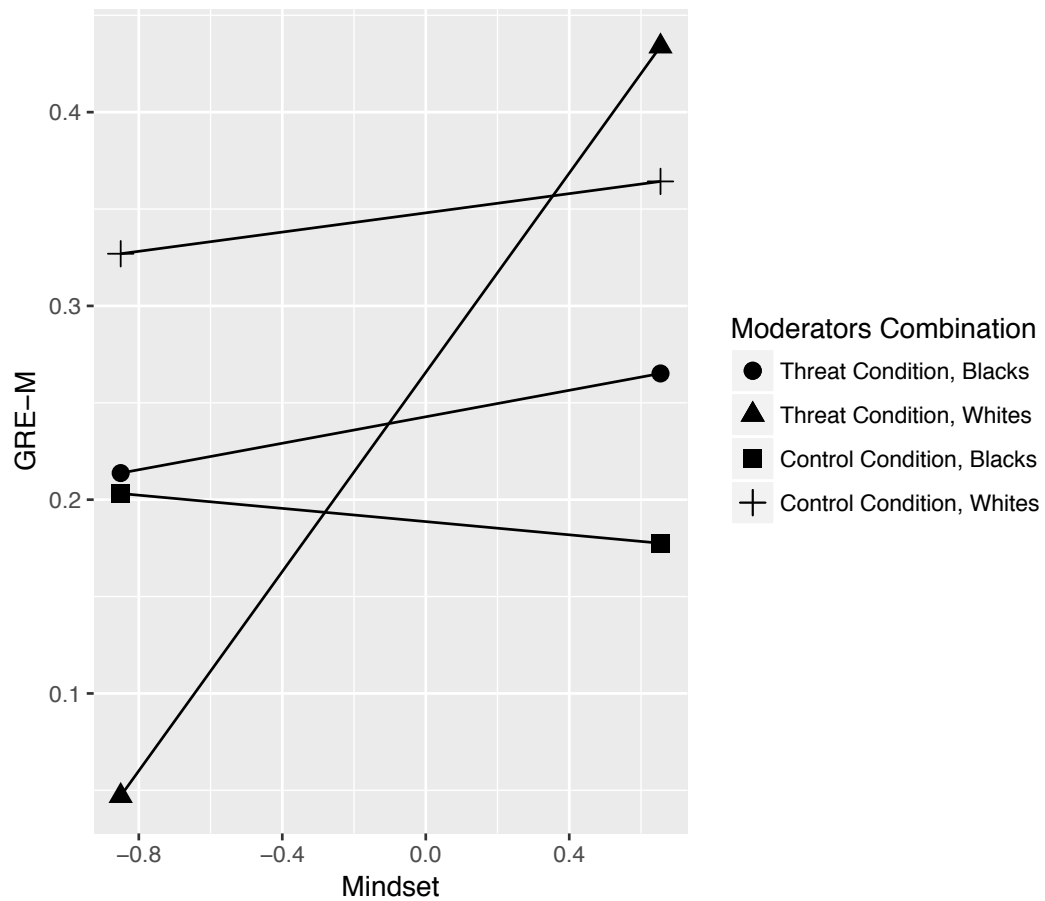


Figure 3.25. Simple Slopes of Mindset moderating Stereotype Threat on GRE-M.

Next, a series of hierarchical regression analyses were conducted in order to test whether mindset moderated stereotype threat on verbal standardized test performance. None of the main effects or their interactions were significant (see Table 3.8 and Figure 3.26). The moderation model was also not significant, $F(7, 23) = .313$, Multiple $R^2 = .087$, $p > .05$.

Table 3.8. Mindset, Threat and Race Effects on GRE-V Full Model Estimates

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.022	0.024	0.939	0.181	0.356
condition	-0.021	0.024	-0.862	-0.172	0.396
mindset	-0.014	0.032	-0.435	-0.086	0.667
Model 2					
race	0.020	0.027	0.768	0.166	0.450
condition	-0.021	0.027	-0.799	-0.175	0.432
mindset	-0.043	0.058	-0.743	-0.265	0.465
race x condition	-0.002	0.027	-0.080	-0.017	0.937
condition x mindset	0.031	0.058	0.532	0.182	0.600
race x mindset	0.023	0.058	0.389	0.138	0.701
race x condition x mindset	-0.029	0.058	-0.502	-0.180	0.621

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

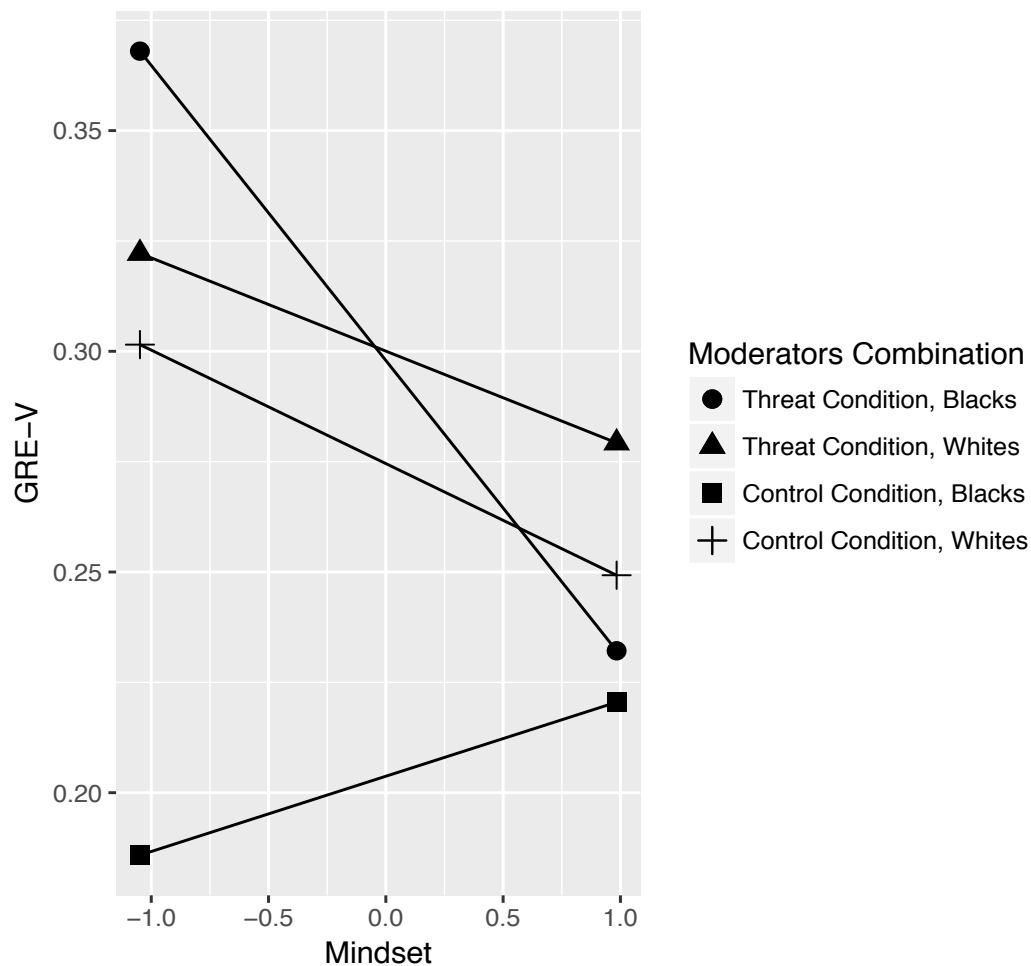


Figure 3.26. Simple Slopes of Mindset moderating Stereotype Threat on GRE-V.

Experiment 2 Summary & Discussion

In line with the results of Experiment 1, there was a lack of strong evidence for stereotype threat effects impacting performance on outcome measures of WMC and standardized test performance. Only in the case of stereotype threat for verbal capacity and intelligence on the RSPAN did these data reveal evidence of a threat effect (such that Black subjects under threat suffered a performance decrease relative to Blacks in the control condition).

In addition, there was weak evidence that stereotype threat effects actually provided students with a small boost (for both White and Black students in some cases). Other than the threat effect observed on the RSPAN, generally, Black subjects tended to receive a small boost on performance under stereotype threat. Overall, students appeared to be resilient to the effects of stereotype threat. Here, it was expected that because this sample comes from a less selective population of undergraduate students that stereotype threat effects would be found in addition to evidence that stereotype threat could be moderated by high trait WMC and mediated by state WMC. Instead, there was no evidence to support the notion that state WMC mediates stereotype threat, as in many cases there was no threat effect revealed. Although, there was a pattern of evidence supporting the hypothesis that trait WMC moderates the effect of stereotype threat on both math and verbal standardized test performance for Black subjects. Moreover, these data revealed that when students were threatened for intelligence and math or verbal capacities, students were able to kick performance up a notch on the second GRE (the GRE that did not match the domain they were directly threatened in). This finding suggests that higher trait WMC protects Black students from the impact of stereotype threat on standardized test performance but the kind of threat and the domain of threat may also matter. It is important to note that the performance gap on the standardized test measures between White and Black students was not

removed, however these data provide evidence that higher WMC may help students with stigmatized identities combat the impact of stereotype threat on important assessments like standardized tests that have the power to determine their admittance to higher education and post-graduate institutions.

In addition to exploring the role of WMC under conditions of stereotype threat, Experiment 2 aimed to uncover whether non-cognitive factors might moderate the effect of stereotype threat on cognitive performance outcomes. Results revealed no *statistically significant* evidence that trait mindset scores moderated stereotype threat effects on performance. Instead, only in the verbal capacity and intelligence condition was a marginally significant trend found such that a more malleable mindset predicted higher scores on the RSPAN (see Table 3.3 and Figure 3.21). However, looking at subsequent performance in this condition on the verbal GRE and math GRE the main effect of mindset was no longer marginally significant and the direction of the data changed such that more fixed mindsets predicted better performance on these measures.

In quantitative capacity and intelligence condition, although non-significant, additional trends in the direction of the data revealed, on the OSPAN, Black students had higher predicted scores under stereotype threat when they had more fixed mindsets. Looking at subsequent performance in this condition on the verbal and math GREs the direction of the data also suggested more malleable mindsets did not predict better performance on these measures. Collectively, these findings reveal that for Black subjects under stereotype threat there was no strong evidence to support the notion that more malleable mindsets predicted higher scores on the WMC or standardized test performance outcomes. An interpretation of these results could be about effort. If you have a more malleable mindset then the theory has suggested that you believe

ability can change and this change probably comes about based on effort, work, and taking on a new challenge. If you have a stigmatized identity it could be that this perspective actually backfires if you feel threatened, tired or overwhelmed. That is, having a more malleable mindset in these situations provides no benefit and performance appears similar to those with more fixed mindsets. For example, if these students already had lower trait WMC compared to those in Experiment 1 they could feel unsure about putting in the effort or the amount of effort that would give their desired outcome (better performance). Alternatively, it could be that having a more malleable mindset under threat requires more vigilance and uses more cognitive resources compared to having a more fixed mindset, hence observing no benefit of malleable mindset on performance. However, this is simply another interpretation that could not be causally supported or confirmed based on these data.

Overall, the stereotype threat manipulation did not appear to consistently induce threat and produce a performance decrement for Black students. As such, many of the results exploring the threat effect in Experiment 2 aligned with the general findings of Experiment 1. However, additional evidence of trait WMC moderating threat effects provided clarifications for why null effects of stereotype threat may have been observed in some cases—for students with high to extremely high trait WMC these additional cognitive resources may provide them with an advantage for combatting the demands of stereotype threat. For these reasons WMC has been demonstrated to be an important cognitive tool for ensuring that students perform their best. However, it is important to note that gaps in standardized test performance remained in most cases in Experiment 2. The remaining gaps in performance between White and Black students suggests there is room to further explore additional measures of interest for ensuring students perform their best such as important non-cognitive factors like mindset.

Moreover, it was unclear whether the results of Experiment 2 were found due to a need for more statistical power or whether these results might be pointing toward replication issues with stereotype threat effects. In the literature, non-cognitive measures of mindset and grit are argued to be important factors for student resilience in the face of academic as well as other types of challenges. It is possible that there simply were not enough participants in the study in order to detect significant stereotype threat effects in addition to recruiting enough subjects to have adequate coverage in trait mindset and grit (refer to supplemental information for these results as they are not discussed in the text here) across their continua. For these reasons the moderation results where the effect of trait non-cognitive measures were explored must be interpreted with caution. Additionally, because these non-cognitive measures were only explored in terms of their predictive power for moderating threat effects on performance this approach does not allow for causal inferences. Because recent reports in the literature suggest that trait grittiness may not be independent from trait conscientiousness, and because of their moderate to strong relationship, both measures were assessed in Experiment 2 and will be assessed in Experiment 3 as well. As such, in the next chapter Experiment 3 will investigate the role of mindset as an intervention for moderating the effect of stereotype threat on cognitive performance, in addition to further examining whether trait non-cognitive factors of grit and conscientiousness might aid performance under stereotype threat.

Chapter 4

A Mindset Intervention to moderate Stereotype Threat Effects on Working Memory

Capacity and Standardized Test Performance

Experiment 3: Will a Mindset Manipulation moderate Threat Effects?

Experiment 3 is designed to assess whether there is a causal relationship of implicit beliefs about ability and threat effects on cognitive performance. The aim of this experiment is to implement a mindset intervention in order to uncover whether this manipulation will have a positive impact in the context of stereotype threat, and more specifically whether this makes a difference in high achieving and highly motivated samples (to be explored in Experiment 3). The question is whether there is a ceiling on how far malleable mindsets can go and whether having a more malleable mindset is more advantageous for some than others. It is expected that the mindset intervention will moderate the effect of stereotype threat on cognitive performance. Based on the general findings of null effects for stereotype threat for race in Experiments 1 and 2, Experiment 3 will explore the stereotype threat effect for gender. Stereotype threat effects for gender are expected such that females in the threat condition will perform more poorly relative to those in the control condition. In addition, implicit beliefs are expected to moderate the threat effect on cognitive performance such that the intervention of a malleable mindset will “spill over (see Inzlicht & Schmiechel, 2012)” on to subsequent task performance—starting from the most immediate task to any additional tasks completed. The idea is that while stereotype threat may tap valuable cognitive resources, the malleable intervention may aid in the allocation of resources whereas a fixed intervention may be “doubly” harmful during contexts of stereotype threat—if threat effects cause performance decrements and having fixed mindsets are not known

to be necessarily helpful, will combining these levels of manipulations be where performance is lowest?

Moreover, what is interesting about this approach is whether this hypothesis will be observed at the individual differences level? That is, in a high achieving and high performing ivy-league sample will the fixed mindset intervention during contexts of stereotype threat prove to be “doubly” harmful based on allocation of WMC resources? Or possibly, the malleable mindset will make little difference for these students as many already appear to be well equipped for combating stereotype threat effects as evidenced in null effects of threat on performance (revealed in Experiments 1). What’s more interesting will be whether under certain conditions such as in high achieving and highly motivated samples that thinking about your ability in a more fixed way is ok, or even possibly beneficial—suggesting that as long as your ability and performance remain high they can positively reinforce such a belief.

Because Experiment 2 only allowed for correlational inference, the purpose of Experiment 3 is to explore whether mindset manipulations moderate stereotype threat effects. Although past research has explored the effect of mindset interventions on stereotype threat effects, threat was never directly manipulated in those studies (see Aronson, Fried & Good; 2002). The present work will both measure and manipulate mindset while also manipulating stereotype threat. In addition, personality and grit will be measured. It is expected that during contexts of stereotype threat the mindset intervention will moderate the effect of stereotype threat on the outcome measures such that more malleable mindsets will perform better than more fixed mindsets under stereotype threat. Because WMC has been implicated as an important cognitive factor, trait and state WMC will be investigated as potential moderators and mediators of the effect of threat on standardized test performance. In addition, the relationships between

baseline mindset, personality and grit measures and threat effects will be considered—exploring whether higher trait scores on mindset, grittiness and/or conscientiousness moderate the effects of threat on performance.

Method

Participants

Ninety-nine subjects were invited to participate in the study. Based on subjects' failure to complete various measures and/or data missingness (missing data, incomplete tasks, incomplete submissions), there was a final sample of seventy-nine subjects (32 White, 9 Black, 22 Asian, 7 Latino and 9 bi-racial or Other)⁴. Participants were students at Princeton University who identified as female. Subjects received credit toward a course requirement or \$16 cash for their participation.

Design

A 2(Condition: Threat vs. Control) x 2(Mindset: Fixed vs. Malleable) factorial design manipulated the effects of gender stereotype threat and mindset on cognitive performance. Only female subjects were invited to participate.

Procedure

Experiment 3 followed the same general procedure as Experiment 2, with the exception of manipulating mindset in addition to measuring it. The mindset manipulation involved participants reading a prompt claiming that ability is either fixed or malleable followed by a written activity reflecting on the information in the prompt⁵. Importantly, another difference in Experiment 3 is that only the quantitative capacity and intelligence threat or the control condition

⁴ Experimenters noted cases when subjects did not complete all portions of the survey measures. Hence, for some of the measures there was some additional variability in data missingness.

⁵ See Appendices for a complete description of the mindset manipulations.

was implemented—there was no verbal capacity and intelligence condition. Also, here, the role of mindset was assessed under conditions of stereotype threat for gender; whereas, Experiment 2 manipulated stereotype threat for ethnicity. In addition, subjects completed survey measures detailed in Appendices A and B⁶.

Tasks

The tasks and procedure of Experiment 3 are similar to those described in Experiment 2, with the exceptions outlined in the design above.

Results

Data Preparation & Analytic Approach

In each of the analyses below, homogeneity of variance assumptions were tested with *Levene's* test and where the test was statistically significant, *Bonferonni* corrections were used and reported below. In cases where *Levene's* tests were non-significant these results were not reported and normal t-tests and/or tests of analysis of variance are reported below.

It is also important to note, in each regression analysis below levels of condition were dummy coded with control = 0, threat = 1.

Summary Statistics & Correlations

The summary statistics and correlations for the measures are presented in table 4.1 below. As expected the measures of WMC of the OSPAN and RSPAN were positively (moderately) correlated. The OSPAN was also positively correlated with the math GRE. Interestingly, both at baseline and post-test mindset had a weak negative trend in predicting RSPAN suggesting that higher scores on trait WMC on the RSPAN were associated with less malleable mindsets post-test. Grit had weak positive trend in predicting RSPAN scores post-test suggesting that higher

⁶ The personality measure of interest (i.e., trait conscientiousness) was completed before the manipulations. The mindset and grit measures were completed both pre and post manipulations.

trait WMC was associated with higher trait grittiness post manipulations. Trait conscientiousness was highly positively related to grittiness and was weakly related to mindset both pre and post. Additionally, trait conscientiousness did not appear to be meaningfully related to any of the other measures (see Table 4.1 below).

Table 4.1. Summary Statistics and Correlations of Variables in Experiment 3

	1	2	3	4	5	6	7	8	9	10
1. ospan	--									
2. rspan	0.464***	--								
3. gremProp	0.323**	0.161	--							
4. gremAttProp	0.301**	0.101	0.727	--						
5. MindsetScorePre	-0.080	-0.207	-0.147	0.037	--					
6. GritScorePre	-0.054	0.066	-0.036	0.013	-0.269*	--				
7. STAI	-0.163	-0.076	-0.112	-0.135	0.109	-0.026	--			
8. MindsetScorePost	-0.112	-0.197	-0.042	0.173	0.799***	-0.114	0.061	--		
9. GritScorePost	0.026	0.162	-0.049	0.034	-0.259*	0.844***	-0.084	-0.121	--	
10. ConscientiousnessScale	-0.001	0.096	0.031	0.109	-0.205	0.714***	-0.125	-0.085	0.705***	--
<i>M</i>	65.530	59.860	0.540	0.78	4.35	2.73	12.67	4.31	3.45	3.5
<i>SD</i>	9.470	10.200	0.190	0.15	1.29	1.38	3.24	1.25	0.63	0.74

Note. ospan = Operation Span, rspan = Reading Span, gremProp = Proportion Correct on Math GRE, gremAttProp = Proportion Attempted on Math GRE, MindsetScorePre = Malleable Mindset Score at Baseline, GritScorePre = Trait Grittiness Score at baseline, STAI = Spielberger State-Trait Anxiety Inventory, MindsetScorePost = Malleable Mindset Score post-test, GritScorePost = Trait Grittiness Score post-test, ConscientiousnessScale = Trait Conscientiousness. *** $p < .001$, ** $p < .01$, * $p < .05$.

Threat & Mindset Effects

Baseline Effects.

Because there are only women in this study, I did not test for group-based differences in baseline working memory assessed via the RSPAN.

Manipulation Check. To examine the degree to which the manipulations caused participants to endorse more fixed or malleable mindsets a 2(Condition: Threat vs. Control) x 2(Mindset: Fixed vs. Growth) factorial ANOVA was conducted on the mindset measure's difference score. The difference score is the post mindset scale scores subtracted from the pre mindset scale scores (before the manipulations). The mindset scales were reverse scored such

that higher scores indicate more malleable mindsets. In this analysis, the main effect of condition (i.e., threat vs. control) was non-significant, $F(1, 74) = .22, p > .05, \eta_{\text{partial}}^2 = .00302$. As expected, the main effect of mindset was significant, $F(1, 74) = 12.52, p < .001, \eta_{\text{partial}}^2 = .145$, indicating that the mindset manipulations made participants more malleable ($M = .21, SD = .70$) or fixed ($M = -.37, SD = .75$). However, there was an unexpected marginally significant interaction, $F(1, 74) = 2.37, p = .13, \eta_{\text{partial}}^2 = .0311$. *Tukey's* pairwise comparisons revealed that the mindset manipulations did significantly affect people's endorsement of these perspectives as expected in the control condition ($M_{\text{diff}} = .809, p = .0027$)—participants endorsed malleable mindsets more in the malleable condition and fixed mindsets more in the fixed condition (see Figure 4.1 below). This difference did not emerge when people were in the threat condition ($M_{\text{diff}} = .306, p = .58$)—people who received the malleable instructions ($M = .21, SD = .70$) did not significantly differ in endorsement from participants who received the fixed instructions ($M = -.37, SD = .75$).

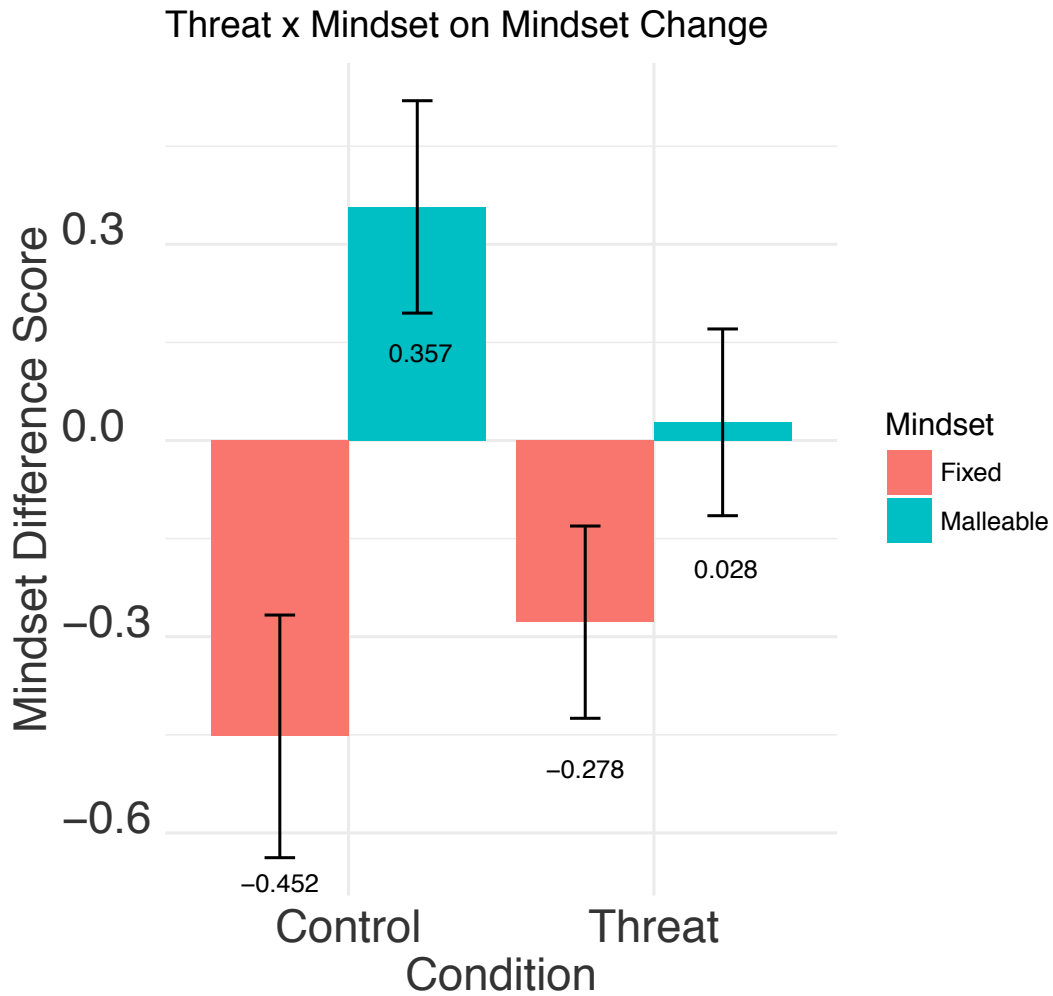


Figure 4.1. Effects of Gender Stereotype Threat and Mindset on Mindset Difference.

Effect of Manipulations on Cognitive Measures.

Next, two 2(Condition: Threat vs. Control) x 2(Mindset: Fixed vs. Growth) factorial ANOVAs analyzed these effects on cognitive performance on WMC via the OSPAN and standardized test performance on the math GRE. Results revealed, for state WMC on the OSPAN, a significant main effect of threat, $F(1, 75) = 5.6, p < .05, \eta_{\text{partial}}^2 = .0693$, such that performance was higher in the threat ($M = 68.5, SD = 7.04$) condition than the control ($M = 63.4, SD = 10.5$). The main effect of mindset, $F(1, 75) = 1.08, p > .05, \eta_{\text{partial}}^2 = .0103$ and the

interaction were not significant, $F(1, 75) = .027, p > .05, \eta_{\text{partial}}^2 = .00036^7$. These results are presented in Figure 4.2 below.

On the math GRE, the ANOVA results indicated a significant main effect of threat, $F(1, 78) = 4.31, p < .05, \eta_{\text{partial}}^2 = .0523$, such that performance was higher in the threat ($M = .59, SD = .20$) than the control ($M = .50, SD = .17$) condition. The main effect of mindset, $F(1, 78) = .002, p > .05, \eta_{\text{partial}}^2 = .000104$ and the interaction were not significant, $F(1, 78) = .318, p > .05, \eta_{\text{partial}}^2 = .0041^8$. These results are presented in Figure 4.3 below.

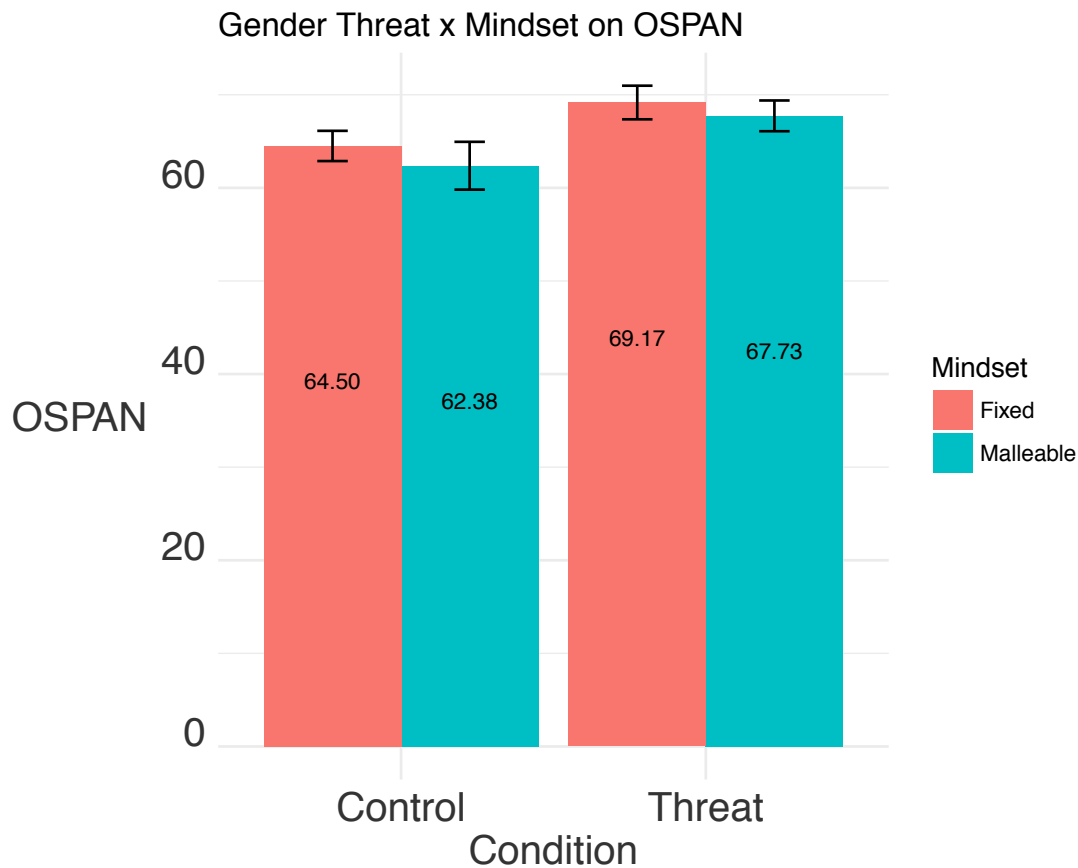


Figure 4.2. Effects of Gender Stereotype Threat and Mindset Interventions on WMC.

⁷ Four subjects working memory scores on the OSPAN were unavailable due to a computer error.

⁸ One subject's data was missing on the math GRE due to a lack of completion.

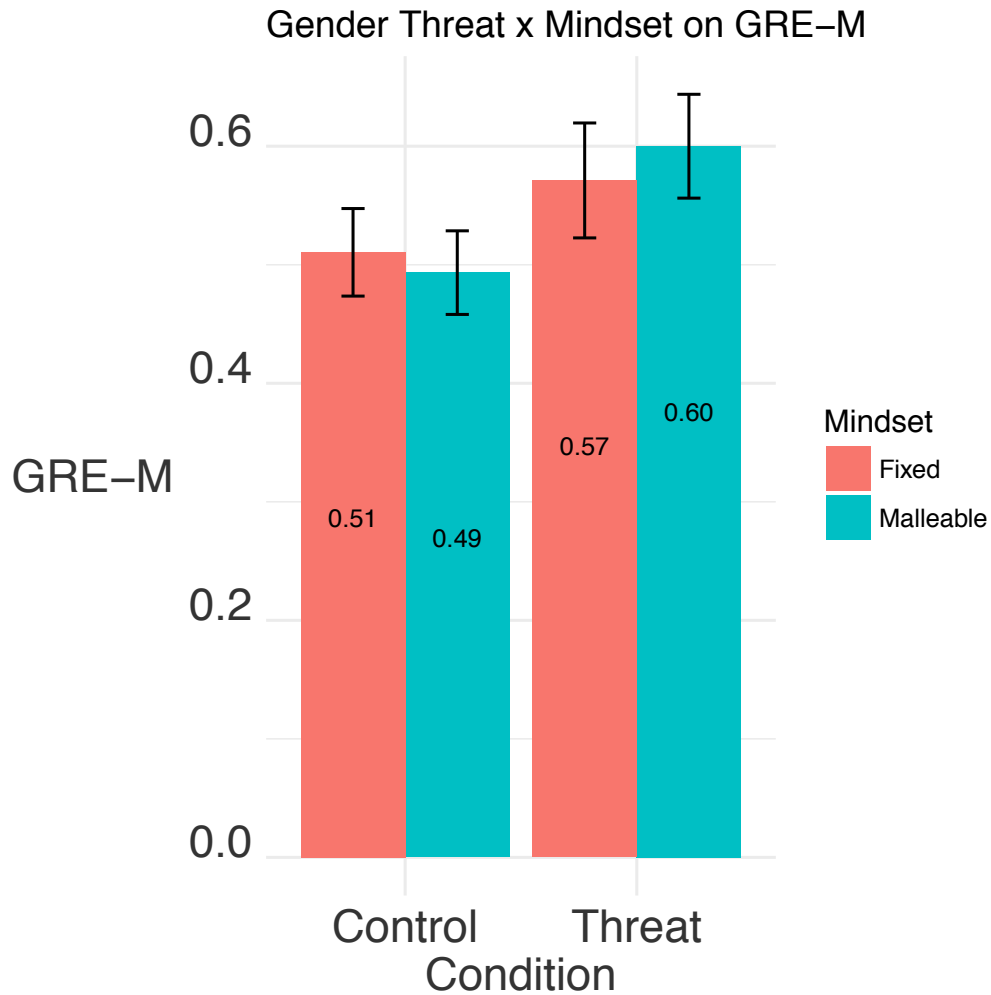


Figure 4.3. Effects of Gender Stereotype Threat and Mindset Interventions on GRE-M.

Next, in line with Experiments 1 and 2, the role of trait and state WMC was assessed under these conditions, however, based on the lack of evidence that the mindset manipulation moderated stereotype threat effects on WMC and standardized test performance, the effect of the mindset manipulation was collapsed across these analyses.

Mediation.

A mediation analysis was conducted with condition as the independent variable, OSPAN, the mediator variable, and GRE-M as the outcome variable. The direct effect of threat on GRE-M was marginally significant, $t(78) = 1.91, b = .078, p = .06$ indicating that the gender threat condition had a predicted .078 increase on the GRE-M for women. The effect of threat on OSPAN was significant, $t(75) = 2.16, b = 3.52, p < .05$, indicating that the threat effect predicted a 3.52 increase on OSPAN. The indirect effect of threat on GRE-M when OSPAN was added to the model was not significant, $t(73) = 1.43, b = .060, p > .05$. In the indirect path, the effect of OSPAN on GRE-M was significant $t(73) = 2.13, b = .0061, p < .05$. *Sobel's* test further supported the lack of evidence for statistical mediation of OSPAN in the effect of gender threat on GRE-M, $Z = 1.53, p = .16$ (see Figure 4.4).

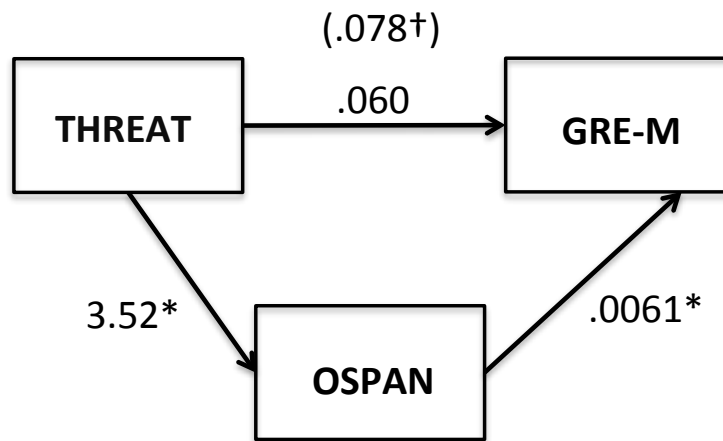


Figure 4.4. Model of WMC Mediating Gender Threat Effect on GRE-M (Princeton).

Bootstrapping methods resampled 1000 simulations to further test for mediation. Results indicated marginally significant effects for the average causal mediation effect, parameter estimate = .022, $CI [-.0049, .0577], p = .13$, as well as the average direct effect of threat on GRE-

M, parameter estimate = .060, $CI [-.0217, .1584]$, $p = .15$. Taken together, these findings reveal a lack of strong statistical evidence that women's WMC (as assessed via the OSPAN) mediated the effect of threat on GRE-M (see Figure 4.5).

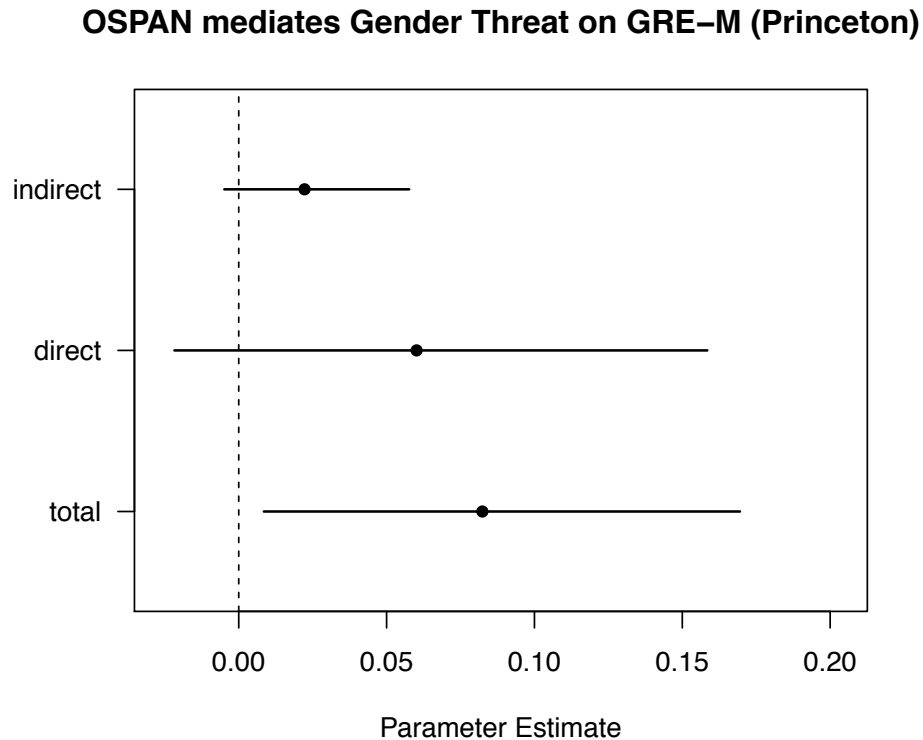


Figure 4.5. WMC Mediation Model for Threat on GRE-M using Bootstrapping (Princeton).

Moderation.

To test whether WMC moderates the effect of gender stereotype threat on standardized test performance, a hierarchical regression was conducted with the outcome math GRE. In the first step, the main effects of RSPAN (mean centered) and condition were investigated. In the second step, the product of RSPAN (mean centered) and condition (dummy coded with control = 0, threat = 1) was added in order to test for moderation.

Looking at math GRE in step one, the effect of RSPAN was not significant, $b = .0025$, $t(73) = 1.22$, $p > .05$. The model, $F(2, 73) = 2.72$, Multiple $R^2 = .069$, $p = .07$ was marginally significant. The effect of condition was marginally significant, $b = .082$, $t(73) = 1.98$, $p = .05$. In step two, the effect of RSPAN approached significance, $b = .0049$, $t(72) = 1.85$, $p = .06$, indicating higher trait WMC predicted higher scores on the math GRE. The effect of condition was significant, $b = .083$, $t(72) = 2.02$, $p < .05$, indicating higher predicted scores on math GRE under threat. The interaction was not significant, $b = -.0057$, $t(72) = -1.41$, $p > .05$. The model approached significance, $F(3, 72) = 2.5$, Multiple $R^2 = .094$, $p = .07$. Taken together, these findings indicate that under threat higher trait WMC did not provide a boost compared to lower trait WMC on the math GRE (see Figure 4.6).

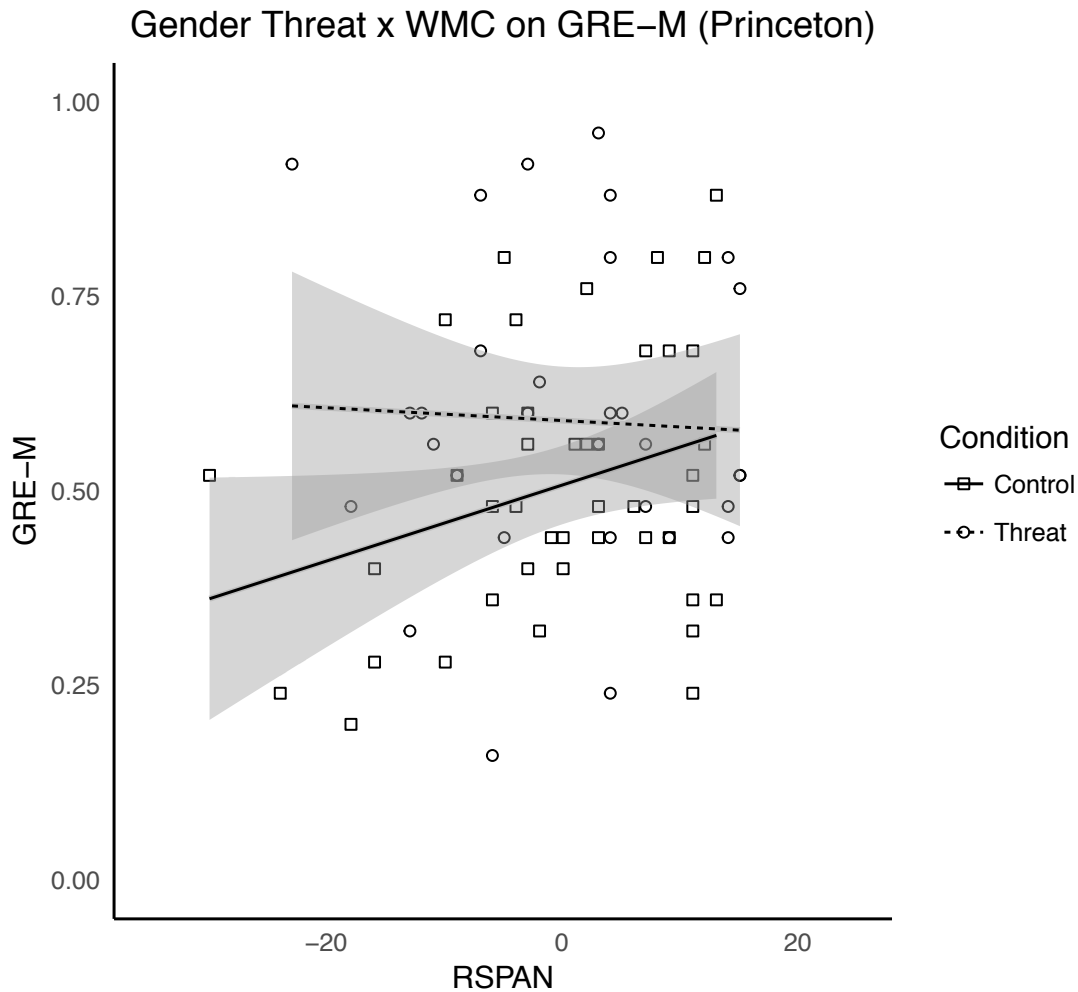


Figure 4.6. RSPAN Moderation Model for Threat on GRE-M

Threat Effects & Non-Cognitive Factors.

Data Preparation & Analytic Approach.

Next, the role of trait mindset⁹ was explored as a possible important non-cognitive factor during contexts of stereotype threat. It is important to note that based on the lack of evidence that the mindset manipulation moderated stereotype threat effects in Experiment 3, the effect of the mindset manipulation was collapsed across these analyses. Specifically, here trait mindset is

⁹ This was the mindset scale score at baseline, that is, before any of the manipulations.

explored as a moderator of the effect of threat on performance outcomes. Mindset measures are typically split based on categories of more fixed or malleable mindsets, but in an effort to preserve statistical power, the mindset measure was kept on a continuous scale. If of interest, the reader is reminded to review the full report of summary statistics for all measures presented in Table 4.1. Other non-cognitive factors of interest—grit and conscientiousness—were also measured. However, the results of these exploratory analyses are not presented here. For the analyses with the grit and conscientiousness measures from this experiment, the reader should consult the supplemental information section.

Mindset Regression Analyses.

Next, A series of hierarchical regression analyses were conducted to assess whether individual differences in trait mindset moderate stereotype threat effects on performance. Under gender stereotype threat, women's performance was expected to be negatively impacted compared to the control condition. It is important to note that because only women were recruited in the study, in each of the regression analyses below the outcome measures of WMC and standardized test performance were regressed on condition ("Control": 0, "Threat ": 1), trait mindset (mean-centered), and their interactions.

In step one, looking at performance on the OPSAN, the effect of mindset was not significant, $b = -.86$, $t(73) = -1.28$, $p > .05$. The effect of condition was significant, $b = 3.67$, $t(73) = 2.23$, $p < .05$, indicating that predicted scores on the OSPAN were higher under threat. The model, $F(2, 73) = 3.25$, Multiple $R^2 = .082$, $p < .05$ was also significant.

In step two predicting OSPAN, the effect of mindset, $b = -.998$, $t(72) = -1.16$, $p > .05$ was not significant. The effect of condition, $b = 3.67$, $t(72) = 2.22$, $p < .05$ was significant, indicating that predicted scores on OSPAN were higher under threat. The interaction was not

significant, $b = .0018$, $t(72) = .57$, $p > .05$. The model approached significance, $F(3, 72) = 2.16$, Multiple $R^2 = .083$, $p = .10$. Taken together, these results reveal a lack of support for trait mindset scores (more malleable mindsets) moderating the effect of threat on OSPAN (see Figure 4.7).

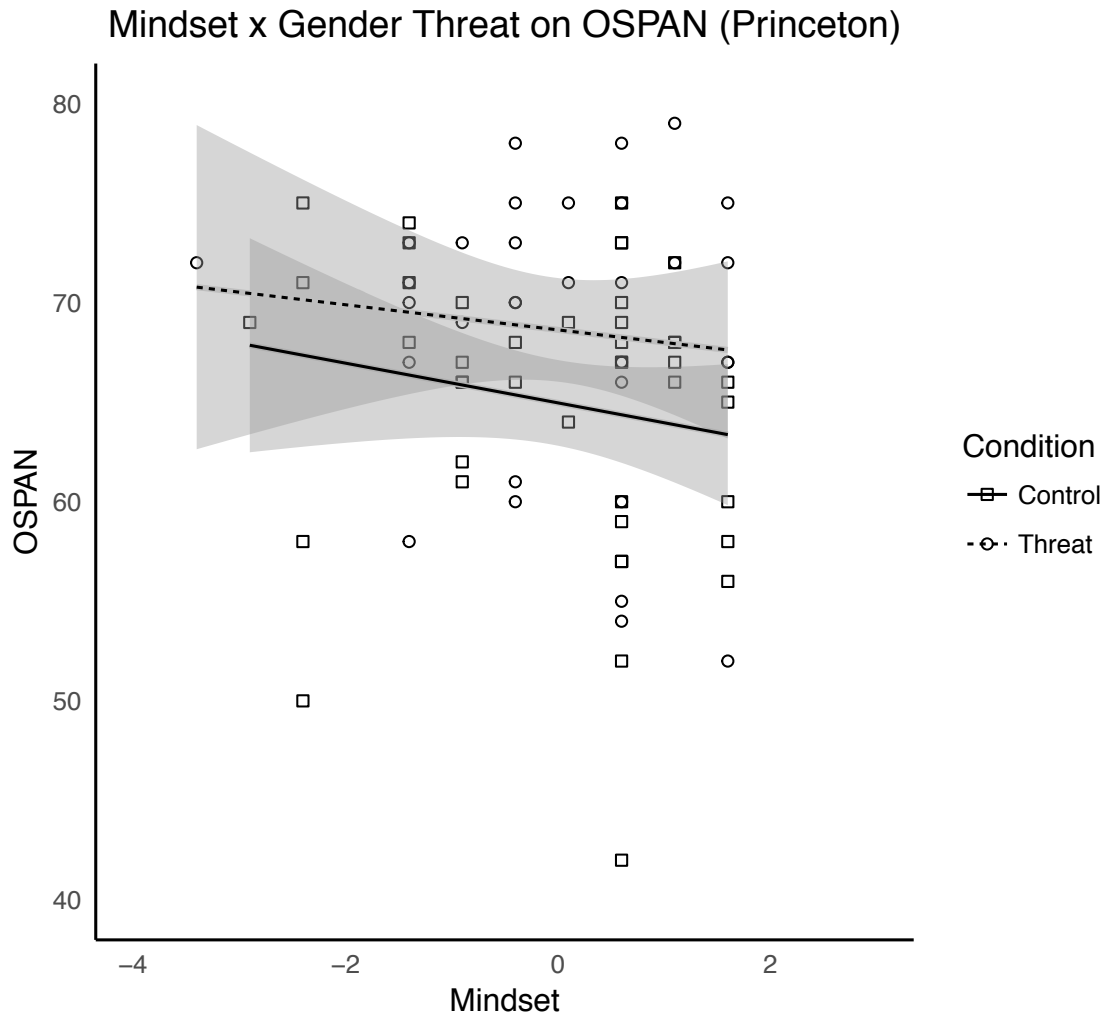


Figure 4.7. Mindset Moderation Model for Threat on OSPAN.

Next, a series of hierarchical regression analyses were conducted in order to test whether mindset moderated gender stereotype threat on math standardized test performance.

In step one looking at the effects on the math GRE, the effect of condition approached significance, $b = .081$, $t(76) = 1.96$, $p = .05$, indicating higher predicted scores on math GRE

under threat. The effect of mindset was not significant, $b = -.026$, $t(76) = -1.53$, $p < .05$. The model, $F(2, 76) = 2.96$, Multiple $R^2 = .072$, $p = .06$ approached significance.

In step two, the effect of mindset, $b = -.014$, $t(75) = -.61$, $p > .05$ was not significant. The effect of condition approached significance, $b = .081$, $t(75) = 1.96$, $p = .05$, indicating that the threat effect predicted higher scores on the math GRE. The model approached significance, $F(3, 75) = 2.21$, Multiple $R^2 = .081$, $p = .09$. However, the interaction was not significant, $b = -.029$, $t(75) = -.868$, $p > .05$, indicating a lack of evidence for trait mindset moderating the effect of threat on the math GRE (see Figure 4.8).

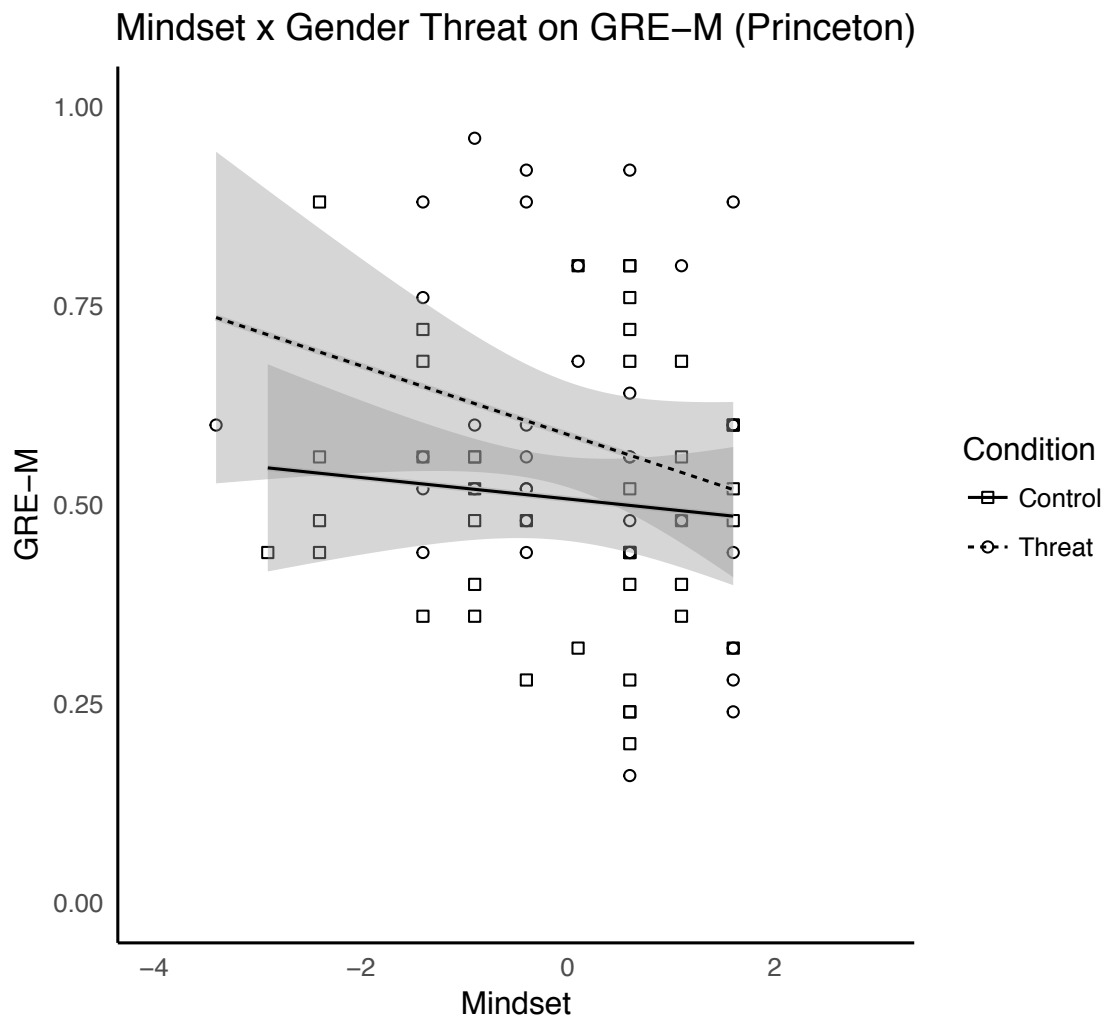


Figure 4.8. Mindset Moderation Model for Threat on GRE-M.

Experiment 3 Summary & Discussion

The summary of evidence gathered in Experiment 3 is somewhat in line with what was revealed in Experiments 1 and 2. Here, we do find some evidence that stereotype threat impacts women's math performance, however the effect of stereotype threat on OSPAN seemed to boost as opposed to decrease performance. This suggests as with Experiments 1 and 2 that the participants were resilient to the effects of stereotype threat. Additional findings of a lack of a performance decrement on the math GRE under stereotype threat suggest that these threat manipulations did not have a detrimental impact on women's math performance.

In general, the manipulations did appear to effectively make participants endorse more fixed or malleable mindsets when stereotype threat was absent, in that students' attitudes became more fixed or malleable from baseline, respectively. In essence, the mindset manipulation worked in the intended direction but only when stereotype threat was absent (in the control condition). Additionally, in this sample of students, the threat manipulation appeared to motivate students in order to perform better than those in the control condition on both the state WMC and the standardized test performance measures. Despite the presence of a similar pattern for state WMC and standardized test performance, there was no strong statistical evidence of state WMC mediating the effect of stereotype threat on standardized test performance. Because the threat effect tended to go in the opposite direction of the prediction, in the cases where there was weak evidence of mediation this was for a boost as opposed to a performance decrement (see Figures 4.4 and 4.5).

Also opposite than predicted, the fixed and malleable conditions had similar performance on the WMC and standardized test performance measures in the threat condition. This suggests that the mindset manipulation did not have the expected moderating effect on stereotype threat.

Under threat participants appeared to have a boost relative to control and instead of the malleable mindset providing subjects with “protection” against threat effects, these data revealed that both fixed and malleable mindset manipulations had equal performance under threat. That is, inducing a fixed mindset did not cause performance to suffer more under threat compared to inducing a malleable mindset.

Taken together, these findings suggest that although the non-cognitive factor of mindset is an important skill in the context of education and achievement, when it comes to stereotype threat mindset may not be as powerful for brief interventions as once believed. These findings tell a story about both the psychological process and mechanisms involved in student performance and achievement. In the present experiment, the students are high achieving and highly motivated. It may have been the case that these students did not need much convincing that effort can bring about positive change or that facing challenges throughout the academic journey is a guarantee. However, in terms of stereotype threat, if anything, the malleable mindset did not appear to differentially improve student performance under stereotype threat. Instead, both fixed and malleable mindset manipulations showed improved performance under stereotype threat compared to the control condition.

Additionally, these data revealed there was no evidence that trait WMC or trait mindset moderated the effect of stereotype threat on standardized test performance. The lack of evidence for trait WMC moderating the effect on standardized test performance was contrary to the findings of Experiments 1 and 2. This may have been the case because the threat was for gender here as opposed to ethnicity in the other studies. In common with Experiments 1 and 2 was the consistent pattern of the threat condition providing a boost for performance compared to the control condition.

The findings from the present experiment reveal a lack of evidence for stereotype threat effects. Additionally, these findings de-emphasize the importance of the influence of non-cognitive factors like mindset because higher scores on this measure appeared to be inconsequential under stereotype threat and did not reveal the predicted benefit of performance enhancements under stereotype threat (see Figures 4.8 and 4.9). Additionally, even when mindset was manipulated the fixed mindset manipulation did not depress performance compared to the malleable manipulation, as expected. It is possible that both fixed and malleable mindset manipulations made subjects reflect on academic performance and instances of challenge and effort and students thought about this more than they did the framing or underlying arguments that intelligence and intellectual ability are either fixed or malleable. In this case, the mindset manipulation could've made subjects a little more fixed or malleable but this was not powerful enough to depress or enhance performance under stereotype threat. It is also possible that because these subjects are high achieving students they can reflect on previous setbacks but not be fully convinced of the importance of focusing on either the fixedness or malleability of intellectual ability but rather, to focus on performing their best. Moreover, in this case, performing better could be easier for this group of students compared to other groups of students. Future work would need to replicate this study with a more diverse and varied student sample in terms of ability and beliefs about ability. Also, future work would also need to add a control condition where there is no mention of mindset.

Chapter 5

General Discussion & Conclusions

The goal of the present project is to move the field forward in our understanding of the intersection of certain social psychological phenomena and their impacts on cognition and cognitive behavior. Specifically, the project first aimed to investigate the cognitive mechanisms underlying the effect of stereotype threat on performance.

Based on the results of Experiment 1, there were null effects of stereotype threat on outcomes of WMC and standardized test performance on sections of the math and verbal GREs. For this reason, there was also a lack of evidence that state WMC mediated the effect of stereotype threat on standardized test performance. Because there were null effects of stereotype threat on the performance outcomes there essentially was no effect to mediate in this case.

However, Experiment 1 revealed evidence that higher trait WMC moderated the effect of stereotype threat on standardized test performance on both the verbal and math GREs. Moderation based on higher trait WMC was only found when students were under stereotype threat for quantitative capacity and intelligence. There was no evidence that the effect of stereotype threat for verbal capacity and intelligence on standardized test performance was moderated by higher trait WMC. Instead, performance in the verbal capacity and intelligence condition was flat under stereotype threat across the range of lower to higher WMC. This suggests that there may be differences in the way that stereotype threat influences performance across different domains in this sample of Princeton students.

Interestingly, the Educational Testing Service reports that the verbal reasoning sections of the Graduate Record Exam require test-takers to draw inferences from incomplete data, in

addition to understanding advanced vocabulary and the relationships between words, whereas; the quantitative reasoning section requires test-takers to understand quantitative information to solve problems and apply basic math skills from elementary algebra, arithmetic and geometry (ETS, 2018, *The GRE General Test*). In addition, PrepScholar—a test preparation website for the GRE—suggests that because the verbal reasoning portions employs advanced vocabulary and the quantitative portion mainly recycles math students were exposed to on the SAT (formerly the Scholastic Achievement Test, now called the Scholastic Assessment Test), the verbal reasoning section may feel more challenging than the quantitative section—assuming that the test-taker is very familiar and comfortable with the math presented in the SAT (PrepScholar, 2018, *GRE vs. SAT: 13 Key Differences*). Based on this information this suggests that the GRE verbal reasoning section may have been perceived as more difficult for these students. This could explain why the higher WMC students were unable to use it to their advantage and perform better under stereotype threat for verbal capacity and intelligence either on the verbal GRE or the math subsection that followed. This could also suggest that the higher trait WMC benefit was essentially used toward completing the task at hand and wasn't available for combatting the effect of stereotype threat in this condition.

Interestingly, those in the control (who were not put under stereotype threat) for the verbal capacity and intelligence condition were able to kick up their performance on both GREs when trait WMC was higher than lower (see Figures 2.9 and 2.10). This provides additional evidence that the effects of stereotype threat for verbal information coupled with a challenging verbal standardized test followed by a math standardized test may have simply overtaxed these participants in a way that they were unable to use the higher trait WMC to their advantage in this situation.

Another interpretation of the moderation differences across threat conditions is that moderation was only found in these data when the baseline measure of WMC was on the RSPAN. When OSPAN was the baseline measure of WMC there was no evidence of moderation. This may suggest that the predictive power of trait WMC based on whether OSPAN or RSPAN was used could bring about dissociations in the effects observed. However, it is unclear whether this could be the case and this interpretation is highly speculative based on the pattern observed in these data.

Because Experiment 1 contained a highly motivated, high achieving, student sample this prompted the question of whether in this context an overarching implicit belief about ability may have also attenuated the effect of stereotype threat on performance? Based on the results of Experiment 1 this was suspected post-hoc based on considering the method of how stereotype threat was induced. By stating that the task was “highly correlated with measures of intelligence,” this could activate a larger and broader identity among these participants, perhaps that of an “ivy-league” student and upon taking the verbal and quantitative sub-sections of the Graduate Record Examination (GRE) this identity domain may have protected everyone from feeling threatened and incidentally functioned as a performance enhancement in some cases rather than a decrement.

Thus, considering the possibility of different implicit beliefs about ability (mindset) the next aim was to directly test the role that cognitive and non-cognitive factors play during contexts of stereotype threat. As such, the extent to which differences in beliefs about ability (as well as other non-cognitive factors explored) both correlate and interact with the effects of stereotype threat on cognitive performance were investigated. The first hypothesis was that implicit beliefs would have an effect under stereotype threat such that more malleable mindsets

would be associated with better performance under threat relative to those with more fixed mindsets (explored in Experiment 2). The second hypothesis was that when implicit beliefs were manipulated as a form of intervention that under stereotype threat, performance would be best for those in the malleable mindset group compared to the fixed group (implemented in Experiment 3).

The results of Experiment 2 replicated the patterns of Experiment 1 in that there were not significant effects of stereotype threat on performance in the intended direction—excluding the effect of verbal capacity and intelligence threat on the RSPAN measure (see Figure 3.3). In addition, general trends in the threat conditions revealed that these subjects tended to have a performance boost in the threat conditions compared to the control. Based on the lack of evidence of stereotype threat effects in the intended direction, subsequently, there was also a lack of evidence for state WMC mediating the threat effect on standardized test performance.

Partially replicating the findings of Experiment 1, there was evidence that trait WMC moderated the effect of stereotype threat on standardized test performance. However, baseline WMC only moderated the effect of stereotype threat on performance on the second GRE that these students completed. That is, in the verbal capacity and intelligence threat condition baseline WMC moderated threat on the math GRE and in the quantitative capacity and intelligence threat condition baseline WMC moderated threat on the verbal GRE. This pattern of results suggests that for these students there was something important about the second GRE such that higher WMC students were able to kick up their performance. This may have been the case because the second GRE was less tied toward the domain specific component of the stereotype threat manipulation—the mention that the task is indicative of “verbal/quantitative capacity.” For these students this may have overshadowed the subsequent information that the

task is “highly correlated with measures of intelligence.” In a sense, these students could kick up their performance on the GREs when the GRE was in a different domain than what they were received stereotype threat for (e.g., taking the math GRE when threatened for verbal capacity and intelligence).

In terms of the non-cognitive measures in Experiment 2, trait mindset scores did not have any significant moderating effects on stereotype threat effects on WMC or standardized test performance. However, in some cases weak trends in the data revealed some evidence that more malleable mindsets predicted better scores on the WMC and standardized test performance measures but this finding was weak and inconsistent for both Black and White subjects. Due to high attrition rates in terms of completing the non-cognitive surveys there was a lower sample size than expected thus these trends in the data must be interpreted with caution (refer to the supplemental information section for full analyses of the non-cognitive exploratory measures of grit and conscientiousness).

In Experiment 3, mindset was manipulated and it was expected that more malleable mindsets would provide a similar benefit to students under stereotype threat for gender as higher trait WMC had provided students under threat for ethnicity in Experiments 1 and 2. Here, results revealed findings opposite of this prediction in that the fixed mindset group performed as well as the malleable mindset group under stereotype threat. Trends in these data also suggested that the threat manipulation provided participants with a boost compared to the control condition. Moreover, looking at trait mindset scores (before the manipulation) also revealed no evidence that trait mindset moderates the effect of stereotype threat. These data suggest the potential positive effects and/or benefits of non-cognitive measures were negligible and suggest that these factors may not play as important of a role under stereotype threat as expected. Additionally,

there was no evidence that WMC or mediated or moderated stereotype threat—instead there was a clear and consistent pattern of a performance boost under stereotype threat.

Taken together the results of Experiments 1-2 provide support for a moderation account for the role that WMC plays during contexts of stereotype threat. Here, across three studies these data suggest that higher trait WMC may help students combat the effects of stereotype threat on performance. However, when higher WMC provides a benefit under threat may also depend on the domain of the task. These data suggest that when Black students are threatened for verbal information in the Princeton sample these students who had higher WMC were not able to use it to their advantage for combatting threat. However when students were threatened for quantitative information both Princeton and Towson students (only on the second GRE which was the math section) were able to use higher WMC to their advantage and perform better under threat compared to lower WMC individuals. Combining these results with those of Experiment 3 these data reveal that cognitive factors play an important role in combatting stereotype threat however, non-cognitive factors may play less of an important a role. It was predicted that more malleable mindsets would help students perform better under stereotype threat compared to more fixed mindsets. Here, the data revealed no difference in performance on WMC or standardized tests based on having a malleable mindset or a fixed mindset under stereotype threat.

Limitations & Future Research

The present work was unable to replicate the basic stereotype threat effect for race and gender in the form of a performance decrement this provides additional evidence for replication issues. Whether this was the case on the basis of the complexity of the language used in manipulation (i.e., “this task is a measure of quantitative/verbal capacity and is highly related to measures of intelligence”) or based on a sample size issue future investigations may find it more

useful to replicate the effect based on more direct language used in the original study (i.e., “this task is diagnostic/non-diagnostic of ability”). Here because of the complexity of the language used in the threat manipulation it is unclear whether participants consistently focused on one part of the manipulation more than others. Because the present work was designed to closely mimic previous investigations of threat effects where WMC was explored the manipulation and language was modeled closely after those studies (i.e., Schmader & Johns, 2003; Regner et al., 2010) as opposed to the original Steele and Aronson (1995) study. In order to have a better understanding of how threat operates one must first capture the effect and demonstrate that it can be pushed around based on these subtle differences in language and/or differences in the level of implicitness vs. explicitness with which the manipulation is administered.

Generally, it would be to researcher’s advantage in this area to recruit samples that exceed the convention of 80% power. Due to issues of replicability of the effect and the small effect sizes reported in reviews (Spencer, Logel & Davies, 2016) and meta-analyses of stereotype threat (see Flore & Wicherts, 2014) the sample size needed in order to detect these effects may be much higher than originally anticipated. Future researchers should consider this carefully when planning and designing future studies in this area. This was especially the case when aiming to have adequate statistical power in capturing the range of beliefs and attitudes in the scales of the non-cognitive factors.

Additionally, based on the consistent finding of trait WMC moderating stereotype threat across task domains and two university samples future work should focus attention on finding ways to conserve precious mental resources, especially for students who struggle based on stigmatization and/or educational inequity. As recent work in the WMC literature finds that WMC may be enhanced through training (see Jaeggi et al., 2008; Jaeggi et al., 2011a; Jaeggi et

al., 2013; Redick et al., 2013) future work should consider this and additional forms of cognitive intervention that may be helpful for students at the individual level. For example, mindfulness techniques and other forms of self-regulation have been shown to improve cognitive functioning and may improve working memory as well (Mrazek et al., 2013; Morrison & Jha, 2015).

Investigating whether these types of interventions are effective for helping students of stigmatized identities boost academic performance should be fruitful for future work in this area. Moreover, mindfulness techniques and practices may be ideal because they may provide benefits to mental and physical health in addition to cognitive ones.

Expected Contribution & Conclusion

The paper opened with a quote by W.E.B. Du Bois' because it demonstrates several important points about the nature of one's beliefs about ability and performance across cultural, social, racial and even historical lines. In a way, Du Bois' argument is an all-encompassing perspective that applies not only to the Black Americans, who have been and continue to be stigmatized, but additionally has implications for any person who understands what it feels like to be stigmatized. The quote acknowledges the legacy of environmental and/or social structures at work in the process of stigmatization allowing for potential disruptions in performance and achievement.

The current work sets the stage with a more nuanced and novel re-conceptualization of stereotype threat that considers both the experimental and the differential perspectives while including non-cognitive measures as concrete additions to previous cognitive models, viewing the threat effect from more of a measurement-focused perspective. In *The Handbook of Theories of Social Psychology*, Carol Dweck once stated: "I have been particularly interested in beliefs with strong motivational properties. It might be interesting from an intellectual standpoint that

people can come to different conclusions about the nature of themselves and others, but it becomes even more intriguing if the different conclusions make a difference for the goals people pursue and the outcomes they experience in their lives.” This remark highlights the importance of not only the discrepancy between the way others view our ability and how we view it but also whether these views and beliefs themselves play an important role in future success and achievement. Because this process may be at work during our day-to-day lives, exploring the interaction between cognitive and non-cognitive factors in achievement is just as important to consider today as it was when Du Bois’ first brought this notion of *double consciousness* to our attention. Although the present work did not find strong empirical evidence supporting the power motivational properties in the form of non-cognitive factors the present results still provide hope.

One can only imagine that blocking out negative views about us regarding ability is compounded in complexity when certain situational pressures are activated in the environment. The aim, then, becomes to help those who feel stigmatized by finding ways to resolve these issues. The present work provides consistent evidence that the negative impacts of having a stigmatized identity can be overcome through effectively directing cognitive resources toward the task at hand. As such, finding ways to focus and preserve mental resources such as practicing mindfulness and self-regulation in both short and longer-term contexts, may be beneficial. Ideally, by focusing mental resources efficiently and with consistency should promote better performance and produce greater long-term success.

Although issues of inequality and gaps in education and achievement remain, the current work helps provide a better understanding of how these issues function from one student to the next. The comprehensive approach of combining experimental and differential approaches in the stereotype threat literature will help the field continue to uncover how, when and why threat

operates and how to better combat it. Ultimately, this approach and the findings revealed in this line of research are valuable for mapping out future forms of intervention. With efforts in the field to design effective forms of intervention that are individually “targeted, tailored, and timely (see Cohen, Garcia & Goyer, 2017),” the comprehensive approach presented in the current work should be useful. The hope is that this work will be both informative and impactful—smoothing the road of achievement and success for future students and world leaders, in spite of negative views based on their identity or over-generalizations about their capability.

References

- 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, 29-46.
- Aronson, J. (1999). The effects of conceiving ability as fixed or improvable on responses to stereotype threat. Unpublished Manuscript, New York University, New York.
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38, 113-125.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–90). New York: Academic Press.
- Bandura, M., & Dweck, C.S. (1985). The relationship of conceptions of intelligence and achievement goals to achievement-related cognition, affect and behavior. In E. T. Higgins, & A. W. Kruglanski, *Motivational Science: Social and Personality Perspectives*. Psychology Press.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637.
- Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and “choking under pressure” in math. *Psychological Science*, 16, 101-105.
- Beilock, S.L., & DeCaro, M.S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 983–998.

- Beilock, S.L., Rydell, R.J., & McConnell, A.R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, 136, 256–276.
- Binet, A. (1909/1973). *Modern ideas on children*. Paris: Flammarion.
- Blackwell, L.S., Trzesniewski, K., & Dweck, C.S. (2007). Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, 78, 246-263.
- Borsboom, D., Romeijn, J., & Wicherts, J. M. (2008). Measurement invariance versus selection invariance: is fair selection possible? *Psychological Methods*, 13, 75-98.
- Cheryan, S., & Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: The psychological hazards of "model minority" status. *Psychological Science*, 11, 399-402.
- Cohen, G. L., Garcia, J., & Goyer, J. P. (2017). Turning point: Targeted, tailored, and timely psychological intervention. *Handbook of competence and motivation (2nd ed.): Theory and application*. New York, NY: Guilford.
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769-786.
- Conway, A.R.A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science*, 7, 547-552.
- Conway, A.R.A., & Kovacs, K. (2013). Individual differences in intelligence and working memory: A review of latent variable models. *Psychology of Learning and Motivation*, 58, 233-270.

- Conway, A. R. A., & Kovacs, K. (2015). New and emerging models of human intelligence. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6, 419-426
- Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, 113(3), 492.
- Cury, F., Da Fonseca, D., Zahn, I., & Elliot, A. (2008). Implicit theories and IQ test performance: A sequential mediational analysis. *Journal of Experimental Social Psychology*, 44, 783-791.
- Daker, R., & Conway, A.R.A. (unpublished manuscript). The Varying Effects of Stereotype Threat, Lift, and Priming on Cognitive Performance.
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Du Bois, W.E.B. (1903). *The Souls of Black Folk: Essays and Sketches*. Chicago: A.C. McClurg & Company.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, 92(6), 1087.
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT-S). *Journal of personality assessment*, 91(2), 166-174.
- Duckworth, A.L., & Yeager, D.S., (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44, 237-251.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child development*, 82(1), 405-432.

- Dweck, C. S. (2011). Implicit Theories. in P.V. Lange, A. Kruglanski, & T. Higgins. (Eds.1) *The Handbook of Theories of Social Psychology*.
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY US: Random House.
- Dweck, C.S. (1999). *Self-theories: their role in motivation, personality, and development*. USA: Taylor & Francis
- Dweck, C.S. & Leggett, E.L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- Dweck, C. S., & Sorich, L. A. (1999). Mastery-oriented thinking. In C. R. Snyder (Ed.), *Coping: The psychology of what works*. New York: Oxford University Press.
- Educational Testing Service. (2018). *The GRE General Test*. Retrieved from: <https://www.ets.org/gre/institutions/about/general/>.
- Education Week, Vol. 35, Issue 28, Page 6
- Elliott, A. J., & Dweck, C.S. (2005). *Handbook of Competence and Motivation*. New York: The Guilford Press.
- Flore, P.C., & Wicherts, J. M. (2014). Does stereotype threat influence performance of stereotyped domains. A meta-analysis. *Journal of School Psychology*, 53, 25-44.
- Flynn, J. R. (1984). "The Mean IQ of Americans: Massive Gains 1932 to 1978." *Psychological Bulletin*, 95, 29–51.
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working Memory Skills and Educational Attainment: Evidence from National Curriculum Assessments at 7 and 14 Years of Age. *Applied Cognitive Psychology*, 18, 1-16.

- Good, C., Rattan, A., & Dweck, C.S. (2008). Development of the sense of belonging to math survey of adults: A longitudinal study of women in calculus. Unpublished manuscript.
- Goodwin, J.C. (2008). *A history of modern psychology*. (Eds. 3). USA: Wiley & Sons
- Heckhausen, J. & Dweck, C.S. (1998). *Motivation and Self-Regulation across the lifespan*. United Kingdom: Cambridge University Press.
- Gould, S.J. (1981). *The Mismeasure of Man*. New York: W. W. Norton & Company.
- Herrnstein, R. & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Simon & Schuster.
- Higgins, E. T., Friedman, R. S., Harlow, R. E., Idson, L. C., Ayduk, O. N., & Taylor, A. (2001). Achievement orientations from subjective histories of success: Promotion pride versus prevention pride. *European Journal of Social Psychology*, 31, 3-23.
- 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.
- Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, 7, 450-463.
- Inzlicht, M. (2016). “*Reckoning with the past*.” Blog entry.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105, 6829–6833.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011a). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, 108, 10081–10086.

- Jaeggi, S. M., Buschkuhl, M., Shah, P., & Jonides, J. (2013). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, 42, 464-480.
- Jensen, A. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1-123.
- Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion—Is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science*, 21, 1686-1693.
- Johnson-Laird, P. N. (2009). Reasoning. In P. Rabbitt (Ed.), *Inside Psychology: A Science over 50 Years*. (pp. 167-77). Oxford: Oxford University Press.
- Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*.
- Kray, L.J., & Haselhuhn, M. P. (2007). Implicit negotiation beliefs and performance: Experimental and longitudinal evidence. *Journal of Personality and Social Psychology*, 93, 49-64.
- Leggett, E. L. (1985, March). *Children's entity and incremental theories of intelligence: Relationships to achievement behavior*. Paper presented at the annual meeting of the Eastern Psychological Association, Boston.
- Licht, B. G., & Dweck, C. S. (1984). Determinants of academic achievement: The interaction of children's achievement orientations with skill area. *Developmental Psychology*, 20, 628-636.
- Lippmann, W. (1922). *Public Opinion*. New York: Harcourt Brace.
- Logel, C., Iserman, E. C., Davies, P. G., Quinn, D. M., & Spencer, S. J. (2009). The perils of double consciousness: The role of thought suppression in stereotype threat. *Journal of Experimental Social Psychology*, 45, 299-312.

- Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, 51, 93-120.
- Marteau, T. M., & Bekker, H. (1992). The development of a six item short form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, 31, 301-306.
- McIntosh, P. (1988). "White Privilege and Male Privilege: A Personal Account of Coming to See Correspondences Through Work in Women's Studies." Wellesley: Center for Research on Women.
- Morrison, A. B., & Jha, A. P. (2015). Mindfulness, attention, and working memory. In *Handbook of mindfulness and self-regulation* (pp. 33-45). Springer, New York, NY.
- Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness training improves working memory capacity and GRE performance while reducing mind wandering. *Psychological science*, 24(5), 776-781.
- Mueller, C. M. & Dweck, C. S. (1998). Intelligence praise can undermine motivation and performance. *Journal of Personality and Social Psychology*, 75, 33-52.
- Nettles, Michael. (2017). Why the Whole Child Matters. Educational Testing Service Open Notes. Retrieved from: <http://news.ets.org/stories/whole-child-matters/>
- Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments—A comprehensive review of the literature. *Teaching in Higher Education*, 8, 195–209.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.

- Paunesku, D., Walton, G., Romero, C., Smith, E., Yeager, D., & Dweck, C. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science*, 26, 784-93.
- Pinel, E. C. (1999). Stigma consciousness: the psychological legacy of social stereotypes. *Journal of personality and social psychology*, 76(1), 114.
- Plaks, J., Stroessner, S., Dweck, C. S., & Sherman, J. (2001). Person theories and attention allocation: Preference for stereotypic vs. counterstereotypic information. *Journal of Personality and Social Psychology*, 80, 876-893.
- Plaks, J. E., & Stecher, K. (2007). Unexpected improvement, decline, and stasis: A prediction confidence perspective on achievement success and failure. *Journal of Personality and Social Psychology*, 93, 667-684.
- Plaks, J. E., Grant, H., & Dweck, C.S. (2005). Violations of implicit theories and the sense of prediction and control: Implications for motivated person perception. *Journal of Personality and Social Psychology*, 88, 245-262.
- PrepScholar, 2018, *GRE vs. SAT: 13 Key Differences*. Retrieved from: <https://www.prepscholar.com/gre/blog/gre-vs-sat/>
- Raven, J. C. (1990). *Advanced Progressive Matrices: Sets I, II*. Oxford: Oxford University Press.
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D., Hambrick, D. Z., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142, 359–379.

- Regner, I., Smeding, A., Gimmig, D., Thinus-Blanc, A., Monteil, J., & Huguet, P. (2010). Individual differences in working memory moderate stereotype-threat effects. *Psychological Science, 21*, 1646-1648.
- Ritchie, S. (2015). *Intelligence: All that matters*. Hodder & Stoughton.
- Rogoff, B. E., & Lave, J. E. (1984). *Everyday cognition: Its development in social context*. Harvard University Press.
- 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.
- Schimmack, U. (2017). *Hidden Figures: Replication Failures in the Stereotype Threat Literature*. Blog entry. Retrieved from: <https://replicationindex.wordpress.com/2017/04/07/hidden-figures-replication-failures-in-the-stereotype-threat-literature/>.
- Schimmack, U. (2016). The Replicability-Index: Quantifying Statistical Research Integrity. <https://wordpress.com/post/replication-index.wordpress.com/920>
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194-201.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440- 452.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.

- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201-292.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4-28.
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual review of psychology*, 67, 415-437.
- 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.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613.
- Steele, C. M. (2010). *Whistling Vivaldi: And Other Clues to How Stereotypes Affect Us*. New York: W.W. Norton & Company.
- Sternberg, R. (2005). On Intelligence and Ability. In A. J. Elliott, & C.S. Dweck, *The Handbook of Competence and Motivation*. New York: The Guilford Press.
- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, 77, 1213-1227.
- Stricker, L. J. (1998). Inquiring about examinees' ethnicity and sex: Effects on AP Calculus AB examination performance (College Board Rep. 98-1, ETS Res. Rep. 98-5). New York, NY: College Board.

- Stricker, L. J., & Ward, W. C. (1998). Inquiring about examinees' ethnicity and sex: Effects on Computerized Placement Test performance (College Board Rep. 98-2, ETS Res. Rep. 98-9). New York, NY: College Board.
- 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.
- Tough, P. (2014). *Who Gets to Graduate?*. New York Times.
- Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology*, 9, 507-533.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498-505.
- Walton, G. M. (2013). The myth of intelligence: Smartness isn't like height. In D. Allen & R. Reich (Eds.) *Education, Justice, and Democracy* (pp. 155-172). Chicago, IL: The University of Chicago Press.
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92, 82-96.
- Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes among minority students. *Science*, 331, 1447-1451
- Wheeler, S.C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: a review of possible mechanisms. *Psychological Bulletin*, 127, 797- 826.

- Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: a question of measurement invariance. *Journal of Personality and Social Psychology*, 89, 696-716.
- Wicherts, J. M., & Millsap, R. E. (2009). The absence of underprediction does not imply the absence of measurement bias. *American Psychologist*, 64, 281.
- Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice*, 29, 39-47.
- Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid *p*-Hacking. *Frontiers in Psychology*, 7, 1832. <http://doi.org/10.3389/fpsyg.2016.01832>
- Yeager, D., Trzesniewski, K. H., Tirri, K., Nokelainen, P., & Dweck, C.S. (2009). Adolescents' Implicit Theories Predict Desire for Vengeance After Peer Conflicts: Correlational and Experimental Evidence. *Developmental Psychology*, 47, 1090-1107.
- Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of educational Research*, 81, 267-301.
- Zajonc, R. B., Heingartner, A., & Herman, E. M. (1969). Social Enhancement and Impairment of Performance in the Cockroach. *Journal of Personality and Social Psychology*, 13, 83-92.

Appendices

Appendix A

Experiment 2 Additional Tasks

Non-Cognitive Items.

A survey of non-cognitive items was administered after the experiment, ranging from 1 week to about 1 month after the session. The survey included several questions about implicit beliefs about general cognitive ability, in addition to personality measures (e.g., BFI and grit/perseverance items).

Implicit Beliefs Scale. The standard implicit beliefs scale was used to get a sense for how flexible students are in their beliefs about ability. The scale includes two questions asking subjects about their implicit beliefs about intelligence and ability. The questions are designed to deem whether participant's belief about ability in the domain is more fixed or more malleable.

Grit Scale. The grit scale was used to gauge student levels of enduring effort over time. The scale includes eight items about perseverance and dedication to task completion with few distractions (e.g., "I finish whatever I begin."). The questions are designed to reveal how long participants will stick with a task and focus on completing it in spite of difficulty or distractions.

Personality Scale. The personality measure was included to get a thorough assessment of students' temperament. The scale included 44 items that assessed the "big five" aspects of personality: extraversion, openness, agreeableness, neuroticism and conscientiousness. The questions are designed to reveal how high participants score on each of these items. The personality scale is of special interest for exploring non-cognitive factors and their relationship to cognitive performance.

School Identification Measure. The school identification measure asked a series of questions about how much students identify with their school. Because activating different identities is known to exacerbate or alleviate stereotype threat effects this measure was included to control for possible differences that may help/harm performance under stereotype threat.

Appendix B

Experiment 3 Tasks

Based on the procedure described in Chapter 4 a detailed summary of these tasks is outlined below.

Reading Span. The automated RSPAN will be included in order to assess baseline/trait WMC. It requires making veridical judgments for sentences while remembering a list of letters (e.g., making a veridical judgment for the sentence “when at last his eyes opened, there was no gleam of triumph, no shade of anger” Yes or No? R). Participants were instructed to read the sentence, answer yes or no, and then remember the letter for later recall. At the end of a list of such trials, participants were asked to recall the letters in serial order (total scored using the partial unit method, see Conway et al., 2005). Participants received three to seven letters per trial and three sets of each trial length, totaling 15 trials total—yielding a maximum score of 75.

Fixed Mindset Manipulation. The fixed mindset manipulation informed subjects that intellectual ability has been shown to be innate and unchangeable and used a crafted example of a research article that claims to demonstrate this information (see Blackwell et al., 2007; Yeager & Walton, 2011). Participants then completed a short activity where they detailed an instance where they worked hard but increased effort did not improve their performance. Participants

were randomly assigned to either the fixed or growth manipulation conditions prior to the stereotype threat or control conditions.

Growth Mindset Manipulation. The growth mindset manipulation informed subjects that research shows that abilities are malleable as people continue to learn and grow and that with effort intellectual ability can be improved. The manipulation used a crafted example of a research article demonstrating this to strengthen the effect (see Blackwell et al., 2007; Yeager & Walton, 2011). Participants then completed a short activity where they detailed an instance where they needed to work hard in order to improve. Participants were randomly assigned to either the fixed or growth manipulation conditions prior to the stereotype threat or control conditions.

Stereotype Threat Manipulation. The threat manipulation procedure was based on Schmader and Johns (2003)—only female participants were recruited and were either assigned to the threat or control condition. In the threat condition subjects were informed that the OSPAN task was indicative of “quantitative capacity” ability and were then informed this capacity was indicative of group ability. Subjects were then instructed to complete a gender identification survey, which consisted of a single question asking participants to indicate their gender. In the control condition the OSPAN task was simply described as a “working memory capacity” measure.

Operation Span. The automated OSPAN was included to assess state WMC when females are under stereotype threat. The task requires completing a series of arithmetic problems while remembering a list of letters. For example, an individual trial in a list would take the following form: $[5 \times 9] + 5 = 45?$ Yes or No? P. Participants were instructed to solve the equation, answer yes or no, and then remember the letter for later recall. At the end of a list of

such trials, participants were asked to recall the letters in serial order (total score was calculated using the partial unit method, see Conway et al., 2005). Participants received three to seven letters per trial and three sets of each trial length, totaling 15 trials total—yielding a maximum score of 75.

GRE-Mathematics. The GRE mathematics measure was included to assess standardized math performance. The subsection consisted of 25 multiple-choice or short answer questions, each requiring mathematical, reasoning and quantitative comparison skills (e.g., Solve the equation $5 = 4x + 3$ for x . Is x greater than, less than or equal to 1?) Participants were given 20 minutes to complete the subsection. The section was taken from free online practice materials provided by the Educational Testing Service. The final score was the proportion of questions correct, i.e., total correct out of 25 possible questions.

Experiment Survey

In addition to the survey items included in Experiment 2 the items detailed below were also included.

SAT Scores. Scholastic Achievement Test Scores were collected by self-report during the experiment survey. These scores will be used to control for potential differences in baseline SAT based on group or gender identity.

GPA Scores. To account for potential differences among high achieving students participants were asked to self-report Grade Point Average Scores during the experiment survey.

Course Domain Identification Survey. To account for individual differences in the levels of student interest with different course domains the survey had students to rate their identification with several course domains (e.g., verbal intensive courses, quantitative intensive courses, STEM courses). The purpose of the survey was to assess how much students identify

with these course domains after the manipulation so that it did not impact their behavior during the experiment.

Manipulation Check. During the experiment survey, students were asked again (they were also asked before the manipulations to secure a baseline measure) about their implicit beliefs regarding ability in addition to questions about academic attitudes and abilities (see Aronson, Fried & Good). The questions were measured on a 6-pt Likert, scale of *strongly agree* to *strongly disagree*.

Supplemental Information

Chapter 3

Threat Effects & Non-Cognitive Factors

A series of regression analyses were conducted to explore whether individual differences in trait grit and trait conscientiousness moderate student performance under conditions of stereotype threat. Stereotype threat effects were expected to only negatively impact Black student performance. However due to the low statistical power in these data both White and Black subjects were included in the moderation analyses with the non-cognitive factors below. Thus, in each of the regression analyses below the hypothesis is that trait non-cognitive factors, grit and conscientiousness will moderate the effect of stereotype threat on the outcomes of WMC and standardized test performance. The outcome measures were regressed on race (“White”: 1, “Black ”: –1) condition (“Control”: 1, “Threat ”: –1), trait grittiness (mean-centered) or trait conscientiousness (mean-centered), and their interactions, respectively. These series of analyses are presented in sets below based on the quantitative capacity and intelligence condition followed by the verbal capacity and intelligence condition in order to align with the presentation of the results in Chapter 3.

Quantitative Capacity & Intelligence Condition.

Grit. Looking at the effect of trait grittiness moderating the effect of stereotype threat on WMC via OSPAN, regression analyses revealed no significant main effects of race, condition, grit, nor their interactions were significant (see Table 1 and Figure 1). The moderation model was also not significant, $F(7, 23) = .643$, Multiple $R^2 = .16$, $p > .05$.

Table 1. *Grit, Threat and Race Effects on OSPAN Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.727	3.524	0.206	0.040	0.838
condition	-1.421	3.830	-0.371	-0.078	0.714
grit	4.839	5.177	0.935	0.194	0.358
Model 2					
race	2.191	3.997	0.548	0.119	0.589
condition	-1.142	3.997	-0.286	-0.063	0.778
grit	2.529	8.016	0.316	0.101	0.755
race x condition	4.806	3.997	1.202	0.258	0.242
condition x grit	4.037	8.016	0.504	0.148	0.619
race x grit	8.050	8.016	1.004	0.324	0.326
race x condition x grit	-0.238	8.016	-0.030	-0.010	0.977

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

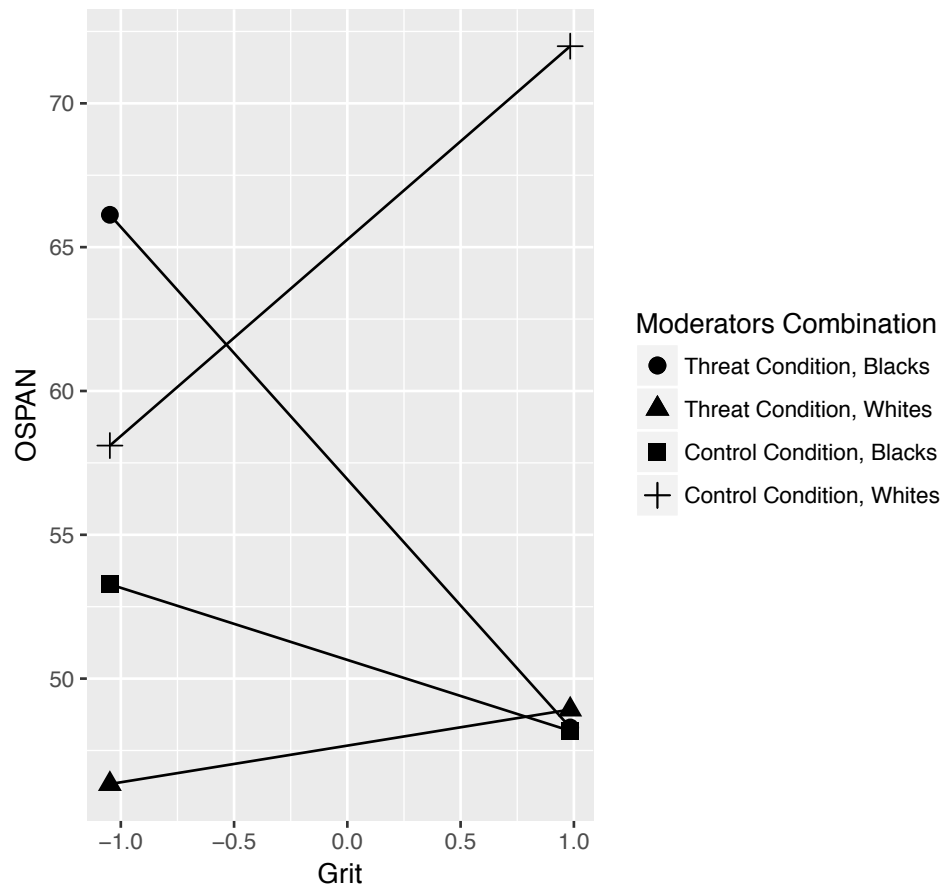


Figure 1. Simple Slopes of Grit moderating Stereotype Threat on OSPAN.

Looking at the effect of trait grittiness moderating the effect of stereotype threat on math GRE, regression analyses revealed only a significant main effect of race such that Whites performed better than Blacks. There were no other significant main effects or interactions (see Table 2 and Figure 2). The moderation model was also not significant, $F(7, 23) = 1.16$, Multiple $R^2 = .26$, $p > .05$.

Table 2. *Grit, Threat and Race Effects on GRE-M Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.054	0.023	2.321*	0.413	0.028
condition	-0.012	0.026	-0.454	-0.089	0.653
grit	-0.025	0.034	-0.713	-0.137	0.482
Model 2					
race	0.050	0.027	1.851†	0.378	0.077
condition	-0.011	0.027	-0.393	-0.081	0.698
grit	-0.063	0.054	-1.177	-0.355	0.251
race x condition	0.020	0.027	0.747	0.151	0.462
condition x grit	0.050	0.054	0.927	0.256	0.364
race x grit	0.059	0.054	1.090	0.330	0.287
race x condition x grit	-0.064	0.054	-1.189	-0.358	0.246

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

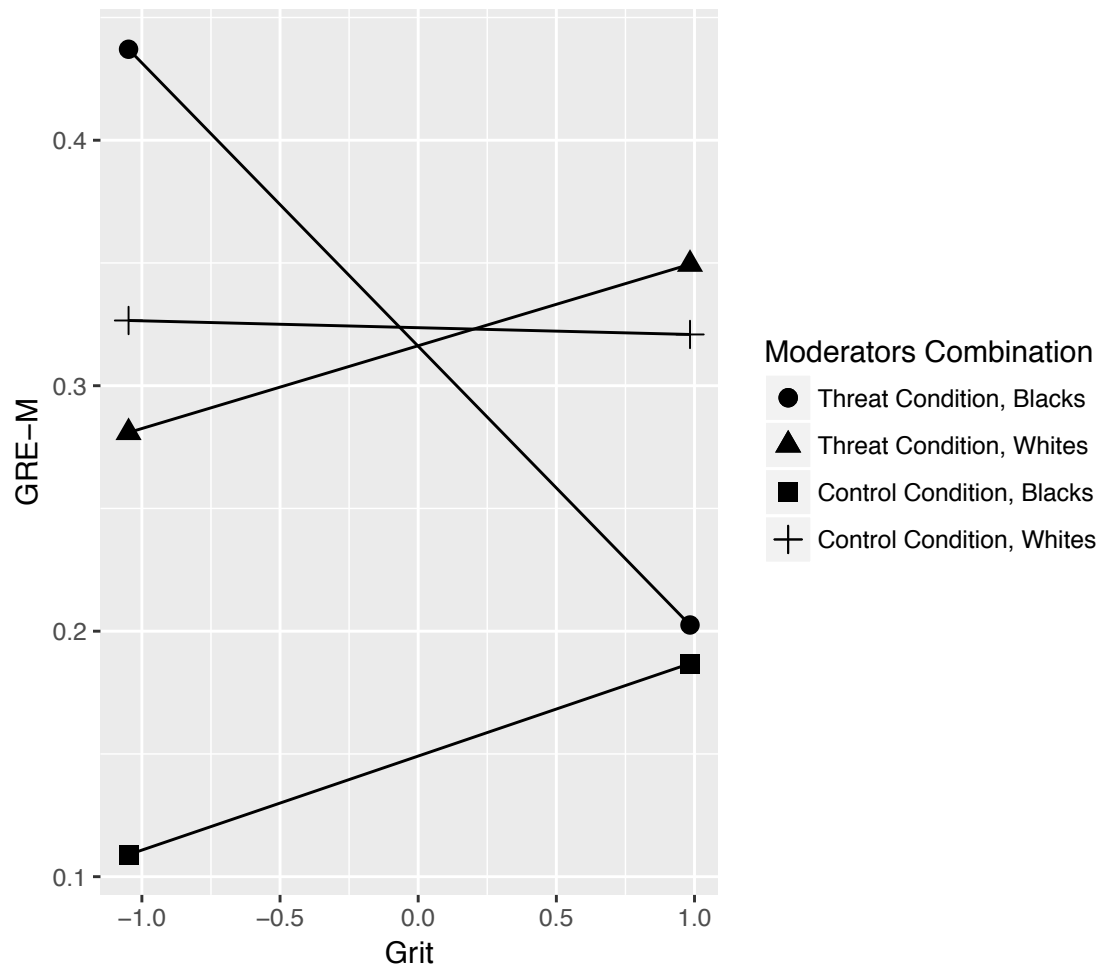


Figure 2. Simple Slopes of Grit moderating Stereotype Threat on GRE-M.

Looking at the effect of trait grittiness moderating the effect of stereotype threat on verbal GRE, regression analyses revealed no significant main effects or interactions. (see Table 3 and Figure 3). The moderation model was also not significant, $F(7, 23) = .329$, Multiple $R^2 = .09$, $p > .05$.

Table 3. *Grit, Threat and Race Effects on GRE-V Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.021	0.023	0.912	0.174	0.370
condition	-0.024	0.025	-0.925	-0.194	0.363
grit	-0.018	0.034	-0.537	-0.111	0.596
Model 2					
race	0.022	0.028	0.804	0.182	0.430
condition	-0.026	0.028	-0.948	-0.217	0.353
grit	-0.013	0.056	-0.224	-0.075	0.825
race x condition	-0.011	0.028	-0.378	-0.085	0.709
condition x grit	-0.005	0.056	-0.086	-0.026	0.932
race x grit	-0.037	0.056	-0.657	-0.221	0.518
race x condition x grit	0.023	0.056	0.417	0.139	0.681

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

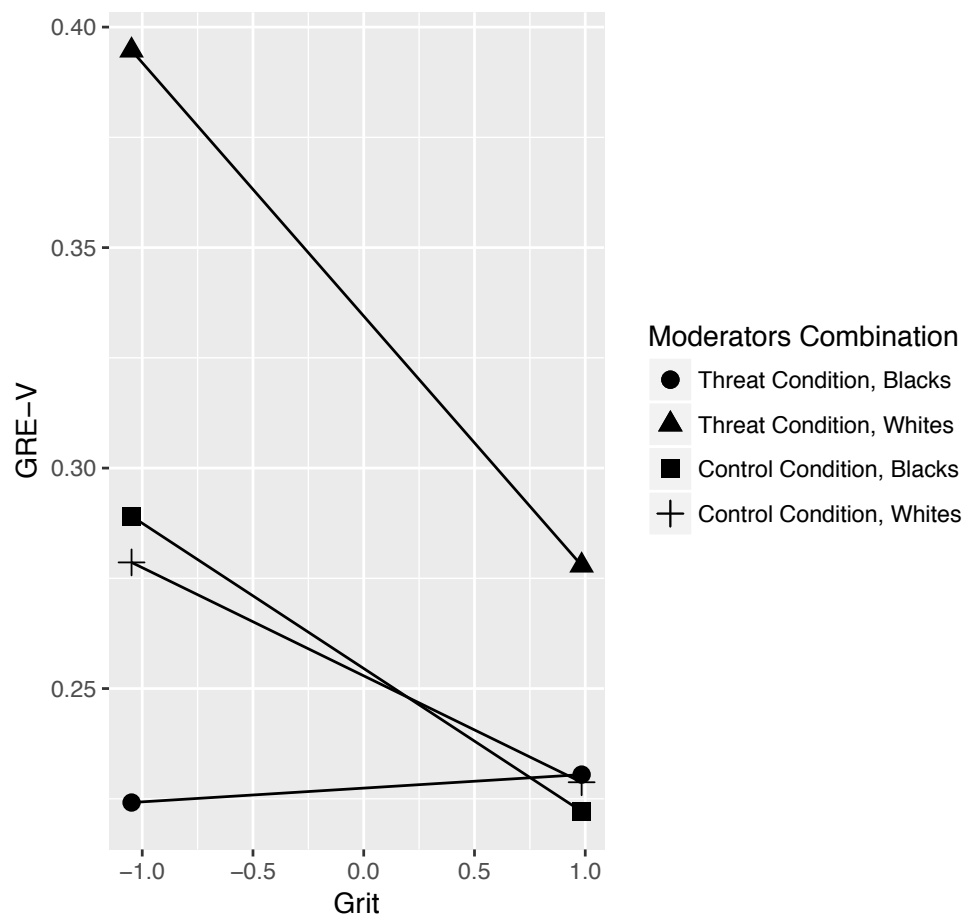


Figure 3. Simple Slopes of Grit moderating Stereotype Threat on GRE-V.

Conscientiousness. Looking at the effect of trait conscientiousness moderating the effect of stereotype threat on WMC via OSPAN, regression analyses revealed no significant main effects or interactions (see Table 4 and Figure 4). The moderation model was also not significant, $F(7, 23) = .719$, Multiple $R^2 = .18$, $p > .05$.

Table 4. *Conscientiousness, Threat and Race Effects on OSPAN Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.805	3.575	0.225	0.044	0.824
condition	-2.602	3.668	-0.709	-0.143	0.484
conscientiousness	1.392	4.392	0.317	0.063	0.754
Model 2					
race	2.539	3.746	0.678	0.138	0.505
condition	-2.056	3.746	-0.549	-0.113	0.588
conscientiousness	6.899	7.985	0.864	0.311	0.397
race x condition	4.146	3.746	1.107	0.222	0.280
condition x conscientiousness	-0.524	7.985	-0.066	-0.023	0.948
race x mindset	2.375	7.985	0.297	0.108	0.769
race x condition x conscientiousness	7.801	7.985	0.977	0.354	0.339

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

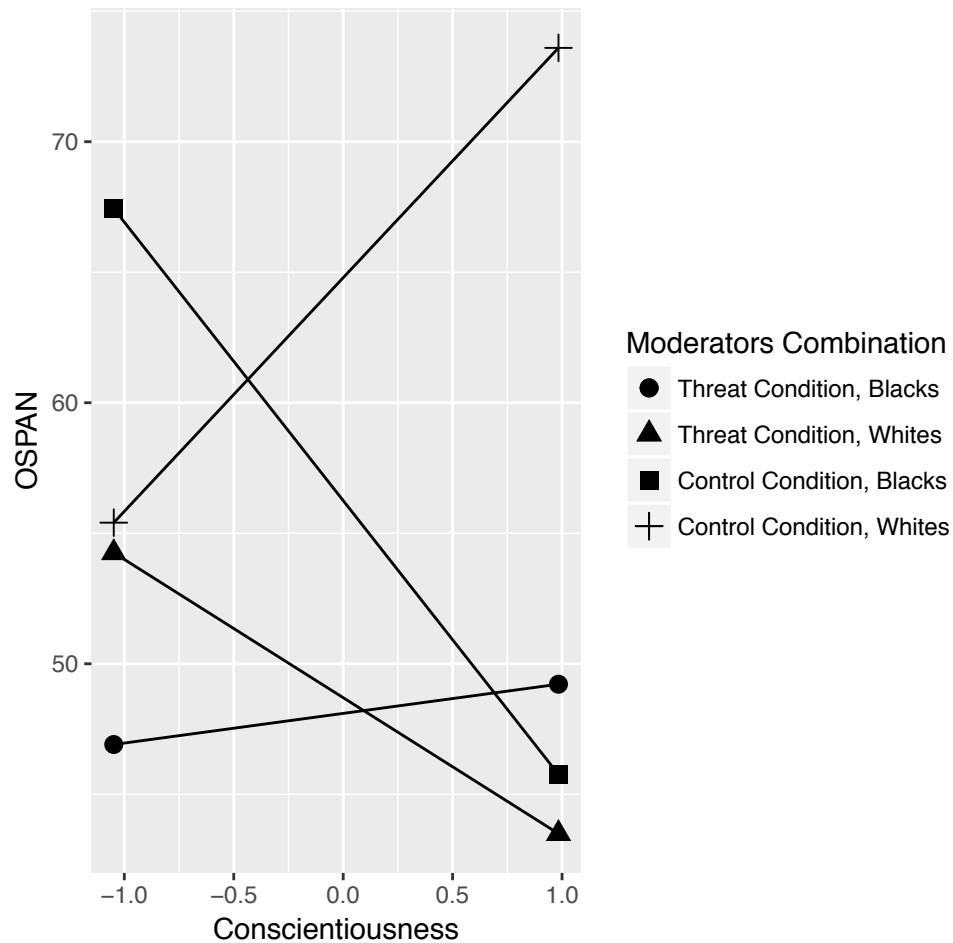


Figure 4. Simple Slopes of Conscientiousness moderating Stereotype Threat on OSPAN.

Looking at the effect of trait conscientiousness moderating the effect of stereotype threat on standardized test performance on the math GRE, regression analyses revealed only a significant main effect of race such that Whites had higher predicted scores on the math GRE. There were no other significant effects (see Table 5 and Figure 5). The moderation model was also not significant, $F(7, 23) = 1.33$, Multiple $R^2 = .29$, $p > .05$.

Table 5. *Conscientiousness, Threat and Race Effects on GRE-M Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.054	0.024	2.293*	0.411	0.030
condition	-0.006	0.024	-0.265	-0.049	0.793
conscientiousness	-0.011	0.029	-0.370	-0.068	0.715
Model 2					
race	0.047	0.025	1.899†	0.360	0.070
condition	-0.007	0.025	-0.295	-0.057	0.771
conscientiousness	0.040	0.053	0.755	0.253	0.458
race x condition	0.024	0.025	0.948	0.177	0.353
condition x conscientiousness	-0.073	0.053	-1.362	-0.441	0.186
race x mindset	0.001	0.053	0.021	0.007	0.983
race x condition x conscientiousness	-0.008	0.053	-0.145	-0.049	0.886

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

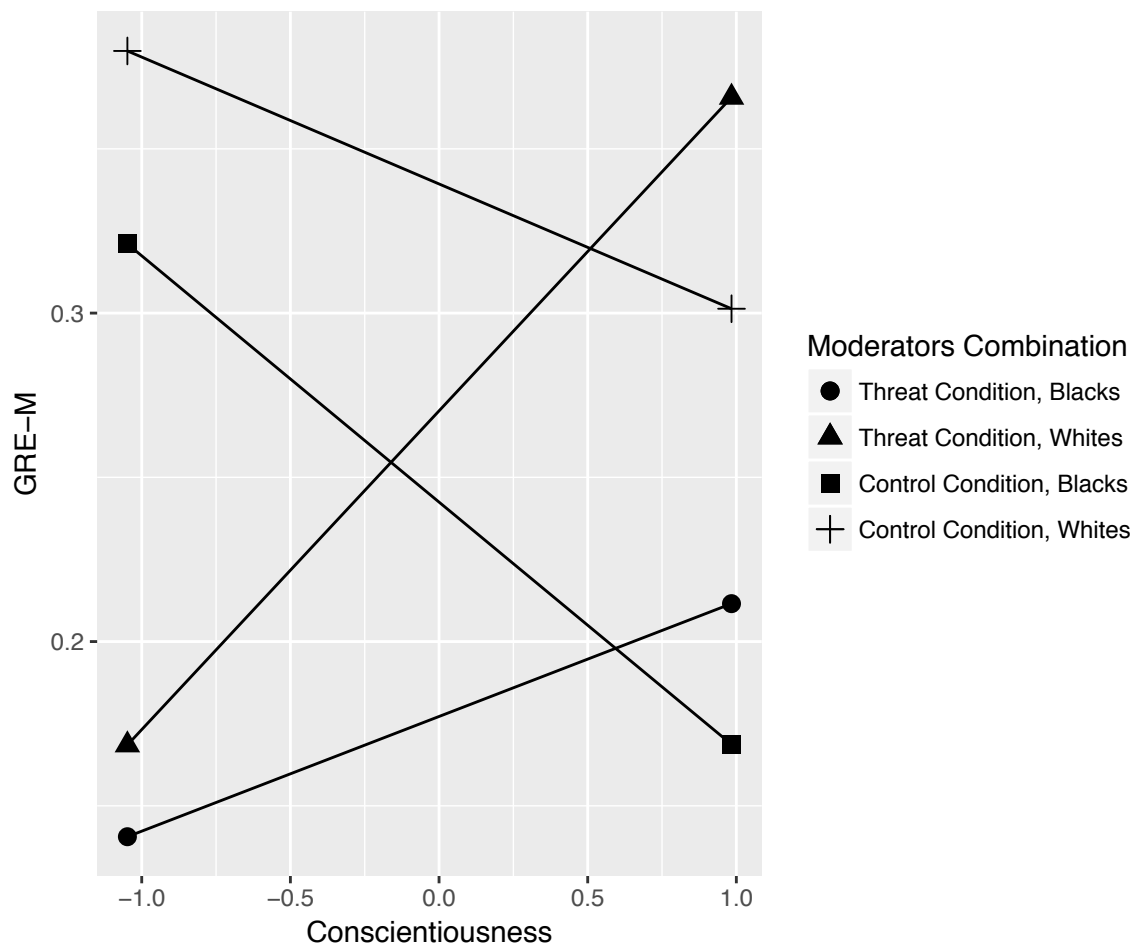


Figure 5. Simple Slopes of Conscientiousness moderating Stereotype Threat on GRE-M.

Looking at the effect of trait conscientiousness moderating the effect of stereotype threat on standardized test performance on the verbal GRE, regression analyses revealed no significant effects (see Table 6 and Figure 6). The moderation model was also not significant, $F(7, 23) = .413$, Multiple $R^2 = .12$, $p > .05$.

Table 6. *Conscientiousness, Threat and Race Effects on GRE-V Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.021	0.024	0.895	0.171	0.379
condition	-0.019	0.024	-0.793	-0.158	0.435
conscientiousness	-0.006	0.029	-0.198	-0.039	0.844
Model 2					
race	0.013	0.026	0.514	0.109	0.612
condition	-0.023	0.026	-0.865	-0.185	0.396
conscientiousness	-0.023	0.055	-0.419	-0.157	0.679
race x condition	-0.009	0.026	-0.361	-0.075	0.722
condition x conscientiousness	-0.004	0.055	-0.073	-0.026	0.943
race x mindset	-0.013	0.055	-0.237	-0.089	0.815
race x condition x conscientiousness	-0.029	0.055	-0.529	-0.200	0.602

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

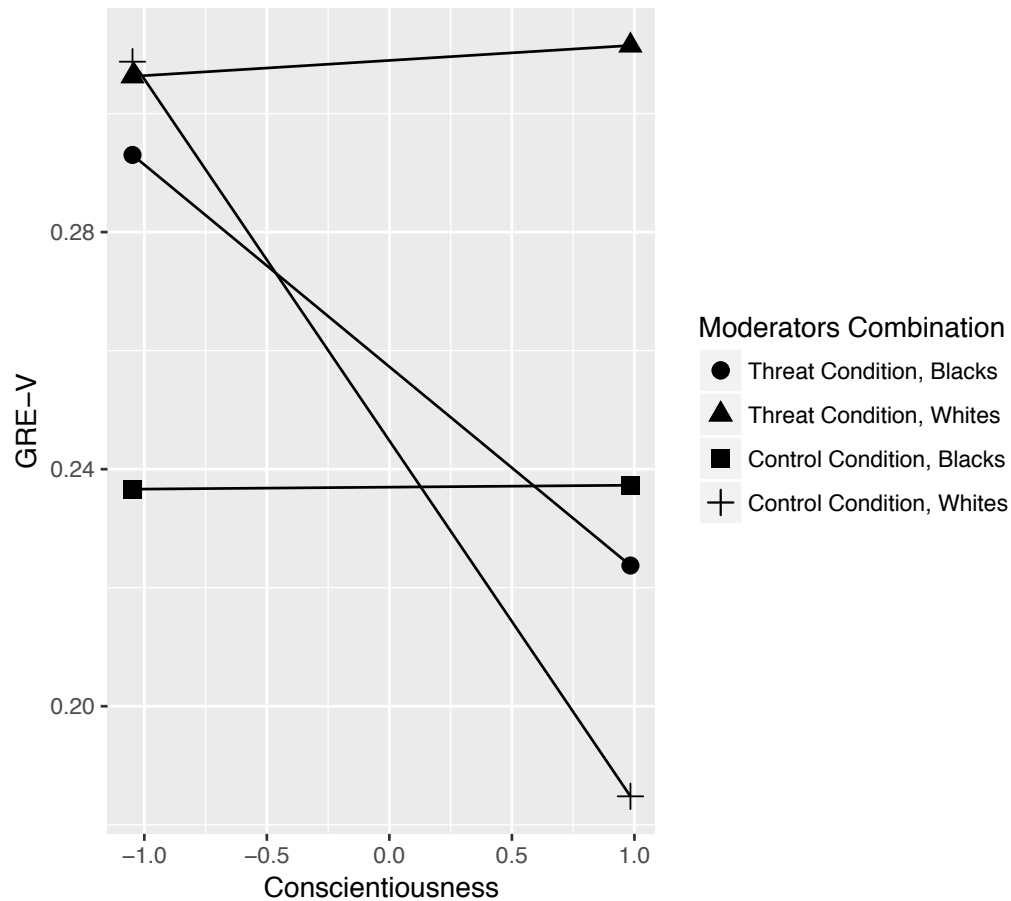


Figure 6. Simple Slopes of Conscientiousness moderating Stereotype Threat on GRE-V.

Verbal Capacity & Intelligence Condition.

Grit. Looking at the effect of trait grittiness moderating the effect of stereotype threat on WMC via RSPAN, regression analyses revealed no significant main effects or interactions (see Table 7 and Figure 7). The moderation model was also not significant, $F(7, 26) = .745$, Multiple $R^2 = .17, p > .05$.

Table 7. *Grit, Threat and Race Effects on RSPAN Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	-0.298	2.921	-0.102	-0.019	0.919
condition	-0.711	2.618	-0.272	-0.051	0.788
grit	0.246	5.613	0.044	0.008	0.965
Model 2					
race	-2.995	3.460	-0.866	-0.192	0.395
condition	-0.728	3.460	-0.210	-0.052	0.835
grit	-2.158	11.885	-0.182	-0.070	0.857
race x condition	-0.697	3.460	-0.201	-0.050	0.842
condition x grit	12.036	11.885	1.013	0.401	0.321
race x grit	6.301	11.885	0.530	0.211	0.601
race x condition x grit	-16.648	11.885	-1.401	-0.557	0.173

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

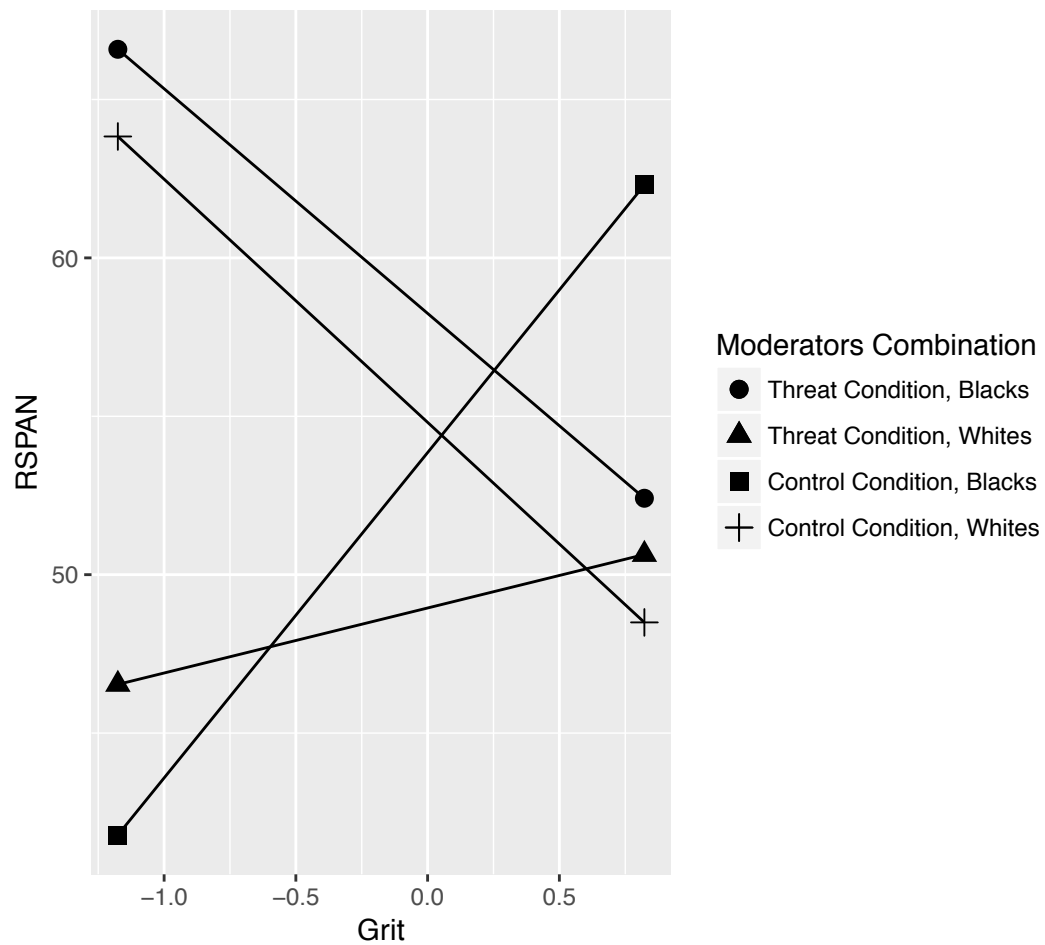


Figure 7. Simple Slopes of Grit moderating Stereotype Threat on RSPAN.

Looking at the effect of trait grittiness moderating the effect of stereotype threat on verbal GRE, regression analyses revealed only the main effect of race was marginally significant such that Whites outperformed Blacks (see Table 8 and Figure 8). There were no other significant effects. The moderation model was also not significant, $F(7, 26) = .707$, Multiple $R^2 = .16$, $p > .05$.

Table 8. *Grit, Threat and Race Effects on GRE-V Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	-0.298	2.921	-0.102	-0.019	0.919
condition	-0.711	2.618	-0.272	-0.051	0.788
grit	0.246	5.613	0.044	0.008	0.965
Model 2					
race	0.079	0.047	1.690†	0.377	0.103
condition	-0.028	0.047	-0.596	-0.148	0.557
grit	0.178	0.161	1.107	0.430	0.278
race x condition	0.031	0.047	0.657	0.165	0.517
condition x grit	0.120	0.161	0.744	0.296	0.464
race x grit	-0.094	0.161	-0.583	-0.232	0.565
race x condition x grit	-0.093	0.161	-0.578	-0.231	0.568

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

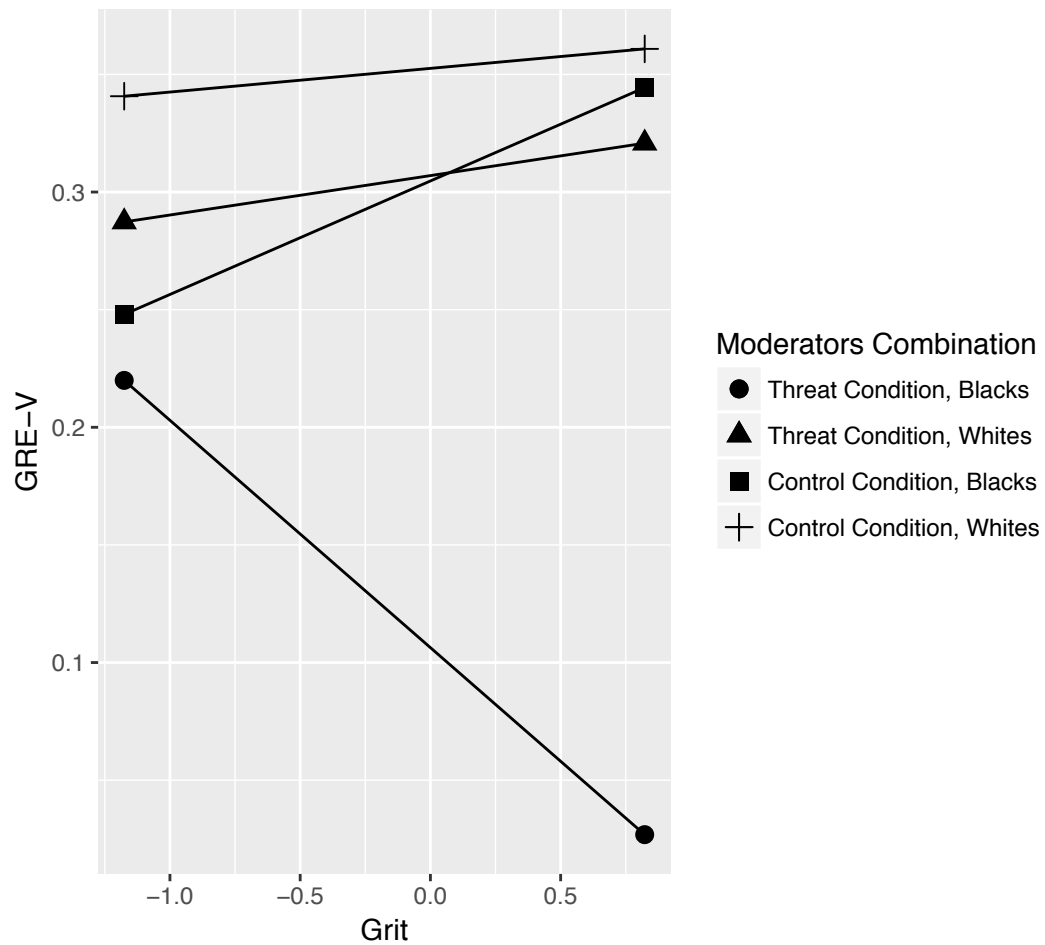


Figure 8. Simple Slopes of Grit moderating Stereotype Threat on GRE-V.

Looking at the effect of trait grittiness moderating the effect of stereotype threat on the math GRE, regression analyses revealed no significant main effects or interaction effects (see Table 9 and Figure 9). The moderation model was also not significant, $F(7, 26) = .426$, Multiple $R^2 = .103$, $p > .05$.

Table 9. *Grit, Threat and Race Effects on GRE-M Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.016	0.028	0.571	0.106	0.573
condition	0.013	0.025	0.541	0.100	0.593
grit	-0.012	0.053	-0.223	-0.040	0.825
Model 2					
race	0.015	0.035	0.435	0.100	0.667
condition	-0.008	0.035	-0.220	-0.057	0.828
grit	-0.110	0.119	-0.929	-0.373	0.362
race x condition	0.032	0.035	0.940	0.244	0.356
condition x grit	-0.006	0.119	-0.052	-0.021	0.959
race x grit	0.115	0.119	0.971	0.400	0.340
race x condition x grit	0.011	0.119	0.096	0.040	0.924

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

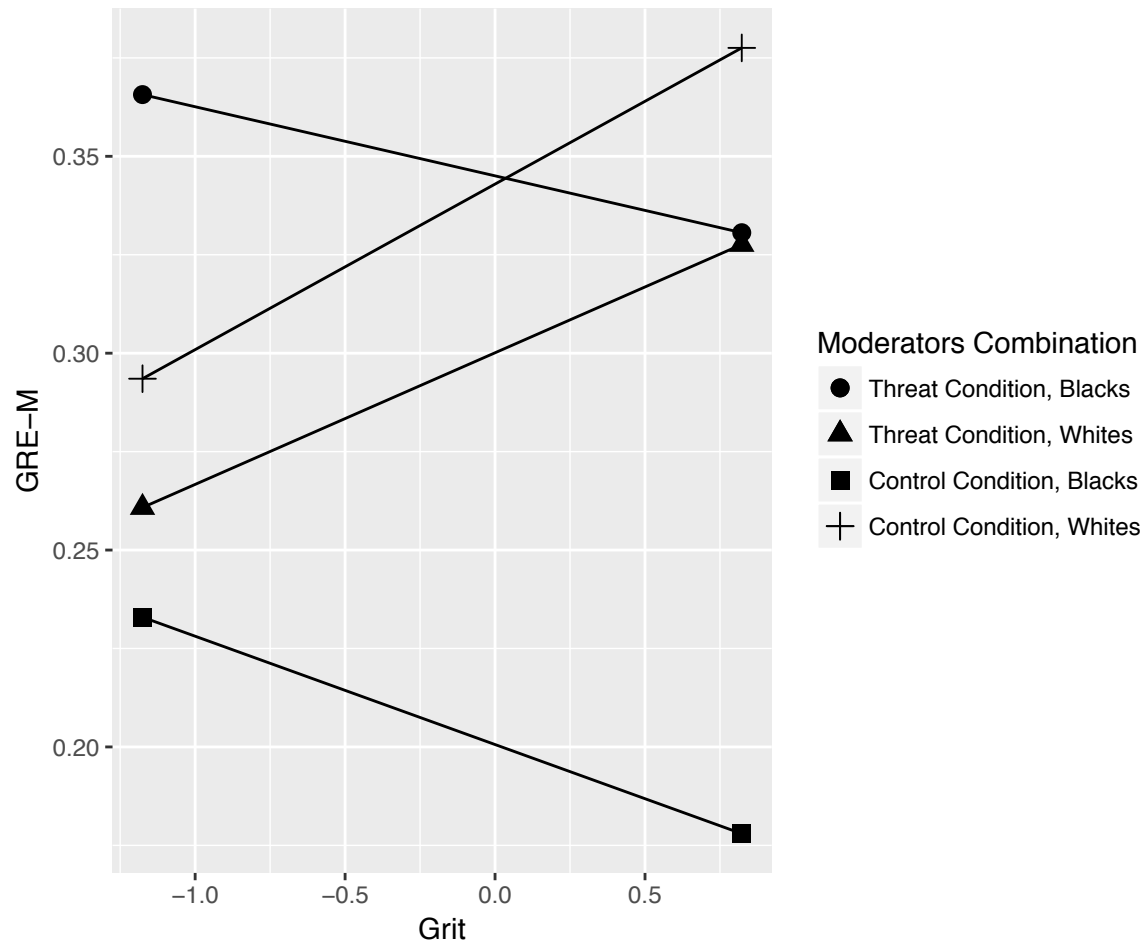


Figure 9. Simple Slopes of Grit moderating Stereotype Threat on GRE-M.

Conscientiousness. Looking at the effect of trait conscientiousness moderating the effect of stereotype threat on WMC via RSPAN, regression analyses revealed a main effect of conscientiousness such that higher scores on conscientiousness were associated with a higher predicted score on RSPAN. There were no other significant main effects or interactions (reported in Table 10 and shown in Figure 10). The moderation model was also not significant, $F(7, 26) = 1.86$, Multiple $R^2 = .33$, $p > .05$.

Table 10. *Conscientiousness, Threat and Race Effects on RSPAN Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	-0.735	2.495	-0.294	-0.047	0.770
condition	-0.471	2.235	-0.211	-0.034	0.834
conscientiousness	14.827	4.446	3.335**	0.520	0.002
Model 2					
race	-2.986	3.081	-0.969	-0.192	0.341
condition	2.020	3.081	0.656	0.145	0.518
conscientiousness	9.718	6.316	1.539	0.341	0.136
race x condition	-3.282	3.081	-1.065	-0.237	0.297
condition x conscientiousness	-6.480	6.316	-1.026	-0.232	0.314
race x mindset	5.831	6.316	0.923	0.207	0.364
race x condition x conscientiousness	7.407	6.316	1.173	0.264	0.252

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

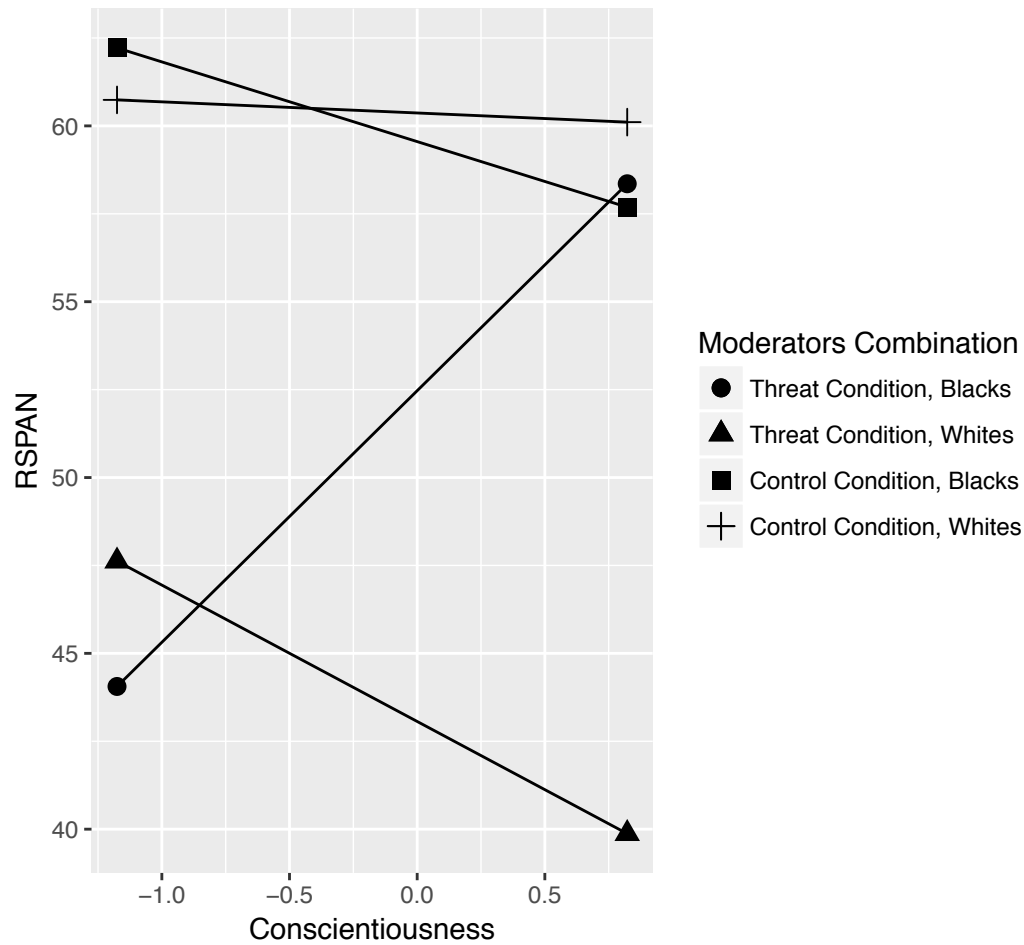


Figure 10. Simple Slopes of Conscientiousness moderating Stereotype Threat on RSPAN.

Looking at the effect of trait conscientiousness moderating the effect of stereotype threat on verbal GRE, regression analyses revealed no significant main effects or interactions (reported in Table 11 and shown in Figure 11). The moderation model was also not significant, $F(7, 26) = .645$, Multiple $R^2 = .15$, $p > .05$.

Table 11. *Conscientiousness, Threat and Race Effects on GRE-V Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.062	0.037	1.658†	0.293	0.108
condition	-0.007	0.033	-0.198	-0.035	0.845
conscientiousness	-0.079	0.066	-1.196	-0.205	0.241
Model 2					
race	0.057	0.047	1.209	0.271	0.237
condition	-0.006	0.047	-0.121	-0.030	0.905
conscientiousness	-0.121	0.096	-1.251	-0.313	0.222
race x condition	0.011	0.047	0.233	0.059	0.817
condition x conscientiousness	-0.033	0.096	-0.337	-0.086	0.739
race x mindset	0.081	0.096	0.844	0.214	0.406
race x condition x conscientiousness	0.005	0.096	0.056	0.014	0.956

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

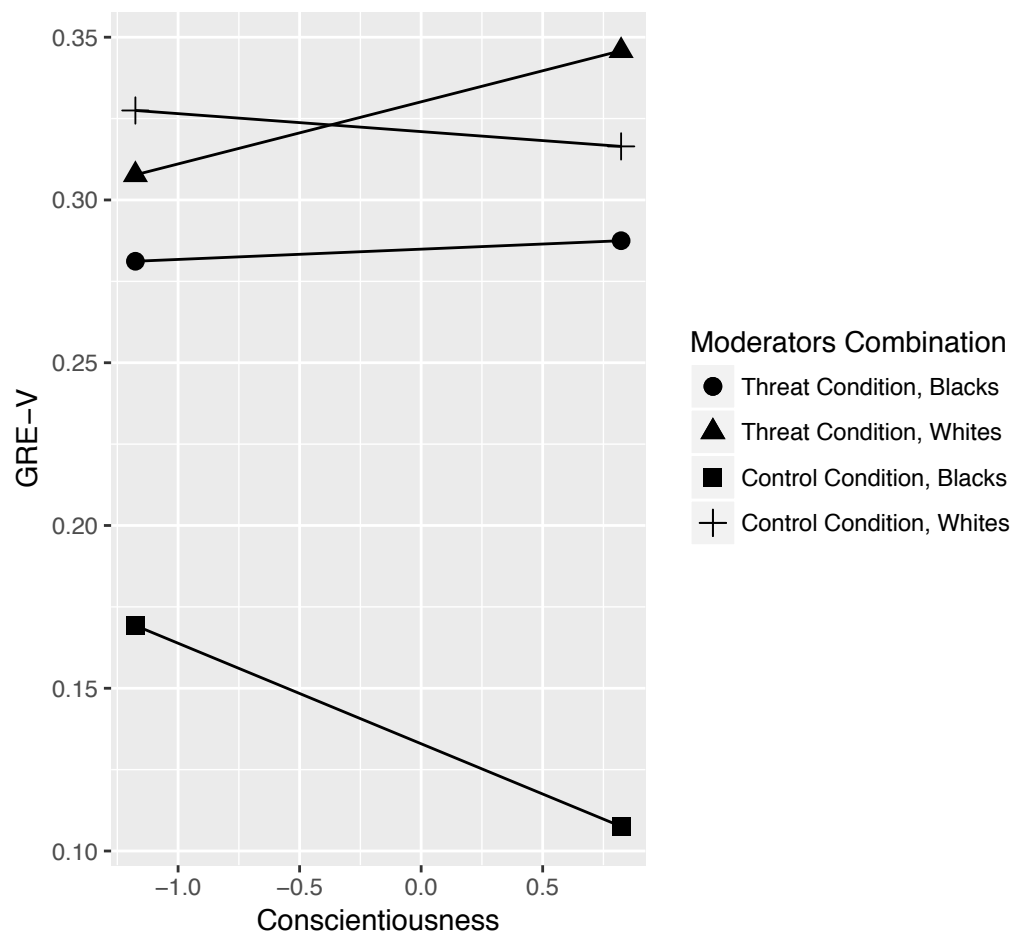


Figure 11. Simple Slopes of Conscientiousness moderating Stereotype Threat on GRE-V.

Looking at the effect of trait conscientiousness moderating the effect of stereotype threat on math GRE, regression analyses revealed no significant main effects or interactions (reported in Table 12 and shown in Figure 12). The moderation model was also not significant, $F(7, 26) = .77$, Multiple $R^2 = .17$, $p > .05$.

Table 12. *Conscientiousness, Threat and Race Effects on GRE-M Full Model Estimates*

	<i>b</i>	<i>SE</i>	<i>t</i>	<i>beta</i>	<i>p</i>
Model 1					
race	0.017	0.028	0.604	0.112	0.550
condition	0.013	0.025	0.538	0.099	0.595
conscientiousness	-0.020	0.049	-0.407	-0.073	0.687
Model 2					
race	0.033	0.033	0.992	0.219	0.331
condition	-0.005	0.033	-0.136	-0.034	0.893
conscientiousness	0.036	0.068	0.535	0.132	0.597
race x condition	0.042	0.033	1.279	0.318	0.212
condition x conscientiousness	-0.020	0.068	-0.294	-0.074	0.771
race x mindset	-0.019	0.068	-0.286	-0.072	0.777
race x condition x conscientiousness	-0.084	0.068	-1.246	-0.313	0.224

Note: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

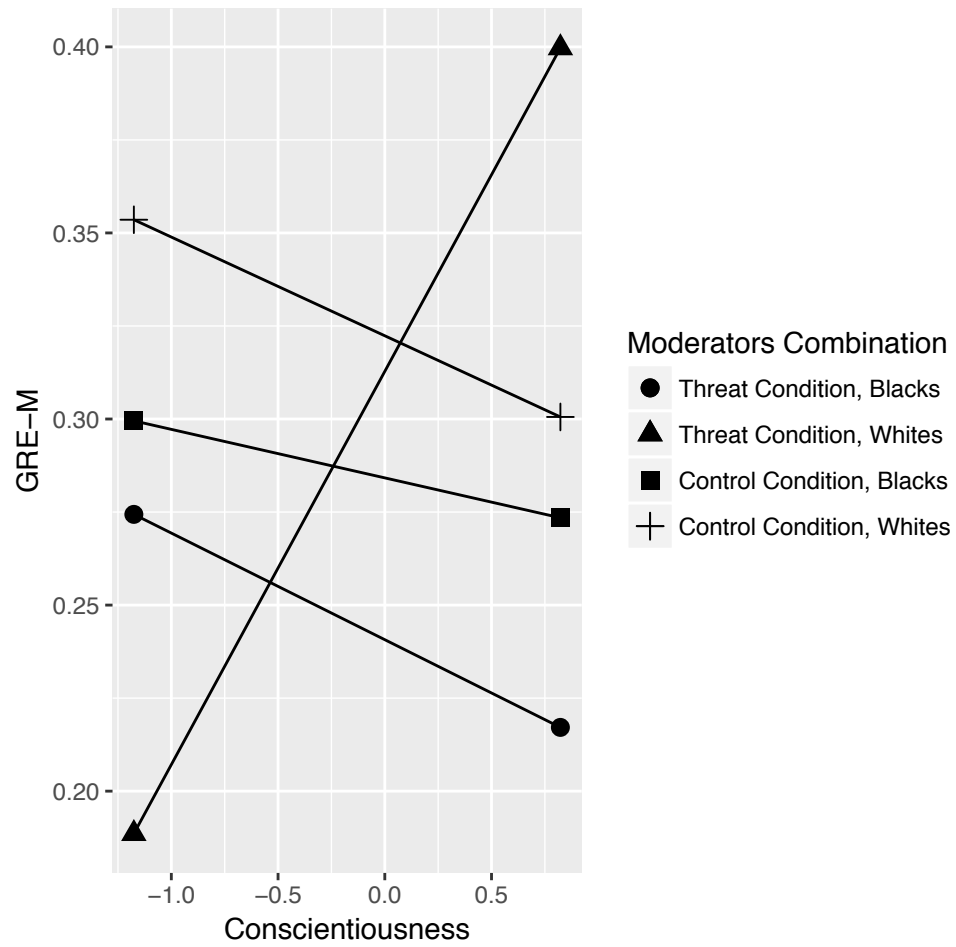


Figure 12. Simple Slopes of Conscientiousness moderating Stereotype Threat on GRE-M.

Chapter 4

Exploration of other non-cognitive factors

Next, A series of hierarchical regression analyses were conducted to explore whether individual differences in trait grittiness and conscientiousness moderate stereotype threat effects for gender on performance. Under gender stereotype threat, women's performance was expected to be negatively impacted compared to the control condition. Based on the lack of evidence that the mindset manipulation moderated stereotype threat effects in Experiment 3, the effect of mindset was collapsed across these analyses. Moreover, it is important to note that because only women were recruited in the study, in each of the regression analyses below the outcome measures of WMC and standardized test performance were regressed on condition ("Control": 0, "Threat ": 1), trait grittiness (mean-centered), or trait conscientiousness (mean-centered) and their interactions.

Grit. In step one, looking at performance on the OPSAN, the effect of grit was not significant, $b = 2.10$, $t(73) = 1.54$, $p > .05$. The effect of condition was marginally significant, $b = 3.25$, $t(73) = 1.97$, $p = .05$, indicating that predicted scores on the OPSAN were higher under threat. The model, $F(2, 73) = 3.65$, Multiple $R^2 = .091$, $p < .05$ was also significant.

In step two predicting OPSAN, the effect of grit, $b = 2.1$, $t(72) = 1.02$, $p > .05$ was not significant. The effect of condition was marginally significant, $b = 3.24$, $t(72) = 1.95$, $p = .06$, indicating that predicted scores on OPSAN were higher under threat. The interaction was not significant, $b = .012$, $t(72) = .004$, $p > .05$. The model approached significance, $F(3, 72) = 2.4$, Multiple $R^2 = .091$, $p = .08$. Taken together, these results reveal a lack of support for trait grittiness moderating the effect of threat on OPSAN (see Figure 13).

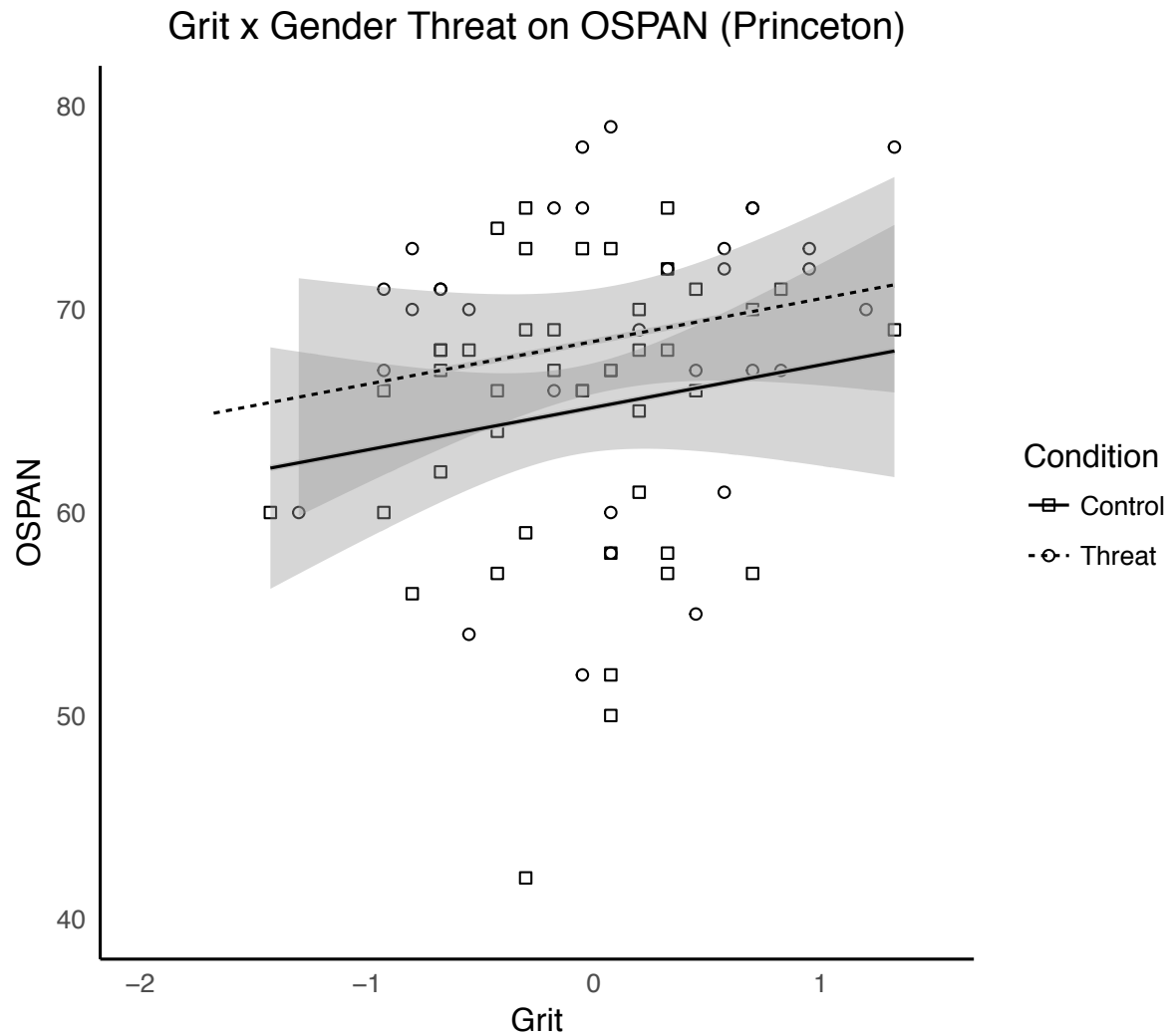


Figure 13. Grit Moderation Model for Threat on OSPAN.

Next, a series of hierarchical regression analyses were conducted in order to test whether grit moderated gender stereotype threat on math standardized test performance.

In step one looking at the effects on the math GRE, the effect of condition approached significance, $b = .078$, $t(76) = 1.86$, $p = .07$, indicating higher predicted scores on math GRE under threat. The effect of grit was not significant, $b = -.0030$, $t(76) = -.088$, $p < .05$. The model, $F(2, 76) = 1.74$, Multiple $R^2 = .044$, $p > .05$ was not significant.

In step two, the effect of grit, $b = -.032$, $t(75) = .55$, $p > .05$ was not significant. The effect of condition approached significance, $b = .080$, $t(75) = 1.90$, $p = .06$, indicating that the threat effect predicted higher scores on the math GRE. The model, $F(3, 75) = 1.31$, Multiple $R^2 = .049$, $p > .05$ and the interaction were not significant, $b = .047$, $t(75) = .686$, $p > .05$, indicating a lack of evidence for trait grittiness moderating the effect of threat on the math GRE (see Figure 14).

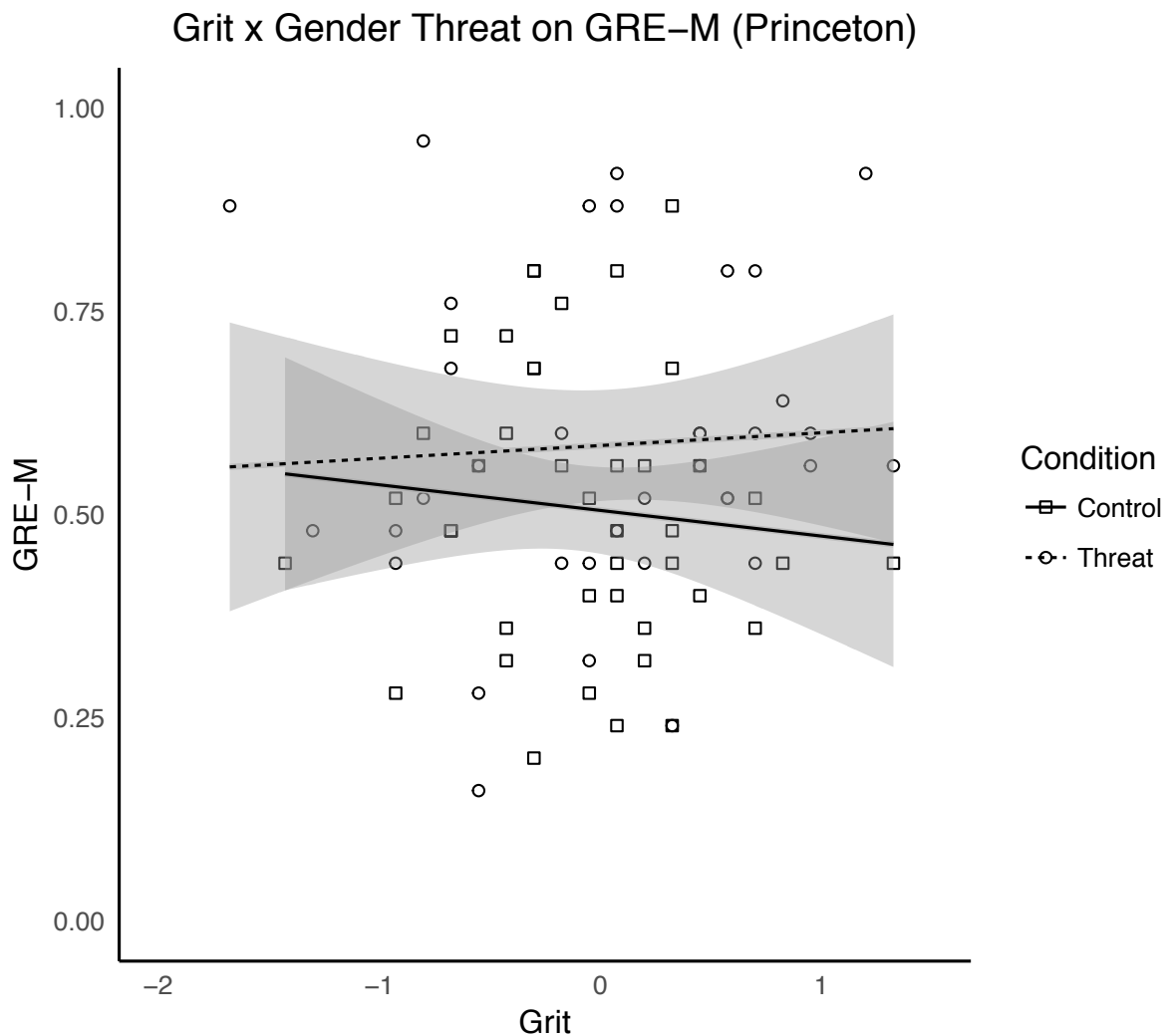


Figure 14. Grit Moderation Model for Threat on GRE-M.

Conscientiousness. Next, a series of hierarchical regression analyses were conducted in order to test whether trait conscientiousness moderated gender stereotype threat on WMC via OSPAN¹⁰.

In step one, looking at performance on the OPSAN, the effect of conscientiousness was significant, $b = 2.43$, $t(68) = 2.05$, $p < .05$ indicating higher conscientiousness predicted higher scores on the OSPAN. The effect of condition was marginally significant, $b = 3.43$, $t(68) = 2.03$, $p = .05$, indicating that predicted scores on the OSPAN were higher under threat. The model, $F(2, 68) = 4.54$, Multiple $R^2 = .12$, $p < .05$ was also significant.

In step two predicting OSPAN, the effect of conscientiousness was not significant, $b = 1.89$, $t(67) = 1.23$, $p > .05$. The effect of condition was marginally significant, $b = 3.47$, $t(67) = 2.03$, $p = .05$, indicating that predicted scores on the OSPAN were higher under threat. The interaction was not significant, $b = 1.06$, $t(67) = .445$, $p > .05$. The model, $F(2, 67) = 3.06$, Multiple $R^2 = .12$, $p < .05$ was also significant. Taken together, these results reveal a lack of support for trait conscientiousness moderating the effect of threat on OSPAN (see Figure 15).

¹⁰ A total of 10 subjects had conscientiousness scales either missing or incomplete.

Conscientiousness x Gender Threat on OSPAN (Princeton)

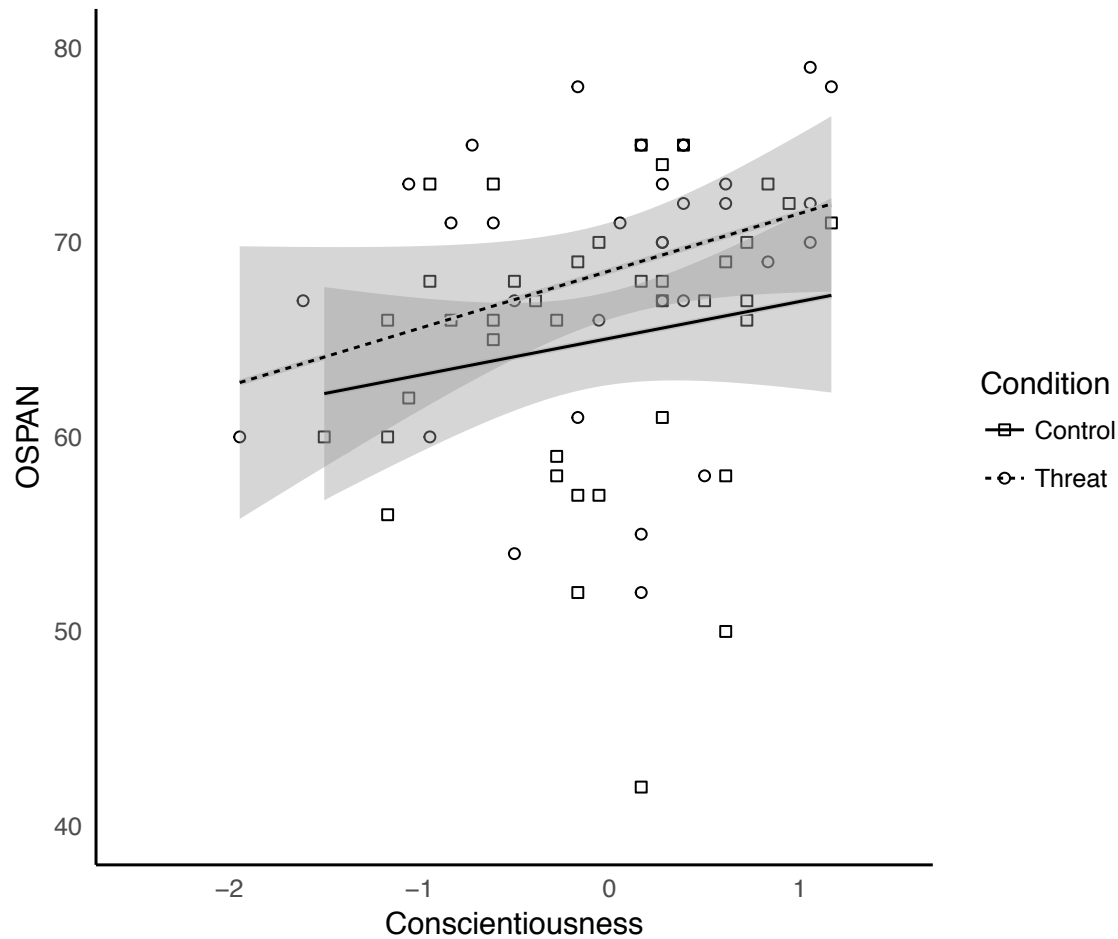


Figure 15. Conscientiousness Moderation Model for Threat on OSPAN.

Next, a series of hierarchical regression analyses were conducted in order to test whether trait conscientiousness moderated gender stereotype threat on the math GRE.

In step one, looking at performance on the math GRE, the effect of conscientiousness was not significant, $b = .016$, $t(67) = .59$, $p > .05$. The effect of condition was marginally significant, $b = .082$, $t(67) = 1.84$, $p = .07$, indicating that predicted scores on the math GRE were higher under threat. The model, $F(2, 67) = 1.94$, Multiple $R^2 = .055$, $p > .05$ was not significant.

In step two predicting math GRE, the effect of conscientiousness was not significant, $b = .033$, $t(66) = .45$, $p > .05$. The effect of condition was marginally significant, $b = .081$, $t(66) = 1.81$, $p = .08$, indicating that predicted scores on the math GRE were higher under threat. The interaction was not significant, $b = -.034$, $t(66) = -.545$, $p > .05$. The model, $F(3, 66) = 1.38$, Multiple $R^2 = .059$, $p > .05$ was not significant. Taken together, these results reveal a lack of support for trait conscientiousness moderating the effect of threat on math GRE (see Figure 16).

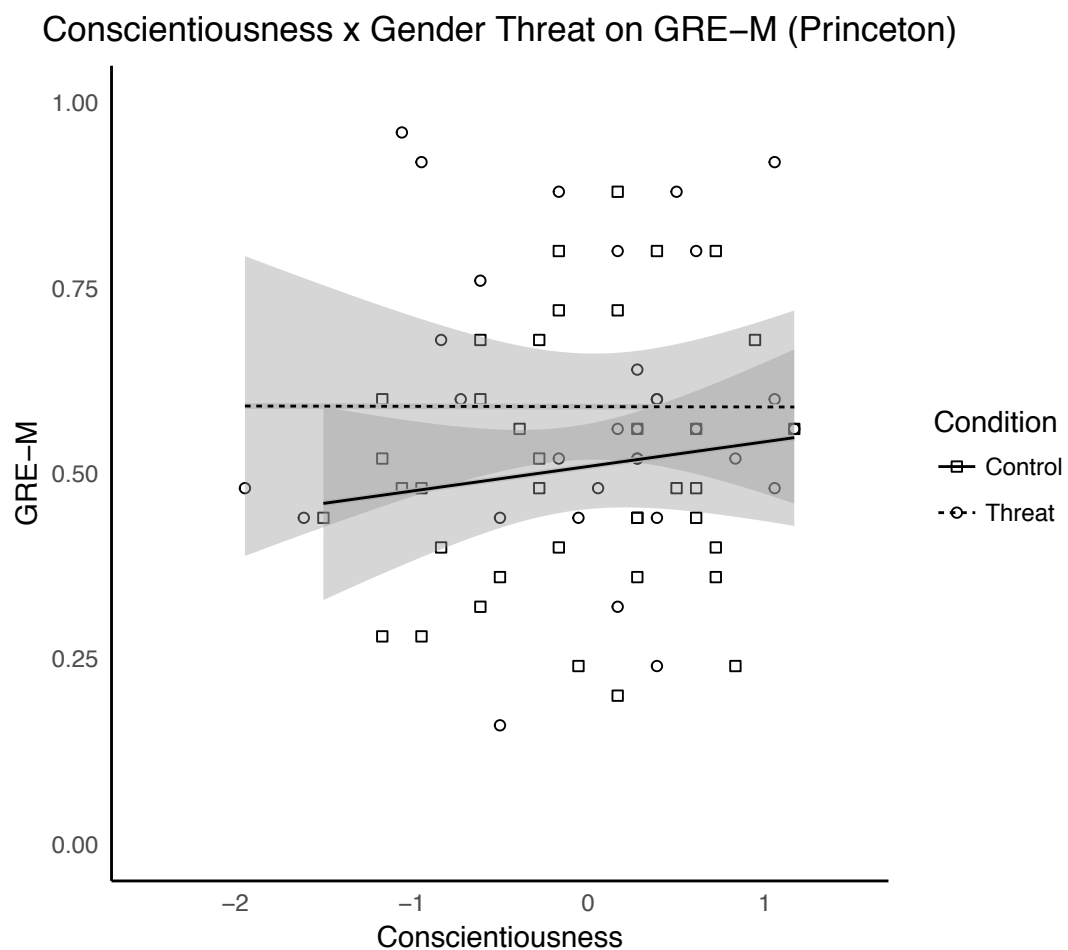


Figure 16. Conscientiousness Moderation Model for Threat on GRE-M.

General Discussion

The results of these exploratory analyses did not reveal any significant influence of the non-cognitive factors of grit or conscientiousness under stereotype threat. Instead, these results suggest that the role of non-cognitive factors during contexts of stereotype threat may be less important than expected. Also, based on the trends in the data it is unclear whether the non-cognitive factors of grit and conscientiousness are actually independent constructs. In both Experiment 2 and Experiment 3 trends in the data revealed strong positive correlations between the grit and conscientiousness measures. Based on this finding it is unclear whether these variables should be treated as independent constructs. Moreover, this notion that grit is not truly independent of the personality variable trait conscientiousness has been suggested in previous work (see Crede, Tynan & Harris, 2017). Although Duckworth (2016) asserts that grit is a distinct construct from personality measures the debate appears to be on-going in the literature (see Duckworth, Peterson, Matthews & Kelly, 2007; Duckworth & Quinn, 2009). In the present exploratory analysis, these measures were analyzed separately and in this way treated as independent but with caution and with the knowledge that some have argued these measures are highly related—bivariate trends in our data revealed moderate to strong direct relationships of these measures in Experiments 2 and 3.

Overall, it was difficult to detect a consistent pattern in terms of the effects that these non-cognitive trait measures have during contexts of stereotype threat. Future studies will need to recruit larger and more powerful samples in order to detect possible effects based on either of these factors. In addition, future studies should also consider the possibility that these constructs are indeed related when thinking about the role that one or both might play during contexts of stereotype threat or in general.

