

**THE IN-BETWEEN: A META-ANALYTIC INVESTIGATION OF STEREOTYPE
THREAT EFFECTS ON MEDIATORS OF THE STEREOTYPE THREAT-
PERFORMANCE RELATIONSHIP**

A Dissertation

by

AMBER LEOLA SMITTICK

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Winfred Arthur, Jr.
Committee Members,	George B. Cunningham
	Kathi N. Miner
	Stephanie C. Payne
Head of Department,	Heather Lench

Graduation: December 2019

Major Subject: Psychology

Copyright 2019 Amber Leola Smittick

ABSTRACT

Stereotype threat is an important phenomenon in educational and organizational settings as it results in lower test scores for minority and women test takers on high stakes tests used for selection, promotional, and educational attainment purposes (Nguyen & Ryan, 2008; Steele, 1997). Although a great deal of research attention has been devoted to investigating stereotype threat effects on test performance, less is known about the processes that underlie these effects. To better understand the underlying processes of stereotype threat, the current study provides a qualitative and quantitative review of the effects of stereotype threat on mediators of the stereotype-threat performance relationship. The dissertation provides and tests a model of stereotype threat's effect on cognitive test performance through several proposed affective, cognitive, and motivational mediators.

The results of the meta-analysis show that most of mediator effect sizes followed a pattern that was consistent with the hypotheses. Specifically, stereotype threatened individuals had higher levels of anxiety and physiological arousal, more distracting (negative) thoughts, and demonstrated increased levels of effort than non-stereotype threatened individuals. They also had lower working memory capacity and decreased performance expectations. However, although the general pattern was in the hypothesized direction across the mediators many of the effect sizes could not meaningfully be interpreted because of large confidence intervals that included zero. Additionally, when looking across the within- and between group design the results were not consistent.

Altogether although many of the study hypotheses were not supported, the results for anxiety and effort do begin to provide some clarity in understanding the processes through which stereotype threat impacts test performance. Furthermore, the lack of an interpretable pattern of

effects across the within- and between-group mediator results highlights a need to continue to move forward in understanding the processes that underlie stereotype threat effects in a more strategic manner.

DEDICATION

This dissertation is dedicated to my parents, Lisa and Joery Smittick. There are no words that can express how extremely grateful I am to you all, so I just say thank you and I love you!

ACKNOWLEDGEMENTS

Mahatma Gandhi said, “Every worthwhile accomplishment, big or little, has its stages of drudgery and triumph: a beginning, a struggle, and a victory,” I am so thankful for the people who not only celebrated with me in triumphs and victories, but who were also there for me through the drudgery and struggles. This process spanned over five years and included new jobs, new states, new houses, and even a marriage and baby and there is no way I would have made it through all these major life events without the support of my family friends.

First and foremost, I have to acknowledge that I would not have even had the opportunity to be in the place without the grace of God. He led me to a graduate school in Texas where I was able to be close to my family and loved me and always brought along the right encouragement when I needed it.

I am thankful to many of the faculty at Texas A&M for being patient with me and giving me the chance to finish this journey. Particularly, my dissertation chair, Winfred Arthur who was always there to support me and provide the feedback and critique, albeit how tough it was, I needed to finish. To my dissertation committee, thank you for hanging in there with me, I know it was probably hard to remember who I was and what the topic was but you all eagerly responded to my emails and were always available to me when I needed you.

To my graduate school friends and cheerleaders. There is no way I would have made it through it all without the breaks in the stress you all provided. There are too many stories and thank yous to include in this section, so we’ll just save it for an epic celebration party.

To my Hogan Assessments colleagues, thank for you supporting me in this process. Allowing me the opportunity to work on my dissertation while I was working and just being genuinely awesome people. I am so lucky to work with so many intelligent and fun individuals!

To my friends who had no idea what a dissertation was or why I was taking so long to complete, thank you for your patiently putting up with my complaints and nerdy talk even though you had no idea what it was about.

To my wonderful and fabulous siblings. Shannon and Joye both are truly fantastic. Thank you for putting up this whiny middle child and loving and supporting me. Roderick, I am not sure you even knew what I was working on, but thank you for being a cool younger brother.

Also, just a random shoutout to two people who will probably never see this. First, to Claude Steele – you put stereotype threat on the map and it was something that I was also interested in and I considered you a force and role model in psychology. When I reached out looking for data for this meta-analysis, you immediately responded to me a lowly graduate student to help get me the information that I needed and though that seems small it meant a lot. Second, thank you Michelle Obama. One, just because she is an awesome and inspiring woman to look up to and has the arms to whose mine are always compared. But, also because I listened to her book *Becoming* as I finished up and her words and listening to her talk about her life was the encouragement that helped me to *finally* wrap this thing up.

Carter, you were not around for most of this process, yet you were still such an inspiration. I love you!

And last to my husband, Brady. Some days during this process I really didn't like you, especially when you quoted random lines from *Whistling Vivaldi* to get me motivated. However, through it all you were the one person who never stopped believing in me and you were my biggest cheerleader. You were willing to do anything for me, including watching a fussy newborn so I could write. I love you and am so thankful for you!

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Winfred Arthur Jr. (advisor) Professors Stephanie Payne and Kathi Miner of the Department of Psychological and Brain Sciences and Professor George Cunningham of the Department of Health and Kinesiology.

All work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by the Graduate Diversity Fellowship) from Texas A&M University.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION.....	iv
ACKNOWLEDGEMENTS	v
CONTRIBUTORS AND FUNDING SOURCES	vii
TABLE OF CONTENTS	viii
LIST OF FIGURES	x
LIST OF TABLES	xii
INTRODUCTION.....	1
STEREOTYPE THREAT	9
Characteristics of Stereotype Threat	9
Moderators of Stereotype Threat	14
MEDIATORS OF STEREOTYPE THREAT	19
A Model of Stereotype Threat Effects	21
Affective Reactions	24
Cognitive Processes	29
Motivational Process	37
Moderators of the Stereotype Threat Mediators	45
THE PRESENT STUDY	51
METHOD	55
Inclusion Criteria	56
Cumulating Results.....	63
Coding Studies.....	63
Coding Sheet and Coding Manual.....	65
Description of Study Variables.....	66
Summary of the Meta-Analytic Data Set.....	68
Meta-Analytic Procedure.....	69
Correction for Measurement Unreliability	70
Computing Effect Sizes	71
Meta-Analytic Computations and Moderator Analyses	71

Subgroup Moderator Analysis	72
Publication Bias	73
RESULTS	75
Affective Mediator Results	79
Cognitive Processes Mediator Results.....	80
Motivational Processes Mediator Results.....	85
Moderator Hypotheses Analyses	88
Summary of Results	102
DISCUSSION AND CONCLUSION	108
Review of Mediator Results.....	111
Practical Implications	114
Limitations	116
Future Research.....	120
Conclusion.....	123
REFERENCES	125
APPENDIX A Stereotype Threat Mediators Coding Sheet	149
APPENDIX B Table of Effect Sizes for Each Mediator.....	153
APPENDIX C Study Effect Sizes Forest Plots by Analysis Level	160

LIST OF FIGURES

Figure 1. Figure 1 from Schmader, Johns, and Forbes (2008), “An integrated process model of stereotype threat effects on performance”.	22
Figure 2. A model of stereotype threat effects on cognitive test performance for stereotyped groups.....	23
Figure 3. A step-by-step model to test the mediators of stereotype threat	52
Figure 4. Flowchart outlining the literature search and screening process for the included studies.	56
Figure 5. Comparison of between-group and within-group designs.....	59
Figure 6. Overall mediator effect sizes	103
Figure 7. Within-group effect sizes by stereotype group.....	106
Figure 8. Between-group effect sizes by stereotype group.....	107
Figure 9. Aggregate within- and between-group design effect sizes.....	110
Figure C1. Forest plot for anxiety within-group effect sizes	160
Figure C2. Forest plot for anxiety between-group results	161
Figure C3. Forest plot for physiological arousal within-group effect sizes.....	162
Figure C4 . Forest plot for physiological arousal between-group effect sizes.....	163
Figure C5. Forest plot for working memory within-group effect sizes	164
Figure C6. Forest plot for working memory between-group effect sizes.....	165
Figure C7. Forest plot for distracting (negative) thoughts within-group effect sizes.....	166
Figure C8. Forest plot for distracting (negative) thoughts between-group effect sizes.....	167
Figure C9. Forest plot for effort within-group effect sizes.....	168
Figure C10. Forest plot for effort between-group effect sizes.....	169

Figure C11. Forest plot for performance avoidance within-group effect sizes	170
Figure C12. Forest plot for performance avoidance between-group effect sizes	171
Figure C13. Forest plot for performance expectations within-group effect sizes.....	172
Figure C14. Forest plot for performance expectations between-group effect sizes	173

LIST OF TABLES

Table 1 Summary of Hypotheses	50
Table 2 Inclusion Criteria Table	57
Table 3 Mediator Description	61
Table 4 Mediator Frequency Counts.....	69
Table 5 Artifact Distribution for Mediator Variables	70
Table 6 Publication Bias Results	74
Table 7 Overall Within-Group Meta-Analytic Results.....	77
Table 8 Overall Between-Group Meta-Analytic Results.....	78
Table 9 Overall Main Hypotheses Within-Group Moderator Results	83
Table 10 Overall Main Hypotheses Between-Group Moderator Results	84
Table 11 Within-Group Meta-Analytic Results by Stereotype Threatened Group	89
Table 12 Between-Group Meta-Analytic Results by Stereotype Threatened Group	90
Table 13 Gender-Based Stereotype Threat Within-Group Meta-Analytic Results by Stereotype Activation Level.....	95
Table 14 Gender-Based Stereotype Threat Between-Group Meta-Analytic Results by Stereotype Activation Level.....	96
Table 15 Race-Based Stereotype Threat Within-Group Meta-Analytic Results by Stereotype Activation Level.....	98
Table 16 Race-Based Stereotype Threat Between-Group Meta-Analytic Results by Stereotype Activation Level.....	99
Table 17 Age-Based Stereotype Threat Within-Group Meta-Analytic Results by Stereotype Activation Level.....	100

Table 18 Age-Based Stereotype Threat Between-Group Meta-Analytic Results by Stereotype

Activation Level..... 101

Table 19 Summary of Results by Hypothesis..... 104

INTRODUCTION

There is a well-documented gap on cognitive test scores between racial minority groups and Whites (Roth, Bevier, Bobko, Switzer, & Tyler, 2001). In 1995, Steele and Aronson proposed the concept of stereotype threat as one phenomenon responsible for the lowered performance of minority group members on tests of cognitive ability. *Stereotype threat* is a form of identity threat experienced by individuals when they feel their behavior could confirm a negative stereotype about a group to which they belong (Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002). The main hypothesis in stereotype threat research asserts that stereotyped individuals will perform more poorly on a task in a threatening situation than in a situation not linked to a negative stereotype. For example, there is a pervasive stereotype that African Americans are intellectually inferior to Whites (Plous & Williams, 1995). Thus, when African Americans are in situations where they are confronted by this stereotype, such as a cognitive ability test setting, African Americans may experience stereotype threat.

Stereotype threat has been proposed as a potentially important phenomenon in testing because it can have deleterious effects on the test performance of members of a number of different groups. Stereotype threat effects have been extensively studied in racial minorities (primarily African Americans) in the cognitive ability domain, and in women in the context of mathematical ability. In addition, stereotype threat has also been documented across multiple domains (e.g., cognitive, mathematical, and athletic ability) as well as in a variety of other groups (e.g., older adults, Asian Americans, and Whites). Despite the number of studies that have documented the negative effects of stereotype threat (e.g., Nguyen & Ryan, 2008; Walton & Cohen, 2003), many questions about the phenomenon remain unanswered; one of the foremost being, “What are the mechanisms underlying stereotype threat’s effect on test performance?”

The purpose of the current study is to examine the effects of proposed mediators in the stereotype threat-performance relationship.

In their seminal article, Steele and Aronson (1995) recognized that mediators of stereotype threat play a critical role in understanding the stereotype threat phenomenon. They asserted that it is not one, but many mechanisms that underlie stereotype threat effects. Furthermore, in their article Steele and Aronson (1995) hypothesized the existence of a number of possible mediators, and since that time the list has continued to expand and become further developed. In line with this recognition, a number of empirical studies have been conducted to investigate the impact of these different mediators (e.g., Ben-Zeev, Fein, & Inzlicht, 2005; Forbes, Schmader, & Allen, 2008; Johns, Inzlicht, & Schmader, 2008; Sawyer & Hollis-Sawyer, 2005). These studies have provided a wealth of evidence for the existence of a number of different mediators. However, these studies have provided limited guidance on the *relative importance* of each of these mediators and insight into whether they each equally contribute to stereotype threat effects. The lack of information on these relationships is due to a number of factors. First, it is highly improbable that one study would be able to measure the entire breadth of mediators and determine the unique effects of each mediator. Additionally, with the many effects linked to stereotype threat, it is also unlikely that only one mediator is operating across all stereotype threat situations. Accordingly, Steele and Aronson (1995) state, “depending on the situation, several of these processes may be involved simultaneously or in alteration” (p. 799). Since multiple mechanisms and processes may be responsible for stereotype threat, an important first step is to build a clear understanding of each of the mediators’ relationship with stereotype threat.

Consequently, the objective of the current study is to identify and discuss the mediators of the stereotype threat-performance relationship and more importantly use meta-analytic procedures to quantify the relationship between stereotype threat and said mediators as examined in the pertinent literature. A summary documentation of the comparative magnitude of these effects has the potential to provide guidance to the design of interventions to reduce the effects of stereotype threat in real-world testing settings.

Schmader, Johns, and Forbes (2008), Smith (2004), and most recently Pennington, Heim, Levy, and Larkin (2016) have taken on the task of providing qualitative reviews for a number of studies involving stereotype threat mediators. Schmader et al. reviewed stereotype threat mediators as a starting point to building a multi-mechanism model linking stereotype threat to performance. Schmader et al.'s model incorporates a number of physiological, cognitive, and affective mediators. In her review, Smith provided an extensive qualitative review of the previously studied stereotype threat mediators; however, because of the lack of any clear evidence for any of the reviewed mediators, she proposed a new mechanism through which stereotype threat may be impacting performance. Pennington et al. (2016) provides a systematic literature review of 45 studies through the lens of a multi-threat framework (Shapiro & Neuberg, 2007). Similar to Schmader et al.'s (2008) review, Pennington et al. (2016) investigated affective, cognitive, and motivational mediators. Pennington et al. (2016) determined that there was support for each type of mediator. However, the empirical support varied and they proposed that it may be due to a number of moderators, including the target of the stereotype threat, the stereotype threat activation cues, and what measures were used to capture the mediator and performance. Altogether each review is helpful in building a framework to understand the processes underlying stereotype threat's effects. However, to fully understand these complex models, an

integral next step is to understand how stereotype threat impacts the mediators proposed in these models. Once a clear link can be made between stereotype threat and its different mediators, then the proposed models can be empirically tested in a more informed manner.

Arguments against Stereotype Threat

Stereotype threat is a popular topic as reflected in the fact that in the past 15 years a number of meta-analysis have been conducted on the subject (Appel, Weber, & Kronberger, 2015; Flore & Wicherts; 2015; Lamont, Swift, & Abrams, 2015; Nguyen & Ryan, 2008; Shewach, Sackett, & Quint, 2019; Walton & Cohen, 2003). While Appel et al. (2015) and Flore and Wichert (2015) focused on groups (immigrants, young girls) outside of the realm of the current study, the remaining meta-analyses do provide information relevant to the current study. Walton and Cohen's (2003), Nguyen and Ryan's (2008) and Shewach et al. (2019) meta-analyses focused on gender and race-based stereotype threat, and Lamont et al.'s (2015) meta-analysis focused on age-based stereotype threat. Walton and Cohen's (2003) meta-analysis included 43 effect sizes, Nguyen and Ryan (2008) meta-analyzed 116, Shewach et al.'s (2019) overall analysis included 181 effect sizes, and Lamont et al. (2015) meta-analyzed 37 effects. Ninety-two of the studies (79%) included in Nguyen and Ryan's (2008) meta-analysis were not included in the Walton and Cohen (2003) meta-analysis, and 132 (62%) of Shewach et al.'s (2019) samples were distinct from Nguyen and Ryan's. Affirming the continued popularity of stereotype threat, a recent search for "stereotype threat" using PsycINFO that was limited to studies published after Nguyen and Ryan (2008) identified an additional 485 articles relevant to the topic.

Despite its popularity, stereotype threat research is not without its critics. One notable criticism centers on the interpretation of stereotype threat research. Sackett, Hardison, and Cullen

(2004) criticized the interpretations of Steele and Aronson's (1995) findings. Sackett et al. (2004) noted that many people in both academia and popular media have misinterpreted Steele and Aronson's findings to mean that stereotype threat completely explained the cognitive ability test score differences between African Americans and Whites by failing to realize that prior SAT scores were used as a control variable in the analyses. Sackett et al. (2004) do not discount Steele and Aronson's findings but simply caution against the broad generalizations and loose interpretations of their findings. Steele and Aronson (2004) also acknowledged that there have been some misinterpretations of the original results from Steele and Aronson (1995). However, they maintain the majority of researchers, including themselves, are not claiming that stereotype threat completely explains the race gap in cognitive ability test scores. Additionally, Steele and Aronson (2004) assert that although their original study controlled for previous SAT scores, a number of follow-up studies that have not used SAT scores as a covariate have replicated the original findings, thus demonstrating the robustness of the effect. Supporting this claim, meta-analytic results support the presence of stereotype threat effects. Nguyen and Ryan (2008) reported *d*-values that ranged from 0.11 to 0.80 for race and gender-based stereotype threat and Lamont et al. (2015) reported *d*-values from 0.28 to 0.52 for age-based stereotype threat.

A second major concern in the stereotype threat literature has been the generalizability of findings outside of the laboratory setting (Cullen et al., 2004; Ryan & Sackett, 2013; Sackett et al., 2004; Sackett & Ryan, 2011; Sackett, Schmitt, Ellingson, & Kabin, 2001; Shewach et al., 2019). Specifically, Ryan and Sackett (2013) outlined a number of factors that are necessary for stereotype threat to occur and questioned whether real-world settings, specifically workplace settings, actually meet these conditions. They state that experiments in specifically designed and controlled laboratory settings create conditions that do not naturally occur in workplace

assessments. However, others (Roberson & Kulik, 2007; Steele & Davies, 2003) suggest that the stereotypes that exist in broader society are still relevant and active in daily workplace experiences. Additionally, many of the laboratory conditions that are considered to be necessary for stereotype threat to occur are factors that are inherent in the workplace setting and may accentuate stereotype threat effects (Steele & Davies, 2003). These conditions include (a) existence and awareness of a negative group stereotype, (b) required performance of a task that is diagnostic or evaluative of a skill that is negatively stereotyped, (c) tasks that are difficult, and (d) the negatively stereotyped individual cares about doing well on the task. Walton, Murphy, and Ryan (2015) suggest that task difficulty may be the only condition of stereotype threat that is not met in the workplace. However, they also note that many of today's jobs are increasing in complexity and the number of challenges faced by the working population are also growing, therefore this concern is likely less relevant. Additionally, while Ryan and Sackett (2013) argue that high domain identification is unique to research participants chosen specifically for stereotype threat studies, Steele and Davies (2003) counter that domain identification is just one factor that strengthens the predicament of feeling one might confirm a negative stereotype. Steele and Davies (2003) further assert that laboratory settings actually underestimate the size of stereotype threat effects because there are a limited number of real-world factors that can be manipulated in a laboratory and they do not always mimic the high-stakes nature of testing situations. They state, "The question for future research is in what forms of employment testing are these effects most significant and what can be done to reduce them" (Steele & Davies, 2003, p. 319). Thus, it seems that both sides of the debate would maintain that the real-world implications for stereotype threat is an area that needs further study, particularly in the realm of developing effective intervention strategies.

Although the focus of the current study is on mediators of stereotype threat, rather than the generalizability of stereotype threat findings to real-world settings, the findings from the current study will provide an important contribution to resolving the ongoing debate. An understanding of the mediators of stereotype threat effects is a critical issue that needs to be addressed for a number of reasons. There is no doubt that stereotype threat effects are robust and detrimental to performance (Lamont et al., 2015; Nguyen & Ryan, 2008). However, although there is a wealth of information on how to activate stereotype threat, what is less known is how to combat its effects. The lack of insight regarding stereotype threat effect mitigation and prevention may be due to the dearth of information on what is triggering the effects. Subsequently this is what interventions need to target. It is impossible to mitigate stereotype threat's negative influence on performance without understanding the mechanisms that lead to that outcome (Ryan & Sackett, 2012).

Ideally, just as doctors must diagnose the reason a patient is experiencing certain symptoms before providing treatment, we must identify what is underlying stereotype threat effects before implementing an intervention to stop the negative effects. To this point, what has clearly been demonstrated is the link between stereotype threat and lower cognitive test performance (Lamont et al., 2015; Nguyen & Ryan, 2008). Although there has been substantial work exploring the mediators of stereotype threat (see Pennington et al., 2016 and Schmader et al., 2008) a clear understanding of these connections is still murky, leaving researchers still unclear on understanding the underlying mechanisms between stereotype threat and performance, the "why". Consequently, we are faced with implementing solutions targeted at stopping the activation of stereotype threat which are ineffective because of the pervasiveness of conditions in everyday life that activate stereotype threat (Steele, 2012).

As previously discussed, stereotype threat is an important phenomenon in educational and organizational settings because its negative effects are linked to lower scores on high-stakes tests that are used for selection, promotional, and educational attainment purposes. This presents a problem as educational and organizational systems strive for diversity. The lowered test scores caused by stereotype threat primarily impact minority test-takers, and consequently hinders their entry and advancement in these settings (Steel et al., 2002). Ryan and Sackett (2013) recognized that while researchers have sought to find strategies to prevent stereotype threat and its effects, what has been missing is a targeted effort at reducing the mechanisms through which performance is harmed. Accordingly, the objective of the current study was to examine the extent to which stereotype threat impacts a variety of proposed mediators by first providing a review of current research that has investigated stereotype threat and its connection to several different mediators and meta-analyzing these stereotype threat-mediator relationships in the extant literature. Subsequently, the findings of this meta-analysis can provide a starting point to addressing, minimizing, and developing strategies to negate stereotype threat effects in real-world testing environments.

Pertaining to the structure of the dissertation, first an overview of stereotype threat and its defining characteristics is presented, followed by a review of the proposed mediators of the stereotype threat-performance relationship. Included in this review is a discussion of several proposed moderator variables for the stereotype threat-mediator relationships. Following these sections, a detailed methodology of the meta-analysis is presented along with the results of the current study. Finally, theoretical and practical implications of the results are discussed, as well as the limitations of the current study and future directions for research in this area.

STEREOTYPE THREAT

In 1995, Claude Steele and Joshua Aronson published one of the first empirical studies on stereotype threat. In this influential study, Steele and Aronson documented that simply being made aware of a negative group stereotype related to cognitive ability, lowered the performance of highly intelligent African American students' on a difficult intelligence test. However, when African American students were told that their performance on the test was not a measure of ability, the students performed significantly better than their peers in the stereotype threat condition ($d = 0.51 - 0.80$). Since this first influential study, a number of studies have documented this same phenomena across a diverse set of groups and in a variety of domains (e.g., Ben-Zeev, Dennehy, Sackman, Olide, & Berger, 2011; Croizet, Desert, Dutrevis, & Leyens, 2001; Marx, Ko, & Friedman, 2009; Shih, Pittinsky, & Ambady, 1999; Stone, Lynch, Sjomeling, & Darley, 1999). Additionally, a large amount of the extant research has been devoted to understanding additional factors involved in the stereotype threat phenomenon, including mediating processes that underlie stereotype threat effects, as well as the moderators that influence the strength of these findings. The following sections provide an overview of stereotype threat, its moderators, and finally a review of the proposed mediators of stereotype threat.

Characteristics of Stereotype Threat

Stereotype threat is a “situational predicament in which individuals are at risk, by dint of their actions or behaviors, of confirming negative stereotypes about their group,” (p. 5; Inzlicht & Schmader, 2012). The possibility of confirming a negative stereotype leads to increased pressure felt by members of stereotyped groups, but instead of increased performance this pressure is associated with decreased performance. For example, African Americans are

susceptible to the stereotype that their race is intellectually inferior, and when they are placed in a situation that is designed to measure intellectual ability (e.g., standardized testing context) they are at risk of confirming the stereotype. Correspondingly, the distress over the possibility of confirming the stereotype is associated with the lower performance of African Americans on intelligence tests (Steele & Aronson, 1995). In contrast, when African Americans are told that an intelligence test is a simple laboratory task, they perform better than when it is described accurately and more similarly to their White peers.

Social identity theory is the foundation of stereotype threat research (Steele, Spencer, & Aronson, 2002). Social identity theory is based on the underlying assumption that individuals hold a number of different social identities (e.g., gender, race, ethnicity, social class, religion, etc; Tajfel & Turner, 1986). The theory posits that situations signal the importance of a particular identity, and a vigilance process then begins. Based on the identity that is made salient, a person is motivated to promote a positive social identity. However, a situation may also signal the possibility that the salient identity is a liability and in danger of a negative appraisal (e.g., negative stereotypes). The possibility of an individual being evaluated based on a negative stereotype about his/her group identity is in direct contradiction and a threat to the individual's underlying motivation to display a positive image of his/her group and therein himself/herself. Put another way, when a negative stereotype is associated with a group with whom an individual identifies, it threatens their personal identity.

The link between identity and stereotype threat can be better understood through a model proposed by Schmader et al. (2008). The model puts forward that stereotype threat effects are the result of incongruence between concepts of the self, group, and domain. Stereotype threat situations cue the misalignment of the characteristics by (1) making group membership salient

(e.g., having women indicate their sex before a math test), (2) then priming a negative group stereotype around the domain (e.g., women being bad at math), but these occur when (3) individuals belief about their ability in the domain is positive, signaling incongruence across the three. Oftentimes this appraisal process can operate outside of consciousness and individuals' responses may vary depending on their level of awareness and how much they work to control this process. Threat effects can also vary depending on the degree of the incongruence and where the incongruence occurs (e.g., self and group concepts, self and domain concepts; Shapiro & Neuberg, 2007).

Additionally, threat can be tied to any of the many identities a person can hold, and threat effects may manifest differently depending on which identity stereotype is active (Shapiro, 2011; Shapiro, Williams, & Hambarchyan, 2013). Identities may differ in their level of concealability, perceived controllability, and ability to change over time. Thus, stereotype threat can be experienced differently depending on these different characteristics of the salient identity.

Steele et al. (2002) highlight three general characteristics of stereotype threat. First, it is situation specific. Any particular situation may send cues that heighten the saliency of a certain identity. For example, a math class may send cues that mathematical ability is important which in turn can make gender identity important for women, due to the stereotype that women are inferior to men in mathematical ability. When an identity is activated, an appraisal process begins to determine how the identity will be evaluated (Murphy & Taylor, 2012). The situation may either signal that there is no threat of negative evaluation or that one's social identity is linked to negative evaluation or stigmatization. In the first case, an individual is able to relax and function normally (Cohen & Garcia, 2008). However, when there is the possibility of a negative evaluation, an individual heightens their vigilance and is "on alert". Correspondingly, Steele

(1997) described stereotype threat as a “threat in the air” felt when a situation signals the saliency of a negative stereotype related to a certain identity.

While traditionally studied in cognitive ability settings, stereotype threat has also been activated in social settings. These effects are often manifested when non-Blacks are afraid of exhibiting behaviors that may confirm the stereotype that those in the racial majority hold racist beliefs (Richeson & Shelton, 2003). Thus, the situation does not have to be limited to the cognitive domain to display stereotype threat effects.

Secondly, because any situation can signal that a particular identity is salient, stereotype threat is a general phenomenon that can be felt by anyone depending on what identity is made salient. For example, although not traditionally associated with negative ability stereotypes, Aronson, Lustina, Good, Keough, Steele, and Brown (1998) made White males cognizant of the stereotype that Whites are inferior in mathematical ability in comparison to Asian Americans. The activation of this stereotype resulted in the White males in the threat condition performing worse on a math test than their counterparts who were not made aware of the stereotype. Thus, while typically studied in low-status groups like racial minorities and women, Aronson et al. (1998) demonstrate that any group can be susceptible to stereotype threat.

The last general characteristic of stereotype threat is the nature of the threat felt is contingent on the content of the stereotype. That is, the nature of the stereotype guides who will experience threat and in what settings. For example, an African American male may experience threat in a classroom setting but not in an athletic setting (e.g., Stone et al., 1999). This relates back to the Schmader et al.’s (2008) model and the principal of incongruence. In the athletic example there may be no incongruence between the self, group, and domain concepts; the athlete identifies as African American, and based on race stereotypes tied to athletics the athlete knows

African Americans are typically skilled in athletics (Stone et al., 1999), they also believe they are a good athlete, thus each concept is in line with each other and stereotype threat does not occur in this domain. And although there are stereotypes associated with African Americans in both academic and athletic settings, they will feel threat in an academic setting but not in an athletic setting because of the content of the different stereotypes.

Correspondingly, when positive stereotypes are made salient in a situation, stereotype boost can occur. *Stereotype boost* is when the presence of positive group stereotypes improves an individual's performance on a task (Shih, Pittinsky, & Ho, 2012). Following the example above, priming the positive athletic stereotype for Black athletes is associated with increased athletic performance. Stereotype boost is different from *stereotype lift* which occurs when an individual's performance improves because of being exposed to negative stereotype of another group unrelated to them (Walton & Cohen, 2003). Both stereotype boost and lift demonstrate how the nature of the stereotype can change the effect it has on an individual.

The key outcome in stereotype threat research studies is test performance, specifically a test measuring some form of ability (e.g., verbal, quantitative, analytical, spatial). The specific type of ability being assessed is a function of the stereotype being targeted (e.g., quantitative ability for women). In general, support for the effects of stereotype threat has been consistent. Anecdotally, Steele and Davies (2003) argue that over 100 studies have found support for the effects of stereotype threat. Two meta-analyses have tested Steele and Davies's assertion in relation to gender and race-based stereotype threat effects. Walton and Cohen's 2003 meta-analysis showed a mean effect size of $d = 0.29$ ($k = 43$) for stereotype threat. Additionally, a subsequent meta-analysis by Nguyen and Ryan (2008) revealed a similar stereotype threat effect size ($d = 0.26$, $k = 116$). Nguyen and Ryan's meta-analysis also investigated the effects of

stereotype threat on women and minority test-takers separately and found the mean effect size for gender-based stereotype was $d = 0.21$ ($k = 44$) and $d = 0.42$ ($k = 72$) for race-based stereotypes. Lamont et al.'s (2015) age-based stereotype threat meta-analysis obtained a mean effect size of $d = 0.32$ ($k = 53$). Although these effect sizes would be considered “small to medium”, each of the meta-analyses found a number of moderators impacted the size of the effects. Moderator analyses revealed that stereotype threat effects ranged from $d = 0.11$ to 0.80. A discussion of the moderators that impact stereotype threat is presented in the following section.

Moderators of Stereotype Threat

As illustrated by the findings of Walton and Cohen (2003), Nguyen and Ryan (2008), and Lamont et al.'s (2015) meta-analyses, there are a number of variables that moderate the effects of stereotype threat. A moderator is a variable that affects the magnitude and/or direction of the relationship between an independent variable and a dependent variable (Baron & Kenny, 1986). In the case of stereotype threat a number of factors have been found to affect the extent to which stereotype threat impacts performance. These include characteristics of the situation (e.g., activation of stereotype threat), the test (e.g., test difficulty), as well as characteristics of the individual (e.g., domain and group identification).

First and foremost, it is important to recognize that stereotypes vary in their meaning and negativity. A stereotype about a group's intelligence can be much more damaging than a stereotype that demeans a group in more trivial ways such as a group's sense of fashion or timeliness. In addition, the ways that these stereotypes are activated may have an effect on performance. Both Walton and Cohen (2003) and Nguyen and Ryan (2008) found that the way in which stereotype threat was activated, whether blatantly, moderately, or subtly, differentially

impacted test performance. The differential impact of activation type and level can be attributed to the process that is initiated. For example, when a stereotype threat is activated by highlighting the minority test-taker's inferiority in the domain (e.g., Aronson et al., 1999) a reactance effect may occur (Kray et al., 2001) and performance may actually improve as the test-takers try to actively combat the stereotype. However, when stereotypes are activated in a less explicit manner (e.g., subtly mentioning the possibility of group differences on a test) the attention of the test-taker shifts away from the test to the interpretation of the information. Consequently, this distraction negatively impacts performance (Nguyen & Ryan, 2008). The effects of a subtle activation of stereotype threat (e.g., having participants indicate their race before a test, [Steele & Aronson, 1995]) can vary, because test-takers may not be aware that stereotype threat has been activated or that an automatic subconscious process that leads to deleterious outcomes has been set into motion.

Walton and Cohen (2003) only investigated two levels of stereotype threat activation and found that blatant explicit activation strategies yielded larger effect sizes ($d = 0.57$) than subtle indirect strategies ($d = 0.29$). Nguyen and Ryan (2008) found the magnitude of stereotype threat effects varied as a function of both activation strategy and the test-taker. Moderate activation strategies were most detrimental to minority test-takers performance ($d = 0.64$) and subtle activation strategies were the worst for female test-takers ($d = 0.24$). These findings suggest that different processes may be activated for minority and female test-takers. This finding is particularly relevant to the current study in the investigation of stereotype threat's effect on mediators.

The extent to which an individual identifies with the domain with which a stereotype is associated can also impact the magnitude of stereotype effects (Steele et al., 2002). When the

domain is a large part of a person's self-concept, then the prospect of failure in that domain can be particularly detrimental. Stereotype threat primes the possibility of failure in these types of circumstances. For example, only women who are highly identified with the math domain would experience stereotype threat because those women who are not identified with the domain would not have their identity threatened by failure on a math test (Cadinu et al., 2003). As a result, participants in stereotype threat studies are often pre-screened and only those with high domain identification are selected to participate in the study (e.g., Brown & Pinel, 2003; Davies et al., 2002). Walton and Cohen's (2003) results demonstrate that the effects of stereotype threat on test performance are especially costly for those individuals who are highly identified with the test domain in comparison to those who do not identify with the test domain. However, Nguyen and Ryan's (2008) meta-analysis found that stereotype threat was more detrimental for women with medium-levels of domain identification, than for women who were highly identified with the domain. Altogether, both meta-analyses demonstrate that domain identification affects the stereotype threat-test performance relationship, albeit at different levels.

In addition to domain identification, the level of group identification can also have an effect on the impact of stereotype threat. The more a person identifies with a group that has a negative stereotype associated with it, the more likely they will experience stereotype threat and have their performance affected by it (Schmader, 2002). Schmader (2002) found that women in a stereotype threat experiment condition who highly identified with their gender performed significantly worse on a math test than women whose identity was not tied to their gender.

The degree of stereotype threat experienced can also be impacted by an individual's ability to cope with the threat (Steele et al., 2002). Coping with stressful situations is largely understood through the transactional stress model proposed by Lazarus and colleagues (e.g.,

Folkman, Lazarus, Dunkel-Schetter, DeLongis, & Gruen, 1986; Lazarus & Folkman, 1984). The transactional stress model posits that the stress experience manifests as a result of the interaction between the individual and the environment (Lazarus & Folkman, 1984). A stressful encounter begins in the assessment of the situation. A *primary appraisal* is the initial assessment of a situation as threatening. A *secondary appraisal* is the individual's assessment of their ability to cope with and handle the threatening situation. In relation to experiences of stereotype threat, the effects of stereotype threat are contingent on both the individual's primary and secondary appraisals. In line with Schmader et al.'s (2008) model, the primary appraisal would be the recognition of incongruence between self, group, and domain while the secondary appraisal would be the process of trying to reconcile the incongruence.

The level of test difficulty is another key moderator in the stereotype threat-performance relationship. Difficulty is considered important because an individual is most likely to feel threatened when the test challenges the individual's ability (Steele et al., 2002). When the challenge is presented, the possibility of failure increases and an individual becomes more aware of the likelihood of confirming an inferiority stereotype. What makes these situations particularly damaging are that difficult tasks require the full mental capacity of an individual, but thoughts focused on the stereotype force their full attention away from the tasks and performance suffers as a result. Nguyen and Ryan's (2008) meta-analysis provides support for the test difficulty hypothesis; their results revealed that the negative effects of stereotype threat on test performance were most pronounced on difficult tests for both minority and female test-takers. Recognizing the moderators of stereotype threat is key to understanding how stereotype threat effects may vary under different circumstances. Taken together, the findings discussed above demonstrate that the characteristics of the test environment, the test-taker, and the test can all

influence the nature of stereotype threat effects. Similar to needing a clear delineation of mediator effects to inform intervention strategies to reduce stereotype threat, a deeper understanding of the moderators of stereotype threat can inform for whom, and in what types of contexts stereotype threat intervention strategies will be the most effective. Additionally, a recognition of the moderators of the stereotype threat-performance relationship can contribute to understanding the nuances of the stereotype threat-mediator relationships. Proposed moderators for the mediators that are investigated in the current study are discussed at the conclusion of the mediator section. Moderators specific to the different mediating mechanisms are discussed in the respective sections of each mediator.

MEDIATORS OF STEREOTYPE THREAT

“And the self-threat it causes — through a variety of mechanisms — may interfere with intellectual performance.” (Steele & Aronson, 1995, p. 797).

The preceding sections have outlined the factors that may affect the experience and effects of stereotype threat. Another important issue and the focus of the current study are the processes that lead to stereotype threat effects. That is, what mechanisms occur between the experience of stereotype threat and its adverse effects on performance? Put another way, what variables mediate the stereotype threat-performance relationship?

Mediators are essential to understanding psychological phenomena because they explain the processes underlying observed relationships (MacKinnon, Cheong, & Pirlott, 2012). A mediator describes *why* a relationship between an independent and dependent variable exists, and *how* the independent variable influences the dependent variable (Baron & Kenny, 1986). The causal chain $X \rightarrow M \rightarrow Y$ provides an illustration of the process by which X (predictor) affects Y (criterion) through M (mediator). For the purpose of the current study X represents stereotype threat and Y represents test performance while the M represents the proposed mediators discussed in the following sections.

The first step in the stereotype threat causal chain is the activation of a negative stereotype (Marx, 2012; Steele & Aronson, 1995). Early work by Steele and Aronson demonstrated that experiences of stereotype threat begin by activating a schema of the particular stereotype. For example, in the Steele and Aronson (1995) study, the African Americans who were about to take an intelligence test (diagnostic condition) were more likely to complete the word fragment _ _CE with the word RACE than their White counterparts. Furthermore, African Americans who were in the diagnostic condition were more likely to complete words in a race-

related manner than the African Americans in the nondiagnostic condition. Additionally, there is no requirement that the activation of stereotype threat has to be an explicit process or occur at a conscious level (Schmader & Beilock, 2011). Nguyen and Ryan (2008) found that it was more subtle activations of stereotype threat that had the biggest impact on women while it was moderately explicit activations that had the biggest impact on the test performance of racial minorities.

Furthermore, work by Marx and Stapel (2006) distinguished between stereotype priming and stereotype threat effects. Although activation is a requirement for both, the distinctive characteristic is in the target of the stereotype; stereotype priming can impact anyone exposed to a stereotype, but stereotype threat effects only impact the victims of the stereotypes. Marx and Stapel refer to this as the “knowing and being” requirement, you must *know* about the group stereotype (e.g., women are bad at math), and *be* a member of the group (women). For example, Marx and Stapel found that both men and women performed worse on a math test after they were primed with words related to not being intelligent. Additionally, in the stereotype threat condition, although all scores were lower, women performed worse than men in the unintelligent prime condition. Thus, while priming and threat effects both require activation, they are distinct experiences with unique outcomes. Altogether, understanding the activation phase is an important first step in understanding the effects of stereotype threat because it signifies the starting point in the process.

Next is determining what occurs after the activation of a negative stereotype. Steele and Aronson (1995) hypothesized a number of mediators that may underlie stereotype threat effects including distraction, anxiety, and withdrawal of effort (via self-handicapping). As work in the field of stereotype threat has grown, the list of potential mediators has also continued to expand

and become further developed. The following section outlines these different mechanisms that are proposed to mediate the effects of stereotype threat. To start, the mediators proposed by Schmader et al. (2008) in the integrated process model of stereotype threat are discussed. Additional mediators, including those reviewed by Smith (2004) and others not included in Schmader et al.'s (2008) model are also discussed.

The following sections discuss the mechanisms that affect performance once stereotype threat is activated. The discussion within these sections are limited to the mechanism undermining performance on cognitive tasks. While it has been demonstrated that stereotype threat can have an impact on cognitive, social, and sensorimotor tasks, for the purpose of this study the discussion is limited to the effects on cognitive tasks only. By restricting the current investigation to stereotype threats' effect on cognitive test performance, it allows the current study to address an issue that is an increasingly relevant and germane topic to organizations. It is important to note that stereotype threat can also affect organizations through other negative outcomes (e.g., lowered job attitudes, increased turnover intentions; Kalokerinos, von Hippel, & Zacher, 2014); however, its effects on these other outcomes are similar to the effects on cognitive tasks but through a different set of mechanisms (Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Schamder et al., 2008).

A Model of Stereotype Threat Effects

Schmader et al.'s (2008) integrated process model of stereotype threat effects on performance outlines a number of paths by which stereotype threat can affect and influence performance on various tasks (see Figure 1). The impetus for the model was the authors' acknowledgment that although anxiety and stereotype threat activation are critical components to stereotype threat effects, these explanations are too simplistic. Thus, the authors proposed a

model that incorporated activation and anxiety but also included a number of other mechanisms through which stereotype threat impacts performance.

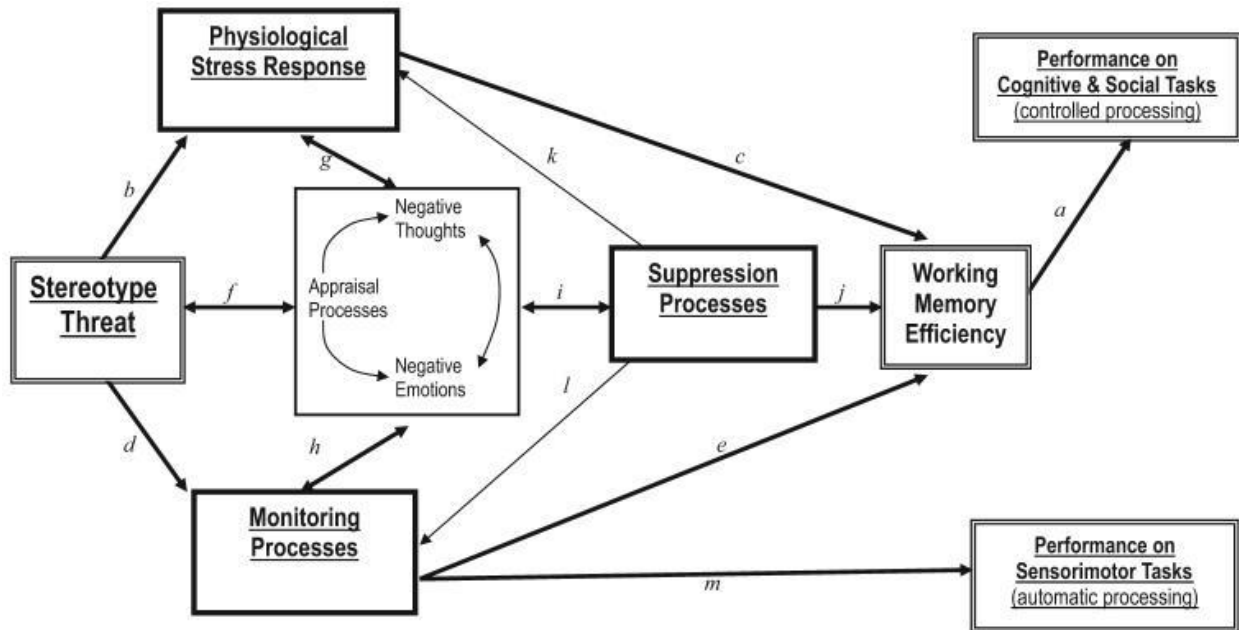


Figure 1. Figure 1 from Schmader, Johns, and Forbes (2008), “An integrated process model of stereotype threat effects on performance”.

Starting from the leftmost side of Figure 1, the model begins with a stereotype threat occurrence that sets into motion one of a number of responses either separately or simultaneously. The next components of the model include physiological reactions (Path *b*), monitoring behaviors (Path *d*), and a cycle of negative thoughts, emotions, and appraisal processes (Path *f*) which lead to attempts to suppress the negative thoughts (Path *i*). The physiological, monitoring, and suppressing behaviors all then affect performance through working memory efficiency (Paths *c*, *e*, and *j*), and working memory is then linked to performance on controlled processes. As noted in the last components in the model (rightmost boxes in Figure 1), a key distinction of this model is that stereotype threat affects performance

through both automatic and controlled responses to stereotype threat (Schmader & Beilock, 2012).

While Schmader et al.'s (2008) model sets an outline for the current meta-analysis, it does not include all proposed mediators of stereotype threat effects. Thus, for the purpose of the current study, Schmader et al.'s integrated process model of stereotype threat effects is considered a useful guide; however, a model which includes additional proposed mediators investigated in the current study is presented in Figure 2.

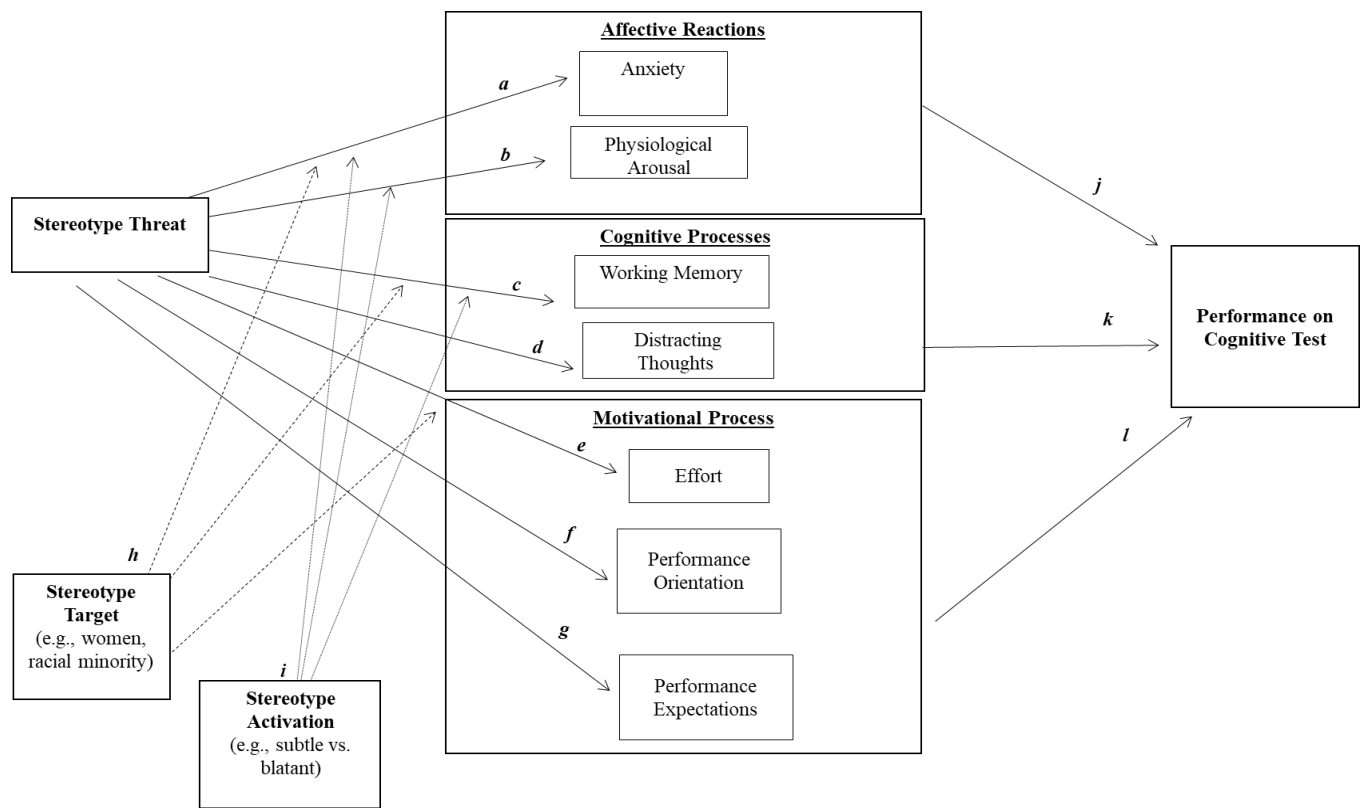


Figure 2. A model of stereotype threat effects on cognitive test performance for stereotyped groups.

It is important to note that Figure 2 has collapsed some of the components of Schmader et al.'s (2008) model into single categories for clarity and conciseness. The model advanced in the

present study organizes the mediators into broad categories of affective, cognitive, and motivational mediators. Additionally, Figure 2 allows for moderators to be built into the mediation model. The goal of the present study is to fill in the links, designated by letters (e.g., *a*), presented in Figure 2, to help build a more comprehensive understanding of stereotype threat effects. Links *a – g* relate to specific hypotheses for each of the moderators, and links *h* and *i* investigate the impact broad moderators may have on the stereotype threat-mediator relationships. The aforementioned links align to specific hypotheses. Links *j – l* are not specifically analyzed in the current meta-analysis but are based on past research and will be discussed further in the discussion. In the sections that follow, each of the links between stereotype threat and the proposed mediators are discussed and the specific hypotheses related to each are presented.

Affective Reactions

Affective reactions are one mechanism via which stereotype threat can negatively affect performance. As discussed previously, stereotype threat represents a stressor. As a stressor, it is hypothesized to lead to a state characterized by heightened arousal, discomfort, anxiety, and distress because of the cognitive inconsistency (e.g., individual expectations not matching expectations of the stereotype image) signaled by the stereotype threat situation (Schmader et al., 2008). These are akin to the feelings that occur when people experience cognitive dissonance (Elliot & Devine, 1994).

Stereotype threat research has investigated stereotype threat's effect on affective reactions through anxiety and physiological reactions (e.g., Bosson, Haymovitz, & Pintel, 2003; Johns, Schmader, & Inzlicht, 2008). Anxiety is defined as “an aversive emotional and motivational state occurring in threatening circumstances” (p. 336; Eysenck, Derakshan, Santos,

& Calvo, 2007). Anxiety is often broken down into emotionality and worry components (Liebert & Morris, 1967). Physiological arousal is considered a part of the emotional component of anxiety (Morris, Davis, & Hutchings, 1981). Individual's perceptions of their physiological reaction to a stressful situation, through increased feelings of nervousness and tension, are indicative of the emotional component of anxiety. The worry component is considered the cognitive component, where individuals are concerned with negative expectations and outcomes from the situation. Thus, although arousal is considered to be a distinct component of anxiety, it is often studied concurrently rather than separately from other anxiety outcomes. As a result, when the effects of anxiety and arousal are studied, their differential effects are indistinguishable from each other. The following sections discuss how each is affected by stereotype threat and subsequently how each affects performance.

Anxiety. Anxiety can be defined in two separate ways, state and trait anxiety. Trait anxiety is a stable personal characteristic related to general levels of apprehension and nervous tension. Contrastingly, state anxiety is a short-term anxiety dictated by the amount and nature of stressors experienced by an individual at any point in time (Eysenck, 1992). For example, the stress elicited by stereotype threat scenarios are considered to initiate state anxiety for minority test-takers. However, certain personal characteristics, like trait anxiety, may make an individual more susceptible to experience state anxiety in a stereotype threat situation. In line with the current study, trait anxiety would be considered a moderator of the stereotype threat-performance relationship, while state anxiety would be considered a mediator of the stereotype threat-performance relationship.

Stereotype threat is considered to be anxiety-provoking situation because of the evaluative component that is a central tenant to the stereotype threat scenario (Walton et al.,

2015). The possibility of potentially confirming a negative stereotype is inherently threatening to individuals and can trigger arousal, distress, and subsequently anxiety (Schmader et al., 2008). Therefore, the first link in the proposed model (Path *a*), proposes that individuals exposed to stereotype threat will have higher levels of anxiety.

Hypothesis 1: Anxiety will be higher in stereotype threatened individuals.

Physiological Arousal. Furthermore, stereotype threat research has broken out stereotype threat's anxiety effects further, and also considers physiological arousal responses (e.g., Bosson, Haymovitz, & Pinel, 2004; Johns et al. 2008). Corresponding with the emotional definition of anxiety, Johns et al. (2008) describe the arousal reaction as affective in nature. For stereotype threat, the arousal elicited by a stereotype threat situation sends "distress" signals to activate vigilance and control behaviors (Forbes, Schmader, & Allen, 2008). This is evidenced by increased activation of the sympathetic nervous system.

Stereotype threat elicits physiological responses on two fronts. First, at a neurological level, when individuals are in threatening situations the region of the brain related to self-control is affected. When an individual is confronted with stereotype threat, the anterior cingulate cortex sends "distress" signals to activate vigilance and control behaviors (Forbes et al., 2008). These signals make individuals increasingly aware of how they could commit a test-related error. However, these behaviors are not necessarily needed nor are they useful. In fact, they can be ineffective and draining in a testing situation because they direct attention and energy away from the task at hand.

The second physiological response occurs as stereotype threat activates the adrenal system to respond to a stress experience. Similar to the neurological response, the hormones secreted by the adrenal system put the body in a state of hyper-vigilance. Although these

hormones are useful in helping the body respond to other types of threat (e.g., physical danger), they are less useful in a testing situation and hinder performance by keeping the body on alert and distracting threatened individuals from the test or other performance task (Lupien & McEwen, 1997; Schmader et al., 2008). In most situations an optimal level of arousal is needed for successful performance. The optimal level of arousal is defined by the Yerkes and Dodson (1908) inverted-U, where low and high levels hurt performance, but moderate levels result in peak levels of performance.

Research investigating the link between hormone secretion and decreased task performance has not been extended to stereotype threat. However, Matheson and Cole (2004) provide a first look into how the aforementioned effects may play out in a stereotype threat effect setting. Matheson and Cole investigated group identity threat by exposing individuals to a university-related stereotype with different valences (i.e., positive, negative, neutral). The authors found that threat was related to increased cortisol-levels (i.e., increases in vigilance responses) but only when an individual used emotion-focused coping. However, they failed to obtain a similar effect when individuals endorsed a problem-focused coping style. In fact, cortisol reactivity decreased when problem-focused coping was used, even when threat was present. This study provides some support for the link between stereotype threat, physiological reactions, and performance decrements. Additionally, the Matheson and Cole study demonstrates the nuanced relationship between threat and physiological reactions and suggest the need for additional work in this area to better understand these types of relationships (Schmader et al., 2008).

Stereotype threat research has demonstrated mixed results for whether the anxiety and distress occur at both the conscious and unconscious levels. Direct measures (i.e., when

participants are explicitly asked about their stress levels) have not demonstrated a clear pattern of results. For instance, Spencer et al. (1999) found that women reported more anxiety when presented with a math test than women who were given the test but also told it had no history of differences in scores for men and women. However, Gonzales, Blanton, and Williams (2002) did not find any differences in self-reported anxiety in participants in threat and non-threat conditions. Conversely, Sawyer and Hollis-Sawyer (2005) found a relationship between stereotype threat and self-reported test anxiety but in the opposite direction than what has been hypothesized (i.e., threatened individuals felt less anxious). In her review of stereotype threat mediators, Smith (2004) concluded that anxiety is not a plausible mediator of stereotype threat and its effects. This conclusion, and the conflicting findings mentioned above, may be due to the use of self-report measures of anxiety in stereotype threat conditions. These types of measures may not accurately and fully capture an individual's experiences of stress and anxiety because individuals in threat conditions may be attempting to regulate and discount their anxiety (Johns et al., 2008). On the other hand, although indirect measures of arousal are not necessarily measures of anxiety, they can serve as markers of an individual's emotional state.

Correspondingly, these indirect measures have shown a much clearer link between stereotype threat and increases in stress-based arousal (e.g., Ben-Zeev, Fein, & Inzlicht, 2005; Blascovich, Spencer, Quinn, & Steele, 2001; Bosson et al., 2004). Therefore, stereotype threat may cause stressful reactions that are best measured through indirect and physiological measures as a person may not even be aware of their body's reaction to the threat. The different types of measures are differentiated in the hypotheses addressed in the present study. First, it is proposed that individuals will have increased physiological arousal in stereotype threat situations (Path *b*). Second, it is proposed that because physiological arousal measures are less likely to be clouded

by an individual's motivation and flawed judgement, the stereotype threat effects for physiological arousal will be larger than the effects for more subjective self-report anxiety measures.

Hypothesis 2: Physiological arousal will be higher in stereotype threatened individuals.

Hypothesis 3: The mean effect size for physiological arousal will be larger than the effect size for anxiety for stereotype threatened individuals.

Cognitive Processes

A second mechanism through which stereotype threat can affect performance is via cognitive processes. The demands of the stereotype threat situation are proposed to negatively impact an individual's cognitive processes and prevent them from operating optimally. Specifically, it is proposed that stereotype threat negatively impacts cognitive processes by decreasing working memory capacity (Path *c*) and the ability to effectively manage negative thoughts (Path *d*).

Working Memory. Working memory is an integral part of Schmader et al.'s (2008) integrated process model. It has been proposed to be the most proximal connection to performance on cognitive tasks because of working memory's strong relationship with cognitive ability (e.g., Schmader et al., 2008; Schmader & Johns, 2003). As such, it is considered the model's linchpin, as all other stereotype threat mediators are hypothesized to affect cognitive performance via working memory.

Working memory, also referred to as executive function or executive control, is responsible for controlling attention and behaviors, effortful processing, and self-regulation to achieve performance goals (Engle, 2002). Working memory allows an individual to remain focused on a given task while concurrently retaining information relevant to other tasks. Working memory is distinctive from short-term memory because although they both share

similar memory processes, working memory has an added element of executive attention which allows an individual to keep information in a state that is ready for use (Engle, 2002; Engle, Tuholski, Laughlin & Conway, 1999). Thus, working memory not only assists an individual in remembering certain information, but also simultaneously performing other tasks that require the processing of other interfering information.

Working memory has been measured with a variety of tasks that share the same underlying structure of asking individuals to use recall, while also performing another task requiring attention (Conway et al., 2005; Engle, 2002). For example, in operation span tasks, individuals are presented a mathematical equation and asked to determine whether it is correct or incorrect (e.g., $[2 \times 3] - 5 = 1$). After the equation, individuals are then presented with a word that they are asked to remember for later. The equations and words are presented in sets of five. Once the sets are completed, individuals are asked to recall the five words previously presented (Turner & Engle, 1989). Participants complete multiple sets during a session. At the end of a session a record is typically made of how many words a participant recalled, the number of correct and incorrect responses to the equation, and the amount of time spent on the different equations. The number of words a participant recalls is the operationalization of working memory capacity. Although seemingly similar to purported “multitasking”, working memory is considered to be the processing and storage capacity that underlies an individual’s ability to effectively switch between tasks (i.e., “multitasking”) rather than being the actual multitasking behavior (Colom, Martinez-Molina, Shih, & Santacreu, 2010)

Working memory is relevant to stereotype threat research because stereotype threat is hypothesized to undermine performance by diverting an individual’s attention from the cognitive task. Stereotype threat is hypothesized to do this by disrupting working memory. Facing

stereotype threat directs an individual's attention away from the task and instead on negative stereotypes and disconfirming these stereotypes. Not surprisingly, a number of studies have demonstrated links between stereotype threat and working memory (e.g., Beilock, Rydell, & McConnell, 2007; Inzlicht, McKay, & Aronson, 2006; Jamieson & Harkins, 2007; Schmader & Johns, 2003). For example, Schmader and Johns (2003) found that both women and Latinos/Latinas performed worse on a working memory task when it was described as a measure of quantitative capacity or general intelligence, than when it was described as a working memory task. Additionally, in a follow-up third study Schmader and Johns demonstrated that decreases in working memory mediated the relationship between stereotype threat and math performance for the women in their sample. The findings of the Schmader and Johns study provide evidence for a direct link between stereotype threat, working memory, and test performance. Furthermore, other studies have investigated the effects of stereotype threat on working memory using different operationalizations and have found similar effects (e.g., Croizet et al., 2004). The resources provided by working memory are especially needed for complex tasks that require the ability to identify the relevant information to solve problems (Schmader et al., 2008). Correspondingly, the effects of stereotype threat are only found when stereotyped groups are solving complex problems and not easy ones (e.g., Specter et al., 1999).

Working memory has been identified as an important causal mechanism in the stereotype threat-performance relationship because of its link to cognitive task performance (Beilock, Rydell, & McConnell, 2007; Kane et al., 2004). However, its link to stereotype threat may be through other mechanisms (Schmader et al., 2008; which are discussed in subsequent sections). Thus, although it is posited here that there is a relationship between stereotype threat and working memory, there are likely other mediators that are more proximal to the stereotype threat

experience that may have stronger relationships with stereotype threat. However, the link between stereotype threat and performance is still an important link to consider. Consequently, it is proposed that individuals exposed to stereotype threat will have a decreased working memory capacity (Path *c*).

Hypothesis 4: Working memory capacity will be lower in stereotype threatened individuals than individuals not exposed to stereotype threat.

The following sections discuss additional mechanisms through which stereotype threat affects performance. However, it is important to keep in mind that working memory is considered to be a central component of stereotype threat's connection to test performance. Thus, although working memory is grouped with the cognitive mediators, it is also linked to other types of mediators including the other cognitive mediators. The cognitive and affective mediators of stereotype threat are believed to impact performance through their relationship with working memory. For example, Eysenk et al. (2007) proposed that anxiety affects test performance because anxiety causes decreased processing efficiency. Correspondingly, processing efficiency is dictated by the central executive component of an individual's working memory system. Processing inefficiency occurs when anxiety impairs components of the central executive system, thus the tie between anxiety and performance is dependent on anxiety's effect on working memory. Altogether, working memory's link with other mediators suggest it is an important component to stereotype threat's effects on performance, independently and in conjunction with other stereotype threat effects.

Distracting (Negative) Thoughts. Schmader et al.'s (2008) model recognizes the impact stereotype threat may have on an individual's thoughts. Particularly, Schmader et al. proposed that stereotype threat may harm cognitive performance through two separate, but related thought

processes, monitoring and suppression. A discussion of how each is connected to stereotype threat is presented next.

Monitoring Processes. The first mediating thought process proposed by Schmader et al. (2008) to be linked to stereotype threat is self-monitoring. Bandura (1991) asserts that self-regulatory processes, like self-monitoring, account for most purposeful actions. Successful self-regulation is dependent on self-monitoring as it provides the evaluation of performance which guides and influences future behavior (Bandura, 1991). While intuitively self-monitoring sounds like a useful behavior to exhibit in a performance episode in a stereotype threat situation, it may have a debilitating effect on working memory and subsequently cognitive task performance (Beilock et al., 2007).

Imbalance between concepts of the self, group, and task domain, set in motion by a stereotype threat experience, trigger the self-monitoring process (Schmader et al., 2008). For example, when a woman is highly identified with the math domain (link between self and domain) but is aware of the negative stereotype about women and lowered math ability (imbalance between self and group), a process to disconfirm and reverse that link between the self and group begins. This process may be evident in a person distancing himself/herself from attributes associated with the group (Steele & Aronson, 1995). Additionally, it is evident in a person's high motivation to avoid failure and thus confirming the group stereotype. The motivation to avoid failure leads to a state where targets of negative stereotype are hyper-vigilant in monitoring their behaviors to detect mistakes and any signs of failure. The self-monitoring process is not limited to performance evaluations. Other evaluations include those an individual makes regarding their own beliefs and privately-held stereotypes. Correspondingly, their thinking on these topics influences the aspects of the performance event to which an individual

pays attention. For example, when a negative stereotype is cued, an individual has made an evaluation of the stereotype which then informs how they appraise their own performance. Thus, when a stereotype threat is triggered, an individual is more focused on errors rather than indicators of success because failures not successes are what inform the negative stereotype they may be trying to disconfirm. The self-monitoring response is not efficient because an individual must attend to a set of questions as they complete their task, “‘Will I do well, consistent with my personal link to the domain?’ or ‘Will I do poorly, consistent with the negative link in the domain suggested by the stereotype?’” (p. 343, Schmader et al., 2008).

Research has provided some support for the role of self-monitoring effects in stereotype threat situations. Beilock et al. (2007) asked women completing complex math problems about their thoughts while completing the problems. They found that women in the stereotype threat condition reported being more worried and conscious of their performance than women completing the same problems in the control condition.

Individuals under stereotype threat are not only monitoring their task-related behaviors, mistakes, and overall performance for signs of failure, they are also staying attuned to how they are coping with the situation. For example, Johns et al. (2008) found that women preparing to take a math test were more attuned to anxiety-related words. Additionally, the self-monitoring process is cyclical because as individuals recognize their own anxiety, they then may interpret it as a signal of poor performance continuing the cycle of looking for failure (Schmader et al., 2008).

Suppression Processes. By definition stereotype threat is a threat to a person’s ego as they face the fear of confirming a negative stereotype associated with some part of their identity (Steele & Aronson, 1995). As mentioned previously, with threat also come feelings of doubt,

stress, and anxiety and thus the need to cope with these new feelings. The previous section noted how these feelings of anxiety and stress may then go into a feedback loop that signal performance failure to the individual trying to avoid confirming a negative group stereotype. Thus, attempting to suppress these negative emotions and feelings may be one way that individuals cope with the threatening situation (Johns et al., 2008). Suppression involves active attempts at controlling negative thoughts and feelings.

Research related to stereotype threat's effects on anxiety can inform stereotype threat's links to cognitive suppression effects. Notably, the research on stereotype threat and anxiety when using explicit self-report measures has produced mixed results. In contrast, using indirect measures to assess anxiety (e.g., non-verbal cues, physiological responses) reveal a clearer picture of the threat-anxiety relationship. One hypothesized reason for the mixed findings, when using the explicit measures, is that individuals may be attempting to deny their anxious feelings, even to themselves (Schmader et al., 2008). Suppression is the process of directing effort to regulate negative and unwanted thoughts and emotions which accompany experiences of threat (Johns et al., 2008). When individuals attempt to deny anxious feelings, the suppression process is set into motion. In contrast to helping an individual cope, the distractive nature of suppression is more harmful than helpful to individuals' performance on cognitive tasks.

The harm caused by suppression is twofold. First, the act of suppressing thoughts is an effortful behavior that exhausts an individual's cognitive resources (e.g., Muraven & Baumeister, 2000; Schmeichel, 2007). Secondly, actively suppressing negative anxiety-related thoughts make the negative thoughts more accessible for the individuals (Wegner, Erber, & Zanakos, 1993).

Both monitoring and suppression thought processes are designed to be helpful for the individuals who engage in them as they help provide a defense from negative thoughts and also

help an individual continue to improve himself/herself. However, in performance situations where stereotype threat has been triggered, monitoring and suppression processes introduce thoughts that are unrelated to completing the task and subsequently hinder performance.

Detecting stereotype threat's effects on monitoring and suppression processes is a difficult task, particularly in the case of suppression. The purpose of the suppression act is to repress the detection of threat effects, thus suppression and threat effects are difficult to disentangle from each other. Johns et al. (2008) used an implicit measure of anxiety to measure suppression. The authors hypothesized that when the experimental participants were told that the implicit measure was designed to measure anxiety, individuals under stereotype threat would make an attempt to avoid anxiety-related words (i.e., suppress anxious thoughts). However, individuals who were under stereotype threat who were not told about the purpose of the measure were hypothesized to be more attuned to the anxiety-related stimuli. Thus, an opposing pattern of results was hypothesized to be exhibited by individuals under stereotype threat depending on the experimental anxiety condition. The experimental manipulations used to detect suppression effects make it somewhat difficult to determine the true population effect because (1) there is typically no control group that is not exposed to stereotype threat, and (2) the stereotype-threatened group is not a pure test of stereotype threat because there is typically an additional manipulation of the anxiety prime. Due to these complexities the current meta-analysis first focuses on the overall effect of stereotype threat on distracting thoughts. However, secondary hypotheses seek to better understand the conditions for which suppression versus monitoring effects may emerge.

Generally, both monitoring and suppression processes are evidence of less than optimal strategies for managing stereotype threat. Although both strategies are designed to help

individuals focus on performance instead of other distractions, in fact negative distracting thoughts are likely more accessible in a stereotype threat scenario (Path *d*).

Hypothesis 5: Stereotype threatened individuals will have more distracting (negative) thoughts than individuals not exposed to stereotype threat.

However, because of the complex nature of monitoring and suppression effects, the emergence of these different strategies is likely dependent on the situation. Self-monitoring suggests that individuals are likely alert for signs of judgment and are monitoring their behaviors for signs of failure and staying attuned to how they are coping with the situation. These monitoring behaviors are likely to be most active in more subtle stereotype threat situations where these threat signs are harder to detect and less clear. Consequently, threatened individuals are more vigilant as they stay alert for these signals.

Hypothesis 6: When stereotype threat is activated using subtle cues, stereotype threatened individuals will have more distracting (negative) thoughts than individuals not exposed to stereotype threat.

More blatantly activated stereotype threat situations will likely trigger suppression behaviors because these situations trigger a clear signal that the present context is threatening and possibly deleterious to performance. Thus, when threatened individuals are explicitly made aware of this harm they are likely more motivated to suppress any thoughts that are related to the negative experience. The work to suppress these negative thoughts will likely lead to threatened individuals demonstrating distracting thought levels at the same level as their non-threatened peers.

Hypothesis 7: When stereotype threat is activated using blatant cues, stereotype threatened individuals and individuals not exposed to stereotype threat will have a similar number of distracting (negative) thoughts.

Motivational Process

There are a number of other possible mechanisms underlying stereotype threat effects that are not tied to executive control functioning and linked to working memory (Schmader et al.,

2008; Smith, 2004). Other motivational focused mediators such as effort, performance orientation, and performance expectations could also play a role in stereotype threat effects. These factors are briefly discussed below.

Effort. Effort is hypothesized to serve as a mediator of stereotype threat in one of two ways. First, an individual could expend less effort as a learned helplessness type of behavior. The negative stereotype signals to an individual that there is little possibility he/she can overcome the negative group stereotype and thus the individual responds by putting little to no effort towards being successful in the domain. Second, effort can affect stereotype threat as an individual feels a determination to prove the stereotype wrong and thus trying so hard that their performance is actually hindered because the effort is not focused. However, and possibly due to these opposing explanations, research has not demonstrated a strong link between effort and stereotype threat in either direction with most studies failing to find any effects (Pennington et al., 2016; Smith, 2004).

A stereotype threatened individual may decrease the effort he/she puts towards a task through self-handicapping behaviors. Self-handicapping allows an individual to protect himself/herself by making external attributions for poor performance (Leary & Shepperd, 1986). For example, a person may claim to have had bad performance on an exam because of loud neighbors who prevented her/him from focusing while they were studying. Test performance for stereotype threatened individuals is hypothesized to be impacted by self-handicapping because it allows the threatened individual to blame something other than an identity-related reason for their failures. Thus, under stereotype threat conditions, minorities and women are predicted to give more self-handicapping reasons to explain their performance. However, although there is some evidence for increased self-handicapping by individuals under stereotype threat (e.g.,

Keller, 2002, 2004; Steele & Aronson, 1995), self-handicapping as a mediator has not yielded conclusive results (Pennington et al., 2016; Smith, 2004).

Jamieson and Harkins (2007) mere effort proposal equates the stereotype threat-performance relationship to the evaluation-performance relationship. That is, a stereotype threat activation is equivalent to a situation informing someone that their performance will be evaluated. Following this line of reasoning, Jamieson and Harkins proposed that similar to findings in the evaluation-performance literature, stereotype threat will only inhibit performance on complex tasks, but not on less complex (easy) tasks. They posit this occurs because when an individual is faced with threat (or evaluation) an individual's effort increases because they are motivated to do well, which in turn increases the likelihood of whatever is the dominant behavior for a given task (Harkins, 2006). For example, on a remote associate task, where participants are asked to find a word that links to three other words (e.g., memory to elephant, lapse, vivid), the participants will likely search for a word that closely relates to one of the triad words. However, the dominant behavior is not always the correct behavior. When the links between the words are more subtle and complex, the dominant behavior is to try and fit a word closely associated with only one of the three words, rather than finding a solution that relates to all three words. Focusing efforts on finding the link between one of the words prevents the activation of the more loosely linked correct answer and reduces performance. When individuals have enough time to correct their error, they can eventually overcome the dominant behavior and direct their effort to finding the correct answer. However, in short-duration or time-pressured tasks (as most high-stakes cognitive tests are), there is not enough time to correct mistakes. That is, a stereotype threat experience increases effort but depending on the complexity and consequently the difficulty of the task that effort may help or hinder performance. Thus, the proposal that effort

differentially affects performance depending on task difficulty provides one potential explanation for the mixed findings of effort as a mediator in stereotype threat conditions. That is, past studies may have relied on results from both easy and difficult tasks and therefore no clear effect for effort in either direction was detected. The effort hypotheses acknowledge that stereotype threat has an effect on effort (Path *e*), but seeks to disentangle these effects.

Hypothesis 8: Stereotype threatened individuals will exhibit increased effort on difficult tests compared to non-threatened individuals.

Hypothesis 9: Stereotype threatened and non-threatened individuals will not differ in levels of effort on easy tests.

The manner in which effort is operationalized may be another reason a clear pattern of effects has not been detected. A number of studies have operationalized effort as the amount of time spent on test items or the latency time before a task decision (e.g., Aronson et al., 1999; Spencer et al., 1999; Steele & Aronson, 1995). However, only Steele and Aronson found any evidence, although it was marginal, for effort as a mediator when it was operationalized this way. Steele and Aronson found that individuals in the stereotype threat condition took longer to complete the first five items of a verbal ability exam. Number of items attempted is another possible proxy for measuring effort (e.g., Johns, Schmader, & Martens, 2003; McIntyre, Paulson, & Lord, 2003; Steele & Aronson, 1995). Completion of more items, even if incorrectly, indicates that individuals were exerting more effort to complete the tasks than individuals who complete fewer items. Self-reported effort is also another way to capture how stereotype threat may impact performance via effortful behavior. However, self-reported findings may be confounded with impression management as individuals do not want to admit to not trying. Taken together, self-report measures may present a less accurate picture of effort exertion than more objective measures used to operationalize effort.

Hypothesis 10: Studies that use self-report measures of effort will produce smaller mean effect sizes for effort than studies that use more objective (e.g., items completed) measures.

Performance Orientation. A number of stereotype threat models have been built on the premise of performance orientation factors being key mechanisms underlying stereotype threat effects (Seibt & Foster, 2004; Smith, 2004). These models have in common the proposition that stereotype threat activates an avoidance or disengagement mindset. The sections that follow discuss how stereotype threat is responsible for disengagement and how it subsequently impacts performance.

In her review of stereotype threat mediators, Smith (2004) concluded that the research searching for mediators of stereotype threat effects could not identify a “reliable” link that explained the effects. Thus, Smith proposed the Stereotype Threat Engagement Process (STEP) model as a way to understand how stereotype threat impacts performance. In the STEP model, Smith proposes that goal-orientation is the catalyst that triggers the performance decrements in stereotype threat situations. Specifically, she states that the stereotype threat situation causes an individual to adopt a performance-avoidance goal orientation, rather than a performance-achievement goal orientation. Stereotype threatened individuals adopt performance-avoidance goal orientations because they are motivated to demonstrate that the negative stereotypes about a group are not relevant to them. In support of her model, Smith (2006) found that women under stereotype threat conditions had lower math performance expectations and that this effect was mediated by the women’s endorsement of an avoidance goal orientation. Additionally, Smith, Sansone, and White (2007) found that women who were highly motivated to be successful were more likely to adopt an avoidance goal orientation and the adoption of the avoidance goal orientation was associated with less interest in computer science. Brodish and Devine (2009)

provided additional support for this model. In their study they found that women who endorsed performance avoidance goals, reported increases in anxiety and worry when exposed to stereotype threat and subsequently performed worse on a math test. These two studies together provide evidence for the link between stereotype threat and goal orientation and presents a case for goal orientation as a possible explanation for stereotype threat effects.

In addition to the STEP model, the regulatory-focus perspective on stereotype threat provides a similar explanation for stereotype threat effects. The regulatory-focus perspective states that people either adopt a promotion or prevention focus and they are most successful when their focus and reward system align (Grimm, Markman, Maddox, & Baldwin, 2009; Higgins, 2000). Similar, to the description in the STEP model, from a regulatory-perspective the stereotype threat situation invokes a prevention or avoidance focus. As discussed previously, individuals become more vigilant in stereotype-relevant situations. This vigilance makes individuals more cautious and rather than being focused on achieving and accomplishing goals, they are more focused on trying to avoid failures, thus signaling a prevention focus (Seibt & Förster, 2004).

The regulatory fit model takes the explanation of stereotype threat further than the STEP model by also adding a proposed solution to stereotype threat through targeting regulatory focus. In the regulatory fit model, performance decreases in stereotype threatened individuals are proposed to be a result of the misalignment of goals and rewards (Grimm et al., 2009; Seibt & Förster, 2004). Thus, because stereotype threatened individuals are prevention-focused, rewards aligned with the prevention focus, rather than the promotion focus, function as a way to alleviate some of the stereotype threat effects.

There has been notable research that has provided some support for the regulatory-focus model (e.g., Barber & Mather, 2013; Grimm et al., 2009; Seibt & Förster, 2004). For example, Grimm et al. (2009) designed a study to test the efficacy of the regulatory fit model. The authors created two different conditions, one where participants gained points for correct responses on GRE quantitative problems and the other where participants were informed that they would lose points for incorrect responses. Grimm et al. hypothesized that women would perform better in the loss condition because it was more aligned with a prevention focus that women under threat would adopt. As predicted, women in the loss condition scored significantly higher than women in the gains condition. Barber and Mather (2013) also obtained similar results when they tested the regulatory-focus model in two separate samples of older adults. When participants under stereotype threat were rewarded based on a prevention-focused structure (i.e., losing poker chips for incorrect answers) they performed better on a working memory task than individuals in the promotion-focused reward system (i.e., receiving poker chips for correct responses). What is important to note here is that the efficacy of stereotype threat interventions are dependent on the mechanism that is targeted. This points to the criticality of knowing what mechanisms are operating and why, when investigating stereotype threat effects. Although intervention strategies are not the focus of the current study, theories to understand the motivational processes underlying stereotype threat are critical to addressing stereotype threat effects. Therefore, the following hypothesis focuses on the mechanism underlying stereotype threat effects related to avoidance behaviors (Path *f*).

Hypothesis 11: Prevention-focus and avoidance orientation will be higher in stereotype threatened individuals compared to non-threatened individuals.

Performance Expectations. Performance expectations is another motivational factor proposed to be impacted by stereotype threat. Performance expectations are inherent in a

stereotype threat setting; stereotype threat signals to an individual that they are not expected to perform well on the task at-hand, and that signal in turn shapes the individuals' expectations of themselves and thus impacts performance. Performance expectations are a notable predictor of future performance (Kovenklioglu & Greenhaus, 1978). By lowering performance expectations, individuals trigger a self-fulfilling prophecy process, where they can only achieve to the level they expect.

Studies that have investigated whether performance expectations serve as mediators of the stereotype threat-performance relationship have obtained mixed results. In one operationalization of performance expectations, Steele and Aronson (1995) measured individual's confidence in their abilities (Study 1), as well as employing an implicit measure to capture the degree of self-doubt threatened individuals were feeling (Study 3). African American's in the stereotype threat condition reported the most feelings of self-doubt; however, performance was not measured and thus no test for mediation could be performed.

Other researchers have employed a variety of other strategies to measure individuals' performance expectations, both before and after the completion of a task. The timing of the performance expectations measure has been demonstrated to impact study results. Stangor, Carr, and Kiang (1998) found that when female participants were asked to estimate their performance on a spatial-ability task, participants in the stereotype threat condition reported lower performance expectations than participants in the control condition. However, both Keller (2002) and Aronson et al. (1999) asked study participants to estimate how they think they did on a math test after the participants had already completed the problems, and found no difference between participants exposed to stereotype threat and those in the control condition. Thus, a mixed picture

emerges for the effects of stereotype threat on performance expectations depending on the timing of the expectation measure.

Test performance is not always included as the dependent variable in studies investigating the effects of stereotype threat. For example, Stangor, Carr, and Kiang (1998) measured performance expectations, rather than actual performance, to evaluate stereotype threat effects. Consequently, in these types of studies no mediation analyses linking the expectations to performance outcomes were conducted.

Performance expectations have also been offered as one way to address the negative impact of stereotype threat. Some researchers have tried providing positive feedback to shape performance expectations and subsequently improve performance as a way to counter the effects of stereotype threat (Candinu et al., 2003; Stangor et al., 1998). However, findings have highlighted that it is important that the feedback match the domain of the stereotype threat for it to have positive effects (Candinu et al., 2003). Taken together, performance expectations as proposed mediators of stereotype threat effects present a likely avenue through which stereotype threat impacts performance (Path *g*); however, it is important to note that the timing of performance expectancy measures may impact the findings.

Hypothesis 12: Performance expectations will be lower in stereotype threatened individuals compared to non-threatened individuals.

Hypothesis 13: Studies that measure performance expectations prior to the completion of the task will produce larger mean effect sizes than studies that measure performance expectations after task completion.

Moderators of the Stereotype Threat Mediators

This section includes a discussion of the moderators that have been investigated across all of the mediators proposed in the previous sections. In the preceding sections a few moderators have been proposed for some of the specific mediator variables (e.g., measure type and anxiety).

The moderators considered below are those that are purported to impact the relationship between stereotype threat and all of the proposed mediators, rather than those specific to the relationship between stereotype threat and one mediator (e.g., mediator measurement timing and distracting thoughts).

Stereotype Threat Target Moderators. The first proposed moderator is the target of the stereotype. In their meta-analysis on the stereotype threat-performance relationship, Nguyen and Ryan (2008) found that the effect of stereotype threat on performance differed depending on whether it was gender-based math stereotype or a race-based intelligence stereotype. In general, the effect sizes for race-based minorities were larger than the effect sizes for women thus, stereotype threat negatively affected racial minorities more than women.

Additionally, the strength of the relationship between stereotype threat and performance was different for women and racial minorities depending on the activation cue. Subtle threat activation hurt women's performance more than blatant activation. In contrast, moderately explicit activation had the worst effects on performance in comparison to explicit or subtle activation for minority test takers. Nguyen and Ryan (2008) proposed that the differences in stereotype threat effects for women and ethnic-minority test takers may be a result of different mechanisms underlying stereotype threat experiences for the groups. For example, when minority test takers were presented with an explicit message to remove stereotype threat effects (e.g., telling minority test takers they typically perform better on the cognitive ability test than majority test takers), the message did not improve test performance; however, when women were presented with these types of messages their performance improved. There are a number of plausible explanations for these different effects. First, minority test takers may feel a pressure to confirm the positive stereotype when they are presented with these types of explicit messages,

and subsequently perform worse because of the added pressure, such as in the model minority phenomenon (Cheryan & Bodenhausen, 2000). On the other hand, it may be that the gendered math stereotype is not considered as negative as the racial intelligence stereotype. Thus, when a message is presented refuting the stereotype that a group is inferior in ability (e.g., mathematics ability, cognitive ability), it is easier for women to accept it than racial minorities. Lastly, the long history of racial prejudice and discrimination may lead racial minorities to be more sensitive and reactive to negative stereotypes and the likelihood of being evaluated based on those stereotypes.

Stereotype threat effects have also been investigated in the older adult population (Lamont et al., 2015), although to a lesser extent than its effects on women and racial minorities. The research exploring stereotype threat's effects on older adults is based on the stereotype that memory and cognitive competence decline as people age (Cuddy & Fiske, 2002). Similar mediators to those investigated for race and gender-based stereotype threat have been explored for age-related stereotype threat (e.g., Chasteen, Bhattacharyya, Horhota, Tam, & Hasher, 2005; Hess, Auman, Colcombe, & Rahhal, 2003; Hess, Hinson, & Hodges, 2009). Similar to the difference between race and gender-based experiences of stereotype threat, the age-based stereotype is not associated with the life-long discrimination and prejudice linked to racial minorities. Recent research by Lamont et al. (2015) quantified the effects of age-based stereotype threat and found it to have a small-to-medium effect with *d*-values ranging from 0.28 to 0.52, depending on various moderators.

Taken together, the aforementioned reasons and past results suggest that age, gender, and race-based stereotype threat produce different experiences of threat, thus it is to be expected that there may be different mediating processes acting in these different experiences. Specifically, it

is proposed that the relationship between stereotype threat and the proposed mediators will be moderated by the target of the stereotype, such that racial minorities will be impacted by stereotype threat to a greater extent than women or older adults. This will be evidenced by larger differences between stereotype threatened and non-stereotype threatened groups on the various mediators for racial minorities than for women and older adults (Path *h*).

Hypothesis 14: Mediator effect size differences will be larger for stereotype threatened racial minorities than stereotype threatened women and older adults.

Stereotype Activation Moderators. Similar to Nguyen and Ryan's (2008) meta-analysis, the second proposed moderator is the stereotype activation cue. Stereotype threat has been manipulated in lab studies in a multitude of ways—from having individuals indicate their race before a cognitive exam (Steele & Aronson, 1995) to having women interact with blatantly sexist men (Logel et al., 2009). These different activation strategies have differential effects on performance, however the relationship between stereotype threat-activation cues and performance is not linear. Stereotype threat is hypothesized to be detrimental because it activates a number of mediators at an unconscious level (e.g., suppression, self-monitoring). Subsequently, subtle levels of stereotype threat activation, such as priming individual's membership in a particular group (e.g., African American, women) before a test, are considered to lead to worse performance than when more blatant activation strategies are used, such as explicitly telling test takers that women or racial minorities typically do not do well on the test. However, Nguyen and Ryan (2008) found that the level of activation differentially affected women and racial minorities. Their findings revealed that women had worse outcomes for subtle activation cues than moderate or explicit stereotype threat activation. On the other hand, moderate activation cues lead to the worse performance outcomes for racial minorities. In line

with these findings, it is proposed that the different forms of activation will have different effects on mediators for different groups (Path *i*).

Hypothesis 15: For women, the mediator effect sizes will be largest when stereotype threat is activated using subtle cues.

Hypothesis 16: For racial minorities, the mediator effect sizes will be largest when stereotype threat is activated using moderate cues.

A summary of all the study hypotheses is presented in Table 1.

Table 1
Summary of Hypotheses

Number	Hypothesis
1	Anxiety will be higher in stereotype threatened individuals.
2	Physiological arousal will be higher in stereotype threatened individuals.
3	The mean effect size for physiological arousal will be larger than the effect size for anxiety for stereotype threatened individuals.
4	Working memory capacity will be lower in stereotype threatened individuals than individuals not exposed to stereotype threat.
5	Stereotype threatened individuals will have more distracting (negative) than individuals not exposed to stereotype threat.
6	When stereotype threat is activated using subtle cues, stereotype threatened individuals will have more distracting (negative) thoughts as individuals not exposed to stereotype threat.
7	When stereotype threat is activated using blatant cues, stereotype threatened individuals and individuals not exposed to stereotype threat will have a similar number of distracting (negative) thoughts.
8	Stereotype threatened individuals will exhibit increased effort on difficult tests compared to non-threatened individuals.
9	Stereotype threatened and non-threatened individuals will not differ in levels of effort on easy tests.
10	Studies that use self-report measures of effort will produce smaller mean effect sizes for effort than studies that use more objective (e.g., items completed) measures.
11	Prevention-focus and avoidance orientation will be higher in stereotype threatened individuals compared to non-threatened individuals.
12	Performance expectations will be lower in stereotype threatened individuals compared to non-threatened individuals.
13	Studies that measure performance expectations prior to the completion of the task will produce larger mean effect sizes than studies that measure performance expectations after task completion.
14	Mediator effect size differences will be larger for stereotype threatened racial minorities than stereotype threatened women and older adults.
15	For women, the mediator effect sizes will be largest when stereotype threat is activated using subtle cues.
16	For racial minorities, the mediator effect sizes will be largest when stereotype threat is activated using moderate cues.

THE PRESENT STUDY

The purpose of the present study is to examine the mediators of the stereotype threat-performance relationship. Mediators are “the generative mechanisms through which the focal independent variable is able to influence the dependent variable of interest” (Baron & Kenny, 1986, p. 1173). The independent variable influences changes in the mediator, which then results in fluctuations in the outcome variable. In the present study, the independent variable is stereotype threat, the dependent variable of interest is cognitive test performance, and the mediators are those mechanisms—specifically affective reactions, and cognitive and motivation processes—investigated in past stereotype threat research. Following Baron and Kenny's four-step approach to establishing the causal chain of a mediator relationship, a number of conditions regarding the relationships between stereotype threat, the proposed mediators, and cognitive test performance must be met (Figure 3). The necessity of including all steps in Baron and Kenny's causal chain approach has been questioned (MacKinnon & Fairchild, 2009; O'Rourke & MacKinnon, 2018); however, the model provides a useful framework for understanding the different components within the mediation process and will be used to as a guide for the current work. Specifically, the current meta-analysis seeks to elucidate the stereotype threat literature by estimating the effects of the various proposed mediators of the stereotype threat-performance relationship.

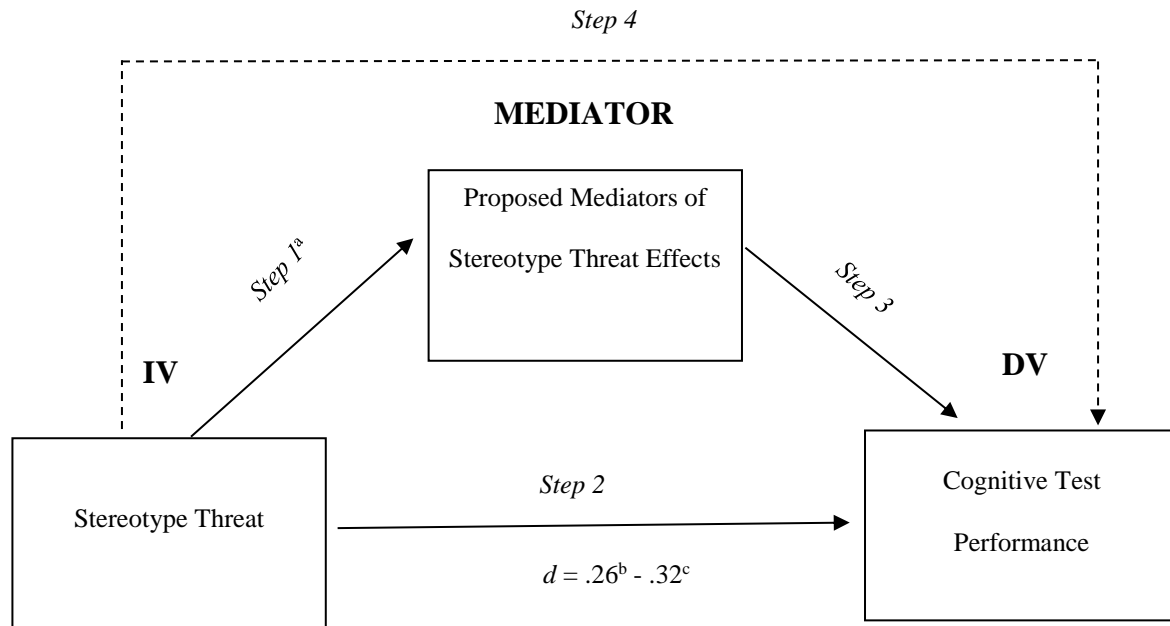


Figure 3. A step-by-step model to test the mediators of stereotype threat a. The current meta-analysis is focused on Step 1 in this model. b. Overall effect size from Nguyen and Ryan (2008). c. Overall effect size from Lamont et al. (2015)

The first condition in Baron and Kenny’s (1986) approach is establishing that a relationship between the independent variable (IV) and the mediator variable exist. Since a number of different mediators of stereotype threat effects have been proposed, examining the magnitude of these relationships (IV-mediator relationship) is the purpose of the present study (Step 1 in Figure 3). Effect sizes will be calculated for the different proposed mediators of stereotype threat. These effect sizes will provide information on the relative strength of stereotype threat’s effect on the various proposed mediators and will provide evidence as to whether the different proposed mediators are mechanisms via which stereotype threat impacts performance.

The second condition to be met to establish mediation, is verifying that there is a relationship between the between the IV and the dependent variable (DV). For the purpose of this study, this is the relationship between stereotype threat and performance (Step 2 in Figure 3).

Although mediation effects may exist in the absence of the IV to DV effect (O'Rourke & MacKinnon, 2018), past research has demonstrated that this condition is satisfied in the context of stereotype threat effects. Specifically, Nguyen and Ryan's (2008) and Lamont et al.'s (2015) meta-analyses established the existence of this relationship. The results of their analyses revealed an overall mean effect size of $d = 0.26$ and 0.38 for the stereotype threat-performance relationship.

After meeting condition two, the third condition is confirming the relationship between the mediators and the DV. In the present study, the third condition requires determining the link between the proposed affective, cognitive, and motivational mediators' effects and cognitive test performance (Step 3 in Figure 3). Although important, the links between the proposed mediator variables and performance outcomes are not the focus of the dissertation and so no effect sizes will be computed. This link is not explored for a number of reasons. First, although each mediator is operationalized similarly across studies, the performance variables in the studies are not oftentimes similar (e.g., math GRE for women, intelligence test for African Americans, etc.), making it difficult to quantify these effects across studies. Additionally, a number of studies do not include measures of performance and only present results for the stereotype threat-mediator relationship (e.g., performance expectations, working memory capacity). Finally, the mediators included in the present study were all initially proposed because of their demonstrated link to performance in studies unrelated to stereotype threat. Consequently, in lieu of computing effect sizes for step three, the current dissertation defers to the extant literature of each mediator as support for the mediator-DV relationship (e.g., Oberauer, Schulze, Wilhelm, & Süß, 2005; von der Embse, Jester, Roy, & Post, 2017).

The final condition in Baron and Kenny's (1986) mediation approach is the examination of the IV to DV relationship after accounting for the mediators. If the IV to DV relationship is no longer significant or its magnitude is significantly decreased, then there is evidence for full or partial mediation. In the present study, the final condition is met when the stereotype threat-performance relationship is no longer significant after accounting for the effects of the proposed affective, cognitive, and motivational mediators (Step 4 in Figure 3). However, in light of Step 3 not being calculated in the current study, the necessary data to test the last relationship are not available. Although this step will not be tested empirically in the present study, it will be considered in the implications and discussions section and is an important future direction for this stream of research.

METHOD

An extensive literature search was conducted to identify relevant studies. First, the process consisted of a computerized search of literature databases, *PsycINFO*, *EBSCO*, and *ProQuest* using search terms to include all mediators, (e.g., “stereotype threat AND mediators”, “stereotype threat AND anxiety”) until all mediators were searched. The full list of search terms are presented in Figure 4. The initial search yielded books, book chapters, peer-reviewed journal articles, thesis, and dissertations related to stereotype threat from 1995 (the publication year of Steele & Aronson seminal article) to 2016 (the year when data collection was completed for this study).

Second, *Google Scholar* was used to search for other articles and materials that may not be published. The *Google Scholar* search returned over 30,000 hits, some of which overlapped with findings from the electronic database search. *Google Scholar* search results were sorted based on relevance to the search terms, and the most relevant items (top 1,000 hits) were reviewed and any original entries were included. Additionally, when authors were contacted to gather necessary data not presented in the available papers, these authors were asked to provide any unpublished articles relevant to the current topic. Lastly, Nguyen and Ryan’s (2008) stereotype-threat meta-analysis references were manually reviewed to ensure a comprehensive list to review was compiled. The search yielded 3,372 hits to be reviewed for potential inclusion in the current meta-analysis.

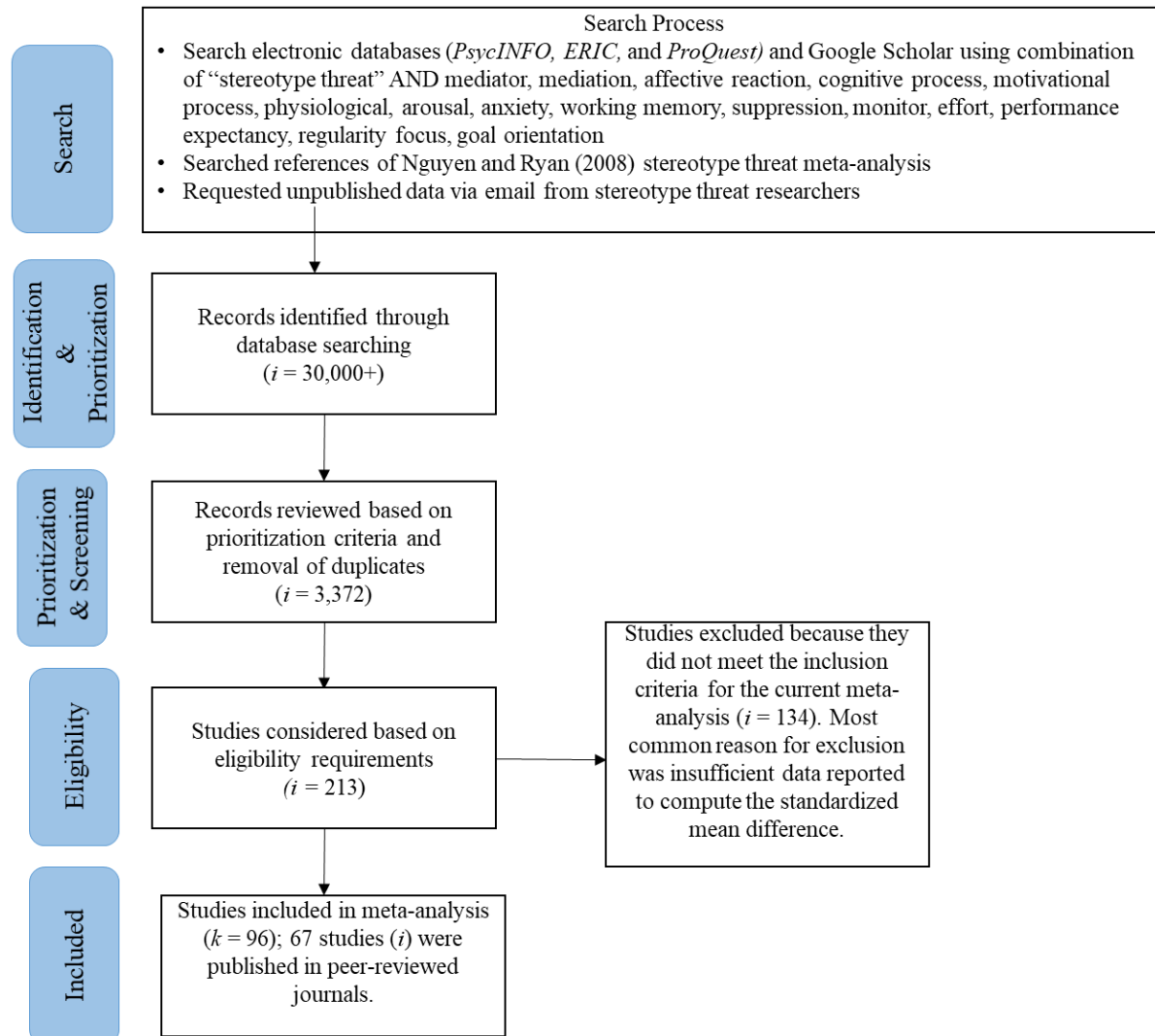


Figure 4. Flowchart outlining the literature search and screening process for the included studies. i = number of sources.

Inclusion Criteria

A number of criteria had to be met for a study to be included in the meta-analysis. These inclusion criteria are described below and summarized in Table 2.

Table 2
Inclusion Criteria Table

Criteria	Definition/Example
1 Is the study designed to test the stereotype threat hypothesis? Does the study include at least two groups, and is one group a clear control group (not exposed to stereotype threat)?	A test of the hypothesis includes a measure of some outcome variable for two groups after one group has been exposed to stereotype threat. The study must include a group affected by stereotype threat (either a constant or primed stereotype) and a group not affected by the stereotype (either through mitigation or lack of exposure).
2 Does the study include a measure of one of the proposed affective, cognitive, or motivational mediators?	The study does not have to include a cognitive ability measure but must include a measure of one of the proposed mediators. For example, a study measuring an affective mediator must include a measure of anxiety or physiological arousal.
3 Is the negative stereotype gender, race, or age based?	Other group stereotypes (e.g., SES) were excluded.
4 Does the study include the effect size or necessary data to compute a sample-weighted <i>d</i> ?	Necessary data includes the effect size (<i>d</i>), sample size, mean, standard deviations, correlation estimates, independent sample t-test estimates, or F-test estimates.
5 Is the study in English or available in English?	Studies that were not available in English were excluded.

Stereotype Threat Paradigm. Only studies designed to test the stereotype threat hypothesis were included in the current meta-analysis. That is, the study must have created an environment where stereotype threat is expected to be operating (e.g., a testing situation where a negative ability stereotype is present for one of the involved groups). Since stereotype threat is operationalized as a categorical variable and not continuous, the study must include (at least) two

groups, one exposed to stereotype threat and one not exposed to stereotype threat. Specifically, there should be one group exposed to a stereotype threat-activated (experimental) environment where stereotype threat had been introduced or manipulated. There should also be a group where stereotype threat had not been activated or where it had been removed (control). The stereotype threat environment could be created in a subtle manner (e.g., simply a testing situation; Schmader & Johns, Study 2) or more blatant manner (e.g., purposefully drawing attention to a negative stereotype; Abrams et al., 2008).

When investigating the stereotype threat phenomenon, generally members of a stereotyped group are randomly assigned to a control or threat condition and analyses are conducted to determine if their mean scores on the measure of interest (e.g., test performance) are different. Figure 5 displays the different groups in between-group and within-group designs. In some studies (e.g., Steele and Aronson, 1995), a stereotyped group for whom the stereotype threat should not be relevant (e.g., Whites, men, young adults) are included as a comparison group (Groups B and D in Figure 5). A within-group design is when members of the same stereotyped group who only differ because of their exposure to the threat are compared to one another (e.g., stereotype-threatened women [Group A in Figure 5] vs. non-stereotype threatened women [Group C in Figure 5]). A between-group design is when the stereotype-threatened group (e.g., women exposed to math-based stereotype threat [Group A in Figure 5]) is compared to the non-stereotyped comparison group (e.g., men exposed to the math-based stereotype-threat [Group B in Figure 5]). The analyses in the current meta-analysis looked at results for both between-group and within-group designs. The critical component is that there must be a clear control group where stereotype threat is absent (e.g., group not exposed to stereotype threat) or a group that is not expected to be impacted by stereotype threat (e.g., White participants). It was

expected that the results would be similar across both within-and between-group designs with the magnitudes for the between-group effect sizes being larger. It should be noted that focus on within- and between-group design results is distinct from within-subject and between-subject designs. The latter speaks to the distinction between repeated measurements of individuals based on exposure to an experimental condition versus single measurement of individuals but with exposure to different experimental conditions. The current focus on within- and between-group designs only constitutes one measure of the mediator but with different comparison groups.

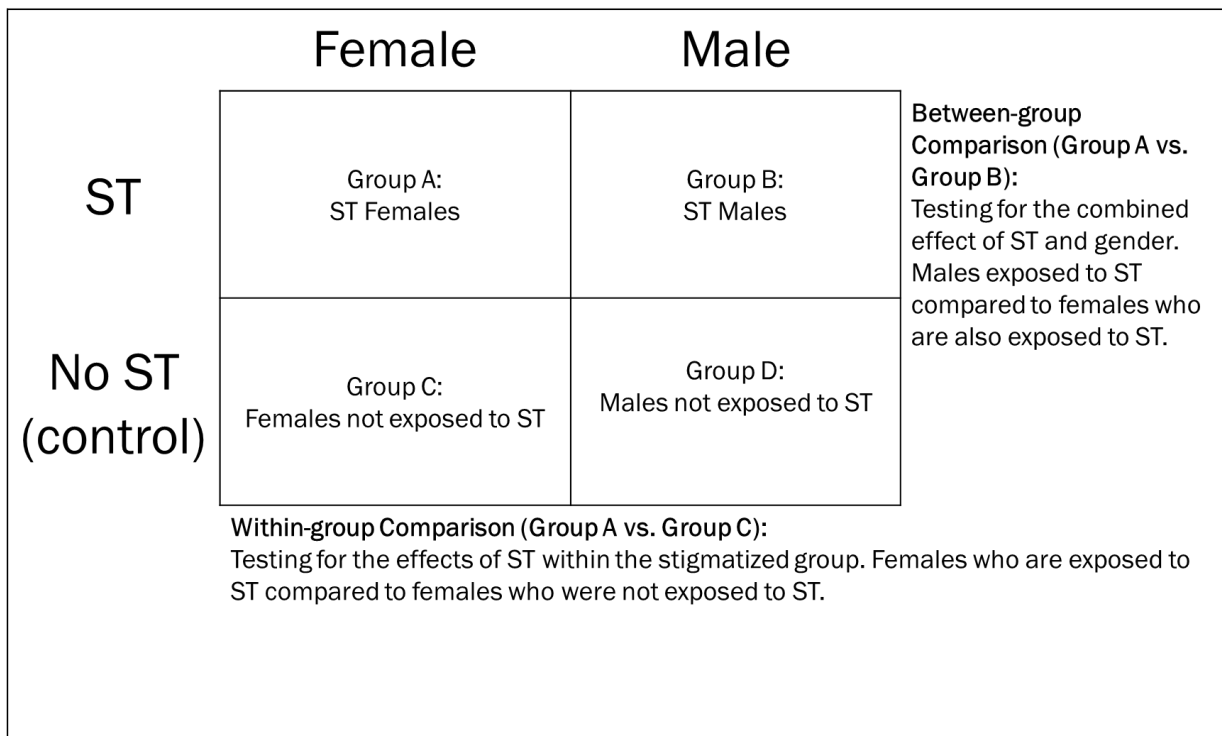


Figure 5. Comparison of between-group and within-group designs.

Measurement of Mediator Domains. The current study is focused on the extent to which stereotype threat impacts a number of proposed mediators. The study will only focus on mediators outcomes, since there is a clear documented effect of stereotype threat on cognitive ability measures (Lamont, Swift, & Abrams, 2015; Nguyen & Ryan 2008) the presence of these

types of measures are not required for the current meta-analysis. Nonetheless, the study must measure stereotype threat effects on one of the proposed mediators (affective reactions, cognitive processes, motivational processes). For example, Spencer et al.'s (1999) study that measured stereotype threat's effect on anxiety was included as it measures an affective effect of stereotype threat. Additionally, the mediator must be distinctly and consistently operationalized in both the stereotype threat and non-stereotype threat condition and must not be the result of an intervention or only measured after a manipulation that is not stereotype threat (e.g., thought suppression instructions, promotion/prevention focus instructions; Keller & Bless, 2008). The mediators included in the current meta-analysis represent effects in three broad domains, affective, cognitive, and motivational. Within each of these domains, a number of mediators were explored. Each of these mediators are defined in Table 3.

Table 3
Mediator Description

Category	Mediator	Definition	Example Measures	Example Study
Affective	Anxiety	Tense, nervous feelings elicited by the stereotype threat manipulation	<ul style="list-style-type: none"> ▪ Worry, emotionality scales ▪ State anxiety measures ▪ Affect measures (e.g., PANAS) 	Abrams et al. (2008)
	Physiological arousal	Physical symptoms of arousal elicited by a stereotype threat situation. “Distress” signals to activate vigilance and control behaviors (Forbes, Schmader, & Allen, 2008). This is evidenced by increased activation of the sympathetic nervous system.	<ul style="list-style-type: none"> ▪ Blood pressure measures ▪ Testosterone 	Allen & Friedman (2016)
Cognitive	Distracting (negative) thoughts	Distracting thoughts elicited by the stereotype threat manipulation	<ul style="list-style-type: none"> ▪ Measures of distracting thoughts ▪ Measures of stereotype threat activation (after manipulation) ▪ Evaluation apprehension 	Beilock, Jellison, Rydell, McConnell, & Carr (2006)
	Working memory	Individual’s capacity to keep information in a state that is ready to use	<ul style="list-style-type: none"> ▪ Working memory tasks (e.g., digit span tasks, vowel counting tasks) 	Schmader & Johns (2003)
Motivational	Effort	Individual’s drive and motivation to complete a task	<ul style="list-style-type: none"> ▪ Number of items attempted ▪ Self-reported effort 	Steele & Aronson (1995)
	Prevention / avoidance focus	Disengagement/orientation to avoid errors and mistakes on the stereotype threat task	<ul style="list-style-type: none"> ▪ Prevention-focus measures ▪ Avoidance goal-orientation measures 	Keller & Dauenheimer (2003)
	Performance expectations	Belief of how well an individual will perform on the stereotype threat task	<ul style="list-style-type: none"> ▪ Task-specific self-efficacy measures (e.g., I believe that I am capable of performing well on this test.) ▪ Self-report performance expectations (e.g., estimate of test performance) 	Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti (2003)

Negative Race, Gender, and Age-Based Ability Stereotype Threat. The negative stereotype threats included in the present study were for gender, race, or age-based ability stereotypes. That is, stereotypes that referenced certain racial groups (e.g., African Americans) are inferior in intellectual ability, women are inferior in mathematical ability, and older adults are inferior in learning and memory ability. Stereotypes that reference athletic ability (e.g., Beilock, Jellison, Rydell, McConnell, & Carr, 2006), social skills (e.g., Bosson, Haymovitz, & Pinel, 2004), and other groups (e.g., class, Croizet & Claire, 1998) were not included in the present study.

Availability of Necessary Statistics. To be retained for the meta-analysis, primary studies must include an effect size quantifying the difference between the stereotype threatened and control groups or the necessary data (e.g., sample size, mean, standard deviations, correlation estimates, independent sample t-test estimates, F-test estimates) that are required to compute a sample-weighted standardized mean difference (i.e., d) between the stereotype threatened and control groups. There were a number of studies that met the preceding requirements, but did not provide the necessary statistics. For studies, that did not report the required data, the authors were contacted in an effort to obtain the requisite data. Eighteen percent (13 of 71) of contacted authors responded after being asked for the additional data.

Language. Lastly, the study must be written in English or have an English translation available. Since the content of stereotypes may vary based on the country in which the stereotype is embedded in, studies that were conducted outside of the United States were noted and coded as such¹.

¹ Seventy-eight percent of data points were from studies conducted in the United States.

Cumulating Results

Many of the studies that were reviewed included measures for a number of the proposed mediators. Since meta-analyses were run at the mediator level for these studies each mediator effect size was treated as an independent effect size.

There were also some studies that included multiple measures of the same mediator variable. For these circumstances, it was important to determine if it is appropriate to cumulate results within a study (Hunter & Schmidt, 2004). The most common reason for multiple data points in the same study occurred when a mediator was measured multiple times using different operationalizations of the variable (e.g., anxiety measured using the State Trait Anxiety Inventory [STAI] and Discrete Emotions measure). In these situations, the data points are considered as dependent and if treated independently could cause sampling error variance to be underestimated in the meta-analysis and in turn an overestimation of the variance of the true effect size. To ensure the assumption of independence was not violated in these situations, the effect sizes were combined across the mediator measures to create a simple estimate. When intercorrelations between the mediator measures were available linear composites were created; however, in many cases the necessary intercorrelations were not available and a simple average was used.

Coding Studies

Three coders coded the studies that were included in the meta-analysis. The dissertation author was the first (lead) coder. The other two coders were undergraduate research assistants who volunteered to work on the project and were compensated for their time. The second and third coders were both upper-level (i.e., senior) psychology undergraduates who expressed interest in the topic and had previous experience as research assistants working in a psychology

lab. These coders were provided resources to familiarize themselves with the purpose of the current study (e.g., summary of the dissertation topic, Steele & Aronson 1995 article, etc.) and were trained by the first author on important concepts related to the study (e.g., stereotype threat, experimental design, etc.). The coders were also trained on the process for coding studies for the current meta-analysis (i.e., using the coding sheet; see Appendix A). The coders also went through practice coding examples, where they independently coded three to six studies and reviewed their results with the first author/lead coder, comparing their coding results with the lead coder's results, and reviewed instances of disagreement, if any. After the training sessions, each coder worked independently to code their assigned list of studies. Coders were encouraged to take notes on anything that seemed unclear and needed further clarification. These notes were used to discuss and resolved any inter-rater disagreement. Periodic review sessions were held to ensure consistency across the coders.

The variables for all studies in the meta-analyses sample were coded by the lead coder. Additionally, 50% of the studies were also coded by one of the two coders to cross-check for accuracy. Consequently, half of the studies were coded by at least two coders. Inter-rater agreement for continuous variables was calculated for those studies that were coded by two coders. For categorical variables, Cohen's Kappa (κ ; Cohen, 1960) was computed to assess the inter-rater agreement. The Kappa average was .66 ($SD = .34$, $Min = .11$, $Max = .94$). Any disagreement was discussed and resolved. The lowest Kappa was for test difficulty, therefore significant time was spent refining the coding for this variable to achieve a clear and consistent understanding. After this process, codes for this variable for revisited to make sure all coders were aligned on the definition.

Coding Sheet and Coding Manual. Following the search and review process, the data in each retained study were coded based on the data coding manual and coding sheet. This section presents a description of the coding process and the variables that were coded.

Coding Process. A data coding manual and coding sheet were developed for the current meta-analysis (see Appendix A). The coding manual includes a list of all variables relevant to the current meta-analysis. The manual includes a description of each variable, a description of the coding categories (e.g., 1 = blatant, 2 = moderate, 3 = subtle), as well as the corresponding Excel column in the coding form. The coding sheet is an Excel spreadsheet where data were entered for each study. The columns in the coding sheet correspond to the variables described in the coding manual. When information was not available for any particular study, the cell was left empty and considered missing data.

Descriptive Study Information. The descriptive information for all studies was recorded, including author name (s), whether the study was published or unpublished, as well as the publication date for those published studies. Additionally, descriptions of the sample and context of the study (e.g., college students, working adults) was recorded. As mentioned in the inclusion criteria section, the country where the study was conducted was also recorded in the coding form.

Study Statistics. The reported study statistics varied across the studies. The key statistic of interest for the current meta-analysis, is a standardized mean difference (i.e., d) between the stereotype threat and non-stereotype threat condition groups for each mediator value (i.e., ST – Non-ST). In some cases, the d was reported in the study and this was used in the analyses. When the d was not reported, the necessary values to compute the standardized mean difference (d) were coded from each study including sample size, variable cell means and standard deviations,

t-test values, and/or *F*-test values. When neither the *d* nor the necessary data to compute it were not reported in the study, the study authors were contacted to obtain the information. Seventy-one authors were contacted to obtain additional data and 13 responded (18%) with additional statistics needed to compute the necessary effect size.

Description of Study Variables. The following sections provide descriptions and the coding guidelines for the mediator variables included in the present meta-analysis, as well as the moderator variables examined in the study.

Mediator Variable. The type of mediator assessed in each study was coded. For affective reactions, the type of reaction was coded (i.e., anxiety, physiological arousal) along with how it was operationalized in the study. The operationalization was coded as indirect (e.g., blood pressure, skin conductance, facial expression) or self-report. For cognitive processes, the specific process were coded (i.e., negative thoughts [to test monitoring and suppression hypotheses] or working memory). Similar to affective reactions, how the cognitive process was operationalized was coded as an implicit (e.g., stimuli responsiveness), task (e.g., working memory tasks) or self-report measure (e.g., self-reported negative thoughts). Lastly for motivational processes, each mediator was coded (e.g., performance orientation, effort). These mediators were also coded for how the variable was operationalized, objective (e.g., number of items completed) or self-report (e.g., self-reported effort, performance-avoidance measures). The mediator definitions, coding categories, and pertinent examples are further described in Table 3.

Test Difficulty. Test difficulty is considered to be a sample-dependent variable, as a test could be considered more difficult if it is administered to a less capable sample (e.g., Spencer et al., 1999). Therefore, test difficulty was coded based on how the study author(s) characterized the test in their study, using the same three categorization levels (very difficult, moderately

difficult, and easy) as Nguyen and Ryan's (2008) meta-analysis. Generally, study authors explicitly described the test difficulty. For example, Boucher, Rydell, and Murphy (2015) administered a timed math test that was coded as very difficult based on their description of the task (i.e., "ten difficult multiple-choice math word problems", p. 58). Jamieson and Harkins (2009) described specifically choosing items that varied in difficulty (i.e., "We selected problems by first randomly picking 12 problems ... that varied in their solution rates from 35% to 65%", p. 1304).

Stereotype Threat Target. The studies in the current meta-analysis were also coded to gather information regarding the stereotype threat target. As mentioned previously, stereotype threat targets for the present meta-analysis are limited to race-, gender-, and age-based stereotypes. Each study was reviewed to identify the specific type of stereotype that was activated (i.e., race-based, gender-based, or age-based stereotypes).

Stereotype Threat Activation. Similar to the Nguyen and Ryan (2008) meta-analysis, studies were coded for the level of the stereotype threat activation cues. The same three coding categories (i.e., blatant, moderate, and subtle) used in Nguyen and Ryan were used in the present study. *Blatant* cues make negative group stereotypes salient to test takers in a deliberate way (e.g., informing minorities groups of their inferiority in the test domain). *Moderate* cues directly communicate information regarding general patterns related to the stereotype, but leave the message open to interpretation (e.g., stating that a test produces group differences but not stating the direction of the differences). *Subtle* cues never directly convey the relevance of a stereotype to the testing situation, rather characteristics of the testing environment may convey the significance of the stereotype (e.g., asking a test-taker to indicate their race/sex before a test).

Mediator Measurement Timing. The timing of the measurement of the mediator was coded as either before or after performance. Studies that did not include measures of performance were not excluded from the meta-analysis; however, they were simply coded as “no performance” and not included in timing moderator analyses.

Additional (Supplementary) Codes. Studies were also coded for a number of other variables that were not related to specific hypotheses, included study design (i.e., longitudinal versus. cross-sectional designs) test purpose (i.e., laboratory vs. operational), and test setting (i.e., individual vs. group settings).

Summary of the Meta-Analytic Data Set

From the 3,372 hits in the initial search, an initial review abstracts of the data resulted in a total of 213 published and unpublished empirical reports on stereotype threat mediator effects that could be potentially included in the meta-analysis. From the 213 reports identified, 133 reports were excluded because they did not meet the inclusion criteria for the current meta-analysis. The most prevalent reason for reports not being included was due to insufficient data reported to compute the standardized mean difference. Eighty of the reports met all the inclusion criteria specified for the current meta-analysis. The screening process is outlined in Figure 4. The 80 reports included in the meta-analysis contained 96 primary studies. Fifty-four reports (67 studies) were published in peer-reviewed journals. Across the 96 primary studies across there were 216 total effect sizes for the seven mediators investigated in the meta-analysis, 161 within-group effect-sizes and 55 between-group effect sizes. The number of effect sizes and the total sample sizes for each mediator are outlined in Table 4.

Table 4
Mediator Frequency Counts

Mediator	Within-Group <i>k</i>	Within-Group <i>n</i>	Between-Group <i>k</i>	Between-Group <i>n</i>
Anxiety	48	2866	12	814
Physiological Arousal	3	228	1	103
Negative Thoughts	22	1284	6	331
Working Memory	19	1171	6	331
Effort	26	1469	16	772
Performance Avoidance	8	536	2	100
Performance Expectations	27	1840	12	850
Total	153	9394	55	3301

Meta-Analytic Procedure

Meta-analysis provides a rigorous method to provide a quantitative summary of literature. Meta-analyses are considered to extend beyond qualitative literature reviews because they allow for a statistical aggregation of results across a large number of studies and samples. These procedures allow for researchers to go beyond the results of single studies and integrate the findings across these studies to examine the pattern of results and the cumulative effect size that is argued to approximate the population effect.

In the present study, random-effects meta-analytic procedures outlined by Hunter and Schmidt (2004) were leveraged. First, overall meta-analyses were conducted to estimate the magnitude of the effect for each of the mediator variables of interest, and then moderator effects for hypothesized relationships were examined.

Correction for Measurement Unreliability

Psychometric meta-analysis provides a superior approach to understanding a psychological phenomenon because it allows for a quantitative estimate of an effect after accounting for sampling error and error from other statistical artifacts (e.g., measure unreliability). The operationalization of the mediators in the current meta-analysis cause them to be susceptible to measurement error. When this is the case, Hunter and Schmidt (2004) recommend correcting for criterion measurement unreliability to estimate the “true” population score for a variable. However, many of the studies in the current sample did not provide reliability estimates for the mediator variables, thus individual corrections for measurement error was not possible. In the absence of reliability estimates for every study, Hunter and Schmidt recommend using an artifact distribution for meta-analysis. This approach was employed by leveraging the available reliability statistics reported by some of the studies. The artifact distribution for each mediator is reported in Table 5. There were no reliability estimates for the physiological arousal or working memory analyses.

Table 5
Artifact Distribution for Mediator Variables

Mediator	<i>k</i>	<i>M</i>	<i>SD</i>
Anxiety	31	.87	.06
Physiological Arousal	–	–	–
Negative Thoughts	8	.83	.04
Working Memory	–	–	–
Effort	5	.83	.09
Performance Avoidance	4	.81	.09
Performance Expectations	6	.80	.08

Computing Effect Sizes

The standardized mean difference (d) was the effect size used in the current meta-analysis. The standardized mean difference represents the mean difference between scores in a standardized format (i.e., ratio of the mean difference in standard deviation units). Effect sizes alone do not provide support or rejection of a hypothesis and are considered for descriptive purposes (Cortina & Landis, 2011). Although some reference commonly used guidelines for interpreting effect sizes are 0.20, 0.50, and 0.80 for small, medium, and large effect sizes for d , it is important to note the context when interpreting effect sizes (Aguinis, Werner, Lanza, Abbot, Angert, Park & Kohlhausen, 2010). To help better understand the context for interpreting stereotype threat, the effect sizes from the meta-analyses by Nguyen and Ryan (2008), Lamont et al. (2015), and Shewach et al. (2019) were also referenced. These effect sizes ranged from 0.14 to 0.80.

The values for each d used in the current meta-analysis were computed from a variety of inputs. The majority (54%) of d s were computed from sample sizes, means, and standard deviation information that was reported in the study. For 20% of the d s, authors provided the necessary information after they were contacted. The remaining d s used were reported in the study or computed from t -test, F -test estimates, or odds ratios that were reported in the study. The effect size for all the studies are listed in Appendix B; additionally, each effect size is plotted by mediator and analysis level in Appendix C.

Meta-Analytic Computations and Moderator Analyses

Hunter and Schmidt (2004) artifact-distribution meta-analytic procedures were used to cumulate the average population effect size δ (corrected for sampling error and criterion [r_{yy}] measurement error), as well as the variance ($SD\delta$). Confidence intervals (CI) and credibility intervals (CV) were estimated for the effect sizes. These intervals help provide additional

interpretation of the effect sizes. First, confidence intervals help researchers determine if a variable (e.g., stereotype threat mediators) is operating in the population and its effects differ from zero. Confidence intervals are computed using the standard error around the average effect size corrected for sampling error (d) and estimate the degree that sampling error remains in the bare-bones (i.e., sample-size only corrected) effect size (Whitener, 1990). Confidence intervals that do not overlap with zero, suggest that the variable is different from zero and has an impact on the population.

Credibility intervals are generally used when testing for the presence of moderators because they help researchers understand whether effect sizes can be generalized to a broader population. The credibility intervals are computed using the standard deviation for the corrected average effect size (i.e., after both sampling and measurement corrections). When credibility intervals are wide this generally suggests that substantial moderators exist and may impact the interpretation of an effect size (Whitener, 1990).

Subgroup Moderator Analysis

For each hypothesized moderator, the data were categorized into the moderating categories and separate meta-analysis were conducted for each category. To establish the evidence for moderators, one can look to see if the population effect size δ is different across the moderator categories. Additionally, the mean corrected standard deviation should be smaller in the moderator categories than in the standard deviation for the combined categories. To judge if there is evidence for moderators one might look to the credibility intervals, as well as checking the percent of variance accounted for by artifacts. The moderator “rule of thumb” test looks to the percent of error variance to the observed variance. If the percent of variance from sampling error is greater than or equal to 75%, then most of the variance is due to error and it is

unlikely that a true moderator effect exists (Hunter & Schmidt, 2004). In addition to the percent of variance accounted for, if the 90% credibility intervals for the population effect size δ are large and overlap zero, this suggest the presence of substantial moderators (Hunter & Schmidt).

Publication Bias

Publication bias is one concern that can arise when conducting a meta-analysis. Particularly there is a risk that meta-analysis authors could only include studies with significant results, this may not be a true representation of the population effect sizes, and could ultimately bias the meta-analytic estimates (Kepes, Banks, McDaniel, & Whetzel, 2012). To combat the potential for publication bias a thorough literature search was conducted and as described in earlier parts of this manuscript, a number of stereotype threat authors were contacted regarding unpublished results to ensure that a comprehensive dataset was used for the current meta-analysis.

Additionally, a number of recommended analyses to test for publication biases were run using the PUB_BIAS macro for SAS (Rendina-Gobioff & Kromre, 2006). These analyses included Beggs Rank Correlation (Begg & Mazumdar, 1994), Egger's Regression (Egger, Smith, Schneider & Minder, 1997), Funnel Plot Regression (Macaskill, Walter, & Irwig, 2001), and the Trim and Fill method (Duval & Tweedie, 2000a, 2000b).

The results of these analyses are presented in Table 6 and suggest that publication bias was not a major concern for the current study. Most results did not show any evidence for any publication bias, with just a few of the mediator results providing evidence for publication bias. Specifically, for the anxiety within-group analyses the Trim and Fill method for estimating the number of studies that might be concealed by publication bias was significant. Additionally, for both the distracting (negative) thoughts and effort within-group analyses both the Egger Regression and Beggs Rank Correlation results were significant suggesting the presence of

publication bias. However, when the results are taken together across all analyses the concern for publication bias is limited. Further implications for these results are discussed in the Discussion section.

Table 6
Publication Bias Results

Mediator	Study Design	Egger Regression	Begg Rank Correlation	Funnel Plot Regression	Trim and Fill
Anxiety	Within-Group	Not Significant	Not Significant	Not Significant	<i>Evidence for Publication Bias</i>
	Between-Group	Not Significant	Not Significant	Not Significant	Not Significant
Physiological Arousal	Within-Group	Not Significant	<i>Evidence for Publication Bias</i>	Not Significant	Not Significant
	Between-Group	—	—	—	—
Working Memory	Within-Group	Not Significant	Not Significant	Not Significant	Not Significant
	Between-Group	Not Significant	Not Significant	Not Significant	Not Significant
Distracting (Negative) Thoughts	Within-Group	<i>Evidence for Publication Bias</i>	<i>Evidence for Publication Bias</i>	Not Significant	Not Significant
	Between-Group	Not Significant	Not Significant	Not Significant	Not Significant
Effort	Within-Group	<i>Evidence for Publication Bias</i>	<i>Evidence for Publication Bias</i>	Not Significant	Not Significant
	Between-Group	Not Significant	Not Significant	Not Significant	Not Significant
Performance Avoidance	Within-Group	Not Significant	Not Significant	Not Significant	Not Significant
	Between-Group	—	—	—	—
Performance Expectations	Within-Group	Not Significant	Not Significant	Not Significant	Not Significant
	Between-Group	Not Significant	Not Significant	Not Significant	Not Significant

RESULTS

The overall within-group and between-group design results are found in Table 7 and Table 8 respectively. Moderator results are presented in Table 9 through Table 13. The results of the meta-analysis are presented as standardized group differences (i.e., d). Each d was computed by subtracting the non-stereotype threatened group's mediator score from the stereotype threatened group's mediator score (i.e., ST – Non-ST). Therefore, a positive d indicates that the stereotype threat group had higher scores on the specified mediator than the comparison group.

The within-group and between-group design results displayed a very similar pattern. As shown in Table 7, the number of data points (k) for the within-group results ranged from 3 to 48 and the total N s ranged from 228 to 2866. The between-group design results were based on a somewhat smaller number of data points and samples sizes (see Table 8), with the number of data points ranging from 1 to 18 and the total N s from 103 to 850. The effect sizes for both the within- and between-group results covered a similar range of magnitudes and included both positive and negative effects (between-group design results: $\delta = -0.69 - 0.52$, within-group design results: $\delta = -0.39 - 0.37$). The range for the between-group design effects was slightly larger than the range for the within-group design effects.

The variance pattern was also comparable for both designs, although these statistics tended to be larger in the between-group design results. For the within-group design results, the observed standard deviations (SD_d) ranged from 0.30 to 0.94, and true standard deviations (SD_δ) ranged from 0.30 to 0.91. For the between-group design results, the observed standard deviations (SD_d) ranged from 0.24 to 1.10, and true standard deviations (SD_δ) ranged from 0.00 to 1.07. After accounting for sampling and measurement error the true effect sizes (δ) increased, and the

true standard deviations were generally smaller than the observed standard deviation. For the within-group design results, sampling and measurement error artifacts explained 6.02 to 66.32 percent of the variance and 6.47 to 100.00 percent of the variance for the between-group design results. Given the smaller number of data points and lower sample sizes for the between-group design results, the larger variances in comparison to the within-group design results are not too surprising.

The preceding section presented a high-level overview of the pattern of results for the within and between-group designs. The following sections discuss the findings in relation to each of the proposed hypotheses. The results are reviewed for both the within- and between- designs and any differences or similarities are discussed in further detail. The results for a number of the moderator analyses (e.g., activation levels by stereotyped group) are based on smaller sizes and/or single studies. They are presented for thoroughness but should be interpreted with caution.

Table 7
Overall Within-Group Meta-Analytic Results

Mediator	<i>k</i>	<i>N</i>	<i>d</i>	<i>SD_d</i>	δ	<i>SD_δ</i>	% Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Anxiety	48	2866	0.17	0.32	0.19	0.20	66.32	-0.08	0.45	0.09	0.29
Physiological Arousal	3	228	0.12	0.94	0.12	0.91	6.02	-1.59	1.84	-12.73	14.04
Working Memory	19	1171	-0.39	0.48	-0.39	0.41	28.58	-0.93	0.15	-0.64	-0.15
Distracting (Negative) Thoughts	22	1284	0.34	0.60	0.37	0.59	19.64	-0.41	1.16	0.08	0.69
Effort	26	1469	0.12	0.43	0.14	0.37	41.98	-0.35	0.62	-0.05	0.33
Performance Avoidance	8	546	0.05	0.30	0.06	0.2	54.08	-0.23	0.34	-0.23	0.34
Performance Expectations	27	1840	-0.17	0.48	-0.19	0.47	25.70	-0.80	0.43	-0.41	0.03

Note: *k* = Number of effect sizes (*d* values); *N* = total sample size; *d* = sample size weighted mean effect size; *SD_d* = sample size weighted standard deviation of *d* values; δ = mean true effect size; *SD_δ* = standard deviation of effect sizes; % Var = percent variance accounted for in observed *d* values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Table 8
Overall Between-Group Meta-Analytic Results

Mediator	k	N	d	SD_d	δ	SD_δ	% Var	CV_{10}	CV_{90}	CI_L	CI_U
Anxiety	12	814	0.47	0.27	0.51	0.10	88.14	0.37	0.65	0.32	0.72
Physiological Arousal	1	103	0.54	—	0.54	—	—	—	—	0.13	1.02
Working Memory	6	328	-0.69	1.10	-0.69	1.07	6.49	-2.26	0.88	-3.43	0.43
Distracting (Negative) Thoughts	6	331	0.45	0.59	0.50	0.58	21.96	-0.36	1.35	-0.18	1.40
Effort	16	772	0.07	0.71	0.07	0.71	0.00	-0.87	1.02	-0.34	0.50
Performance Avoidance	2	100	-0.07	0.24	-0.08	0.00	100.00	-0.08	-0.08	-14.71	14.71
Performance Expectations	12	850	-0.60	0.64	-0.68	0.68	15.37	-1.60	0.25	-1.29	-0.19

Note: k = Number of effect sizes (d values); N = total sample size; d = sample size weighted mean effect size; SD_d = sample size weighted standard deviation of d values; δ = mean true effect size; SD_δ = standard deviation of effect sizes; % Var = percent variance accounted for in observed d values due to all corrected artifacts; CV_{10} and CV_{90} = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Affective Mediator Results

Hypotheses 1 and 2 predicted that anxiety and physiological arousal would be higher for stereotype threatened individuals and that effect sizes would be bigger when affective reactions were measured by indirect or physiological measures. Support was found for Hypothesis 1; anxiety was higher for stereotype threatened individuals. This result was consistent for both within-group and between-group designs, although the magnitude of the effect was stronger for the between-group ($\delta = 0.51$) design results, than the within-group ($\delta = 0.19$) design results.

Results for Hypothesis 2 should be interpreted with caution because the data available were limited. There were only within-group design three data points (k) and one between-group design data point. For both designs, physiological arousal was higher for stereotype threatened individuals ($\delta = 0.12$ and $\delta = 0.54$); however, the confidence intervals were wide and included zero, thus Hypothesis 2 was not supported.

Hypothesis 3 was not supported; effect sizes were not larger for physiological arousal when compared to measures of anxiety. For the within-group design results effect sizes for anxiety and psychological arousal were both small ($\delta = 0.12$ vs. $\delta = 0.19$), and for physiological arousal the confidence interval included zero. The effect sizes were larger for the between-group design results for both anxiety and physiological arousal ($\delta = 0.46$ vs. $\delta = 0.54$). However, the results should be interpreted with caution for physiological arousal since this only reflects the result of one data point.

Overall, only the results for anxiety were consistent across both within- and between- designs providing some level of support for affective reactions as mediators in the stereotype threat-performance relationship, with the results suggesting that exposure to stereotype threat is related to increased levels of anxiety. The moderator results were mixed across the designs. The

within-group results suggest that there may be additional moderators, while the between-group results suggest that much of the variance between the effect sizes is explained by sampling and measurement error (88.14%).

Cognitive Processes Mediator Results

Hypotheses 4 and 5 focused on the cognitive mediators, working memory and distracting (negative) thoughts. Hypothesis 4 which predicted that working memory capacity would be lower for stereotype threatened individuals than for individuals not exposed to stereotype threat was supported; working memory capacity was lower for the stereotype threatened individuals. These effects ranged from medium to large across the within-group ($\delta = -0.39$) and between-group ($\delta = -0.69$) results. However, the between-group design results should not be interpreted because the confidence interval included zero (95% CI = -3.43 – 0.43). In addition, the large credibility intervals and small percentage of variance accounted for both within- and between-group design results suggest the presence of moderators.

Partial support was also found for Hypothesis 5. As predicted, individuals who experienced stereotype threat had more distracting (negative) thoughts than those who had not been exposed to stereotype threat. This effect was consistent for both the within-group ($\delta = 0.37$) and between-group ($\delta = 0.50$) design results. However, similar to the working memory findings, the confidence intervals for the between-group effect included zero. Additionally, for both the within- and between-group effect sizes their credibility intervals overlapped with zero and variance accounted for was less than 75% suggesting the presence of additional moderations.

Hypothesis 6 and Hypothesis 7 pertained to the presence of moderators for the distracting (negative) thoughts effects and the results are presented in Table 9 and Table 10. There were a small number of data points ($k = 1 - 3$) for the between-group design results, therefore these

results should be interpreted with caution. Hypothesis 6 was focused on the presence of monitoring processes operating and predicted that distracting (negative) thoughts would be higher for stereotype threatened individuals when stereotype threat was activated with subtle cues. Hypothesis 6 was partially supported; distracting (negative) thoughts were higher for stereotype threatened individuals when stereotype threat was activated through subtle cues (within-group $\delta = 0.77$, between-group $\delta = 0.70$). However, the confidence intervals for the between-group effect size included zero. Additionally, there was evidence for moderators for both effects.

Hypothesis 7 was focused on better understanding the suppression process. The hypothesis predicted that there would be no differences in the level of distracting (negative) thoughts between stereotype threatened and non-stereotyped threatened individuals when stereotype was activated with blatant cues. This hypothesis was supported, the effect sizes approached zero for both within-group ($\delta = -0.03$) and between-group ($\delta = 0.11$) analyses and the confidence intervals included zero, suggesting there was no differences in the number of distracting (negative) thoughts between stereotype and non-stereotyped individuals.

Overall the cognitive process results present a mixed picture for stereotype threat's impact on cognitive mediators. Although the within-group effects suggest that mediation likely occurs through a reduction of working memory function and an increase in distracting (negative) thoughts, that same result cannot be interpreted as meaningful for the between-group design findings. The within-group results could help to provide some clarification of the operation of monitoring and suppression processes. Specifically, when threat was activated with subtle cues, there were increased levels of distracting (negative) thoughts for stereotype threatened individuals suggesting that monitoring processes were operating in these scenarios. When

stereotype was activated with blatant measures, similar levels of distracting (negative) thoughts were observed between stereotype threatened and non-threatened individuals; these results suggest the presence of suppression processes in these types of scenarios. Although, the pattern for these effects were consistent across both the within- and between-group design results, the between-group design results should be interpreted with caution because these are based on a small number of data points (i.e., only one study) and the confidence intervals for the effect included zero. Additional data points to test these effects in between-group samples are needed to better understand how these effects may manifest.

Table 9
Overall Main Hypotheses Within-Group Moderator Results

Mediator	Moderator	<i>k</i>	<i>N</i>	<i>d</i>	<i>SDd</i>	δ	<i>SD</i> δ	% Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Distracting (Negative) Thoughts	Subtle	10	474	0.69	0.72	0.77	0.74	17.61	-0.26	1.79	0.17	1.64
	Moderate	4	259	0.47	0.30	0.51	0.17	72.29	0.23	0.8	-0.02	1.17
	Blatant	8	551	-0.03	0.36	-0.03	0.29	44.87	-0.45	0.38	-0.37	0.30
Effort	Difficult Test	5	297	0.20	0.38	0.23	0.32	46.86	-0.26	0.71	-0.31	0.82
	Moderately Difficult Test	15	774	0.18	0.40	0.20	0.33	48.45	-0.24	0.64	-0.05	0.46
	Easy Test	1	29	0.93	—	1.08	—	—	—	—	0.12	3.86
	Self-Report	16	870	0.00	0.41	0.00	0.34	45.71	-0.46	0.45	-0.25	0.24
	Objective	14	742	0.33	0.39	0.38	0.31	52.07	-0.04	0.79	0.12	0.66
Performance Expectation	Before Task	15	1061	-0.23	0.48	-0.26	0.47	24.68	-0.89	0.37	-0.58	0.04
	After Task	5	423	-0.03	0.53	-0.04	0.54	17.47	-0.87	0.8	-0.82	0.73

Note: *k* = Number of effect sizes (*d* values); *N* = total sample size; *d* = sample size weighted mean effect size; *SDd* = sample size weighted standard deviation of *d* values; δ = mean true effect size; *SD* δ = standard deviation of effect sizes; % Var = percent variance accounted for in observed *d* values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Table 10
Overall Main Hypotheses Between-Group Moderator Results

Mediator	Moderator	<i>k</i>	<i>N</i>	<i>d</i>	<i>SDr</i>	δ	<i>SD</i> δ	% Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Distracting (Negative) Thoughts	Subtle	3	145	0.63	0.94	0.70	1.00	10.18	-1.19	2.58	-2.82	14.09
	Moderate	2	137	0.39	0.08	0.42	0.00	100.00	0.42	0.42	-0.42	1.58
	Blatant	1	49	0.10	—	0.11	—	—	—	—	-0.52	0.78
Effort	Difficult Test	6	261	0.13	0.96	0.15	1.03	10.25	-1.37	1.67	-1.04	1.55
	Moderately Difficult Test	4	170	0.39	0.69	0.44	0.71	20.70	-0.71	1.6	-0.82	2.78
	Easy Test	1	29	0.26	—	0.29	—	—	—	—	-0.52	1.28
	Self-Report	11	614	-0.05	0.72	-0.05	0.75	14.47	-1.08	0.98	-0.62	0.50
	Objective	7	226	0.61	0.63	0.70	0.60	34.93	-0.16	1.56	0.02	1.71
Performance Expectation	Before Task	4	235	-1.29	0.78	-1.51	0.94	14.93	-3.06	0.03	-14.07	0.10
	After Task	6	491	-0.38	0.36	-0.42	0.31	42.51	-0.87	0.03	-0.91	0.58

Note: *k* = Number of effect sizes (*d* values); *N* = total sample size; *d* = sample size weighted mean effect size; *SDd* = sample size weighted standard deviation of *d* values; δ = mean true effect size; *SD* δ = standard deviation of effect sizes; % Var = percent variance accounted for in observed *d* values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Motivational Processes Mediator Results

Hypotheses 8, 9, 10, 11, 12, and 13 all focused on motivational process mediators. Specifically, Hypotheses 8, 9, and 10 focused on the effects of effort, Hypothesis 11 focused on performance orientation, and Hypotheses 12 and 13 focused on performance expectations. The overall results show that stereotype threatened individuals exerted more effort. The effects sizes were similar in magnitude for both within- ($\delta = 0.17$) and between-group ($\delta = 0.15$) design results. However, across the results for both designs, the confidence intervals included zero and there is evidence of moderators for these relationships (i.e., low percentage of variance accounted for and credibility intervals overlapping with zero).

Hypotheses 8 and 9 predicted that test difficulty would moderate the stereotype threat-effort relationships. Specifically, Hypothesis 8 predicted that stereotype-threatened and non-stereotype-threatened groups would differ in their effort levels on difficult tests. As the results presented in Table 9 and Table 10 show, although the δ s for both the within- ($\delta = 0.32$) and between-group ($\delta = 0.15$) design results suggest that stereotype threatened individuals exhibit increased effort on difficult test compared to non-threatened individuals, the confidence intervals for both effects overlapped with zero.

Hypothesis 9 predicted that stereotype threatened individuals would exert similar levels of effort as non-threatened individuals on less difficult (easy) tests. However, because of limited data this hypothesis was not tested. Only one data point was available for each design type, therefore the results are presented for thoroughness but should be interpreted with caution. For both the within- and between-group design results, the effect sizes suggest that effort increases for stereotype threatened individuals in comparison to non-threatened individuals on less

difficult tests (within-group $\delta = 1.08$, between-group $\delta = 0.29$). However, each effect size was based on one data point and the confidence intervals were quite large and included zero.

Hypothesis 10 predicted that studies that use self-report measures of effort would produce smaller mean effect sizes for effort than studies that use more objective measures. As the results presented in Table 9 and Table 10 show, this hypothesis was supported. The results for self-report measures of effort were close to zero for both within- ($\delta = -0.02$) and between-group ($\delta = -0.05$) designs, and the confidence intervals included zero. In contrast, when effort was measured via objective measures, the results indicate small to moderate effects between stereotype threat and effort. These results were consistent for both the within ($\delta = 0.38$) and between-group ($\delta = .70$) designs. Taken together the results present a fairly mixed picture when considering stereotype threat's relationship to effort. Particularly, stereotype threat seems to have the largest effect on effort when effort is measured via objective measures. Specifically, there is evidence that stereotype threatened individuals exhibit increased levels of effort, but they may not be actively aware of this increased exertion.

Hypothesis 11 predicted that a prevention or avoidance orientation would be most pronounced for individuals who experience stereotype threat. There were only two data points available to test this hypothesis in the between-group design so these results should be interpreted with caution. Support was not found for Hypothesis 11, individuals who experienced stereotype threat did not have a higher performance avoidance approach than non-stereotype threatened individuals (within-group $\delta = 0.06$, between-group $\delta = -0.08$).

The last set of motivational processes mediators focused on performance expectations. Hypothesis 12 predicted that performance expectations would be lower for individuals who had experienced stereotype threat in comparison to individuals who did not experience stereotype

threat. As the results presented in Table 7 and Table 8 show, performance expectations were lower for stereotype threatened individuals for both the within ($\delta = -0.19$) and between-group ($\delta = -0.68$) design results. Although the pattern of results seem to support the hypothesis the confidence intervals for the within-group effect included zero. Additionally, for both effect sizes credibility intervals were wide and included zero suggesting that there are likely moderators for these relationships.

Hypothesis 13 proposed that the timing of the performance expectation measure would moderate the stereotype threat-performance expectations relationship. Specifically, studies that measured performance expectations prior to the completion of the task would produce larger negative effects than studies that measured performance expectations after the task. However, both within- and between-group design results are mixed (see Table 8 and Table 9). The within-group results do not support Hypothesis 13; the effect size for performance expectations measured before performance was somewhat bigger ($\delta = -0.26$) than the effect size for performance expectations measured after performance ($\delta = -0.04$). However, both effect sizes had confidence intervals that included zero. The between-groups findings present a similar pattern of findings, although the magnitude of effects was larger. When performance expectations were measured before the task those who experienced stereotype threat had lower expectations and it was a large effect ($\delta = -1.51$); when expectations were measured after the task, those who experienced stereotype threat also had lower expectations ($\delta = -0.42$), but the effect size was moderate in size. It is important to note that although the magnitudes seem large, the confidence intervals for both effect sizes included zero.

Overall the results for the motivational mediators were quite mixed. There is evidence that effort might mediate the stereotype threat-performance relationship, but this effect is only

evident under particular circumstances. Specifically, when effort is measured via objective measures the pattern emerges that stereotype threatened individuals exhibit increased levels of effort. This pattern was evident in both the within- and between-group design results. There was no evidence that performance avoidance orientation varied across stereotyped versus non-stereotype individuals. Last, the results were again mixed for Hypothesis 12. Stereotype threatened individuals had lower performance expectations than non-threatened individuals, but this pattern was only observed in the between-group design results. The only results that were consistent across both the within- and between-group designs were the results for increased levels of effort observed for stereotype threatened individuals when effort was measured via objective measures.

Moderator Hypotheses Analyses

Additionally, moderator analyses were conducted to test the final three hypotheses. Hypothesis 14 predicted that mediator effect sizes would be larger for stereotype threatened racial minorities than stereotype threatened women and older adults. Many of the analyses are based on a relatively small number of data points (most analyses are based on *ks* less than 10), and most of the data points had small *Ns*. Notably the between-group design results had the smallest sample sizes and although these results are presented, they should be interpreted with caution. These results are presented in Table 11 and Table 12.

Table 11
Within-Group Meta-Analytic Results by Stereotype Threatened Group

Stereotype Threatened Group	Mediator	<i>k</i>	<i>N</i>	<i>d</i>	<i>SD_d</i>	δ	<i>SD_δ</i>	%Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Women	Anxiety	31	1674	0.11	0.31	0.11	0.15	79.34	-0.08	0.31	-0.01	0.24
	Physiological Arousal	3	228	0.12	0.94	0.12	0.91	6.02	-1.59	1.84	-12.73	14.04
	Working Memory	14	788	-0.39	0.59	-0.39	0.52	21.89	-1.09	0.32	-0.77	-0.04
	Distracting (Negative) Thoughts	15	912	0.31	0.54	0.34	0.53	22.85	-0.36	1.05	0.01	0.71
	Effort	24	1323	0.16	0.39	0.18	0.32	48.60	-0.24	0.6	-0.01	0.37
	Performance Avoidance	7	451	0.06	0.34	0.06	0.25	55.21	-0.30	0.43	-0.28	0.42
	Performance Expectations	17	1006	-0.10	0.46	-0.11	0.42	32.42	-0.68	0.45	-0.38	0.15
Racial Minorities	Anxiety	11	762	0.19	0.30	0.21	0.19	67.83	-0.05	0.47	-0.01	0.43
	Working Memory	2	105	-0.32	0.22	-0.32	0.00	100.00	-0.32	-0.32	-14.04	2.95
	Distracting (Negative) Thoughts	4	226	0.63	0.94	0.70	1.02	8.48	0.97	2.37	-0.95	14.04
	Effort	4	127	0.00	0.68	0.00	0.65	28.82	-1.05	1.06	-1.43	1.44
	Performance Avoidance	1	95	0.02	—	0.03	—	—	—	—	-0.42	0.48
	Performance Expectations	5	450	-0.12	0.58	-0.13	0.6	13.66	-1.06	0.79	-1.03	0.69
Older Adults	Anxiety	6	430	0.40	0.35	0.43	0.27	47.76	0.09	0.83	0.04	0.89
	Working Memory	3	278	-0.42	0.13	-0.42	0.00	100.00	-0.42	-0.42	-0.79	-0.08
	Distracting (Negative) Thoughts	3	146	0.04	0.23	0.04	0.00	100.00	0.04	0.04	-0.6	0.70
	Effort	2	162	0.21	0.73	0.24	0.78	9.58	-2.18	2.65	-14.04	14.04
	Performance	5	384	-0.40	0.46	-0.44	0.45	25.12	-1.14	0.25	-1.26	0.20

Note: *k* = Number of effect sizes (*d* values); *N* = total sample size; *d* = sample size weighted mean effect size; *SD_d* = sample size weighted standard deviation of *d* values; δ = mean true effect size; *SD_δ* = standard deviation of effect sizes; %Var = percent variance accounted for in observed *d* values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Table 12
Between-Group Meta-Analytic Results by Stereotype Threatened Group

Stereotype Threatened Group	Mediator	k	N	d	SD_d	δ	SD_δ	% Var	CV_{10}	CV_{90}	CI_L	CI_U	
Women	Anxiety	8	608	0.49	0.31	0.53	0.22	58.59	0.22	0.83	0.23	0.85	
	Working Memory	3	99	0.30	1.26	0.30	1.21	7.91	-1.97	2.58	-14.04	14.04	
	Distracting (Negative) Thoughts	3	162	0.40	0.28	0.44	0.03	98.78	0.38	0.51	-0.33	1.48	
	Effort	12	612	0.11	0.64	0.13	0.65	19.77	-0.79	1.01	-0.33	0.61	
	Performance	2	100	-0.07	0.24	-0.08	0.00	100.00	-0.08	-0.08	-14.71	14.71	
	Avoidance												
	Performance Expectations	6	502	-0.50	0.4	-0.56	0.37	33.31	-1.11	0.01	-1.16	-0.08	
Racial Minorities	Anxiety	3	140	0.40	0.11	0.43	0.00	100.00	0.43	0.43	0.11	0.77	
	Working Memory	1	36	-0.65	—	-0.65	—	—	—	—	-1.62	0.03	
	Distracting (Negative) Thoughts	2	103	0.90	0.93	1.00	1.03	10.11	-2.16	4.17	-14.07	14.07	
	Effort	5	162	0.24	1.25	0.27	1.36	8.41	-0.182	2.35	-1.71	4.59	
	Performance Expectations	3	138	-0.43	0.66	-0.49	0.67	21.53	-1.75	0.77	-14.13	1.64	
Older Adults	Anxiety	1	66	0.50	—	0.54	—	—	—	—	0.00	1.21	
	Physiological Arousal	1	103	0.54	—	0.54	—	—	—	—	0.13	1.02	
	Working Memory	2	193	-1.22	0.88	-1.22	0.85	3.95	-3.84	1.40	-14.04	14.04	
	Distracting (Negative) Thoughts	1	66	-0.13	—	-0.14	—	—	—	—	-0.70	0.39	
	Effort	1	66	0.03	—	0.03	—	—	—	—	-0.52	0.60	
	Performance Expectations	3	210	-0.95	1.12	-1.09	1.33	5.29	-3.59	1.42	-14.04	3.42	

Note: k = Number of effect sizes (d values); N = total sample size; d = sample size weighted mean effect size; SD_d = sample size weighted standard deviation of d values; δ = mean true effect size; SD_δ = standard deviation of effect sizes; % Var = percent variance accounted for in observed d values due to all corrected artifacts; CV_{10} and CV_{90} = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

The results for anxiety did not support Hypothesis 14. For the within-group findings, the effect size for age-based stereotype threat was largest ($\delta = 0.43$), followed by the effect sizes for race-based stereotype threat ($\delta = 0.21$). The gender-based stereotype threat effect size was smallest ($\delta = 0.11$). It is also important to note that the confidence intervals included zero for the gender-based and race-based effect sizes. The results for the between-group findings were mostly consistent with the within-group findings, with the age-based effect size again being the largest ($\delta = 0.54$). However, for these findings the gender-based effect size was next ($\delta = 0.53$), followed by the race-based effect size ($\delta = 0.43$). Although the between-group effect sizes were larger, they were based on a smaller number of data points. Additionally, the age-based effect size only consisted of one data point and the confidence interval included zero. Overall, these results do not support Hypothesis 14; the race-based effect size was not larger than the age and gender-based effect sizes for the stereotype threat-anxiety relationship.

Only four data points were available to test stereotype threat's effects on physiological arousal, three within-group data points for gender-based stereotype threat and one between-group sample for age-based stereotype threat. Since no race-based effect sizes were available, Hypothesis 14 could not be tested. However, it is interesting to note that the age-based stereotype threat effect was again the largest effect ($\delta = 0.54$), while the gender-based effect was small ($\delta = 0.12$), and the confidence interval included zero. Although not specifically predicted, the results tend to suggest that stereotype threat has the most detrimental effects on older adults' affective reactions. Older adults had the highest levels of anxiety and physiological arousal when exposed to stereotype threat.

The results were mixed across the within- and between-group results for working memory. For the within-group results, all effect sizes were negative and moderate in size with

the age-based effect being the largest ($\delta = -0.42$), followed by the gender-based ($\delta = -0.39$), and race-based ($\delta = -0.32$) effect size. However, the confidence intervals for the race-based effect size included zero. The pattern of the effect sizes for the between-group results were not consistent with the within-group results. Although stereotype threat was associated with decreased working memory capacity for racial minorities ($\delta = -0.65$) and older adults ($\delta = -1.65$), working memory capacity increased for women ($\delta = 0.30$). However, all group-based between-group confidence intervals included zero.

The results were again mixed for stereotype threat's effects on distracting (negative) thoughts. Typically, distracting (negative) thoughts tended to increase in the stereotype threat conditions across each stereotype threatened group. This pattern was true for both gender-based and race-based effect sizes for within-group ($\delta = 0.34, 0.70$) and between-group design results ($\delta = 0.44, 1.00$); however, this was not consistent for age-based findings ($\delta = 0.04, -0.14$). Additionally, only the gender-based within-group stereotype threat effect had confidence intervals that did not include zero. Overall Hypothesis 14 was not supported for the cognitive process mediators. The hypothesized pattern of larger effect sizes for race-based results was not supported and most of the effect sizes were not different from zero.

Hypothesis 14 was not supported for the motivational process mediators. The effect sizes for all groups and analyses had confidence intervals that included zero and again did not follow the hypothesized pattern.

Overall, the results did not support Hypothesis 14. The only mediator where the race-based effect size was larger than the gender and race-based effects was for distracting (negative) thoughts; however, the confidence interval for the race-based effect size included zero. Additionally, there was evidence of additional moderators for many of the mediator effect sizes.

The last set of hypotheses examined moderators for the group-based stereotype findings. Specifically, Hypothesis 15 predicted that the mediator effect sizes would be largest for women when stereotype threat was activated using subtle cues. The results for Hypothesis 15 are presented in Table 13 and Table 14.

The within-group results for gender-based stereotype threat are presented first. The effect sizes were in the hypothesized direction for anxiety, distracting (negative) thoughts, effort, and performance expectations. That is, the effect sizes for subtle activation were larger than the effect sizes for moderate or blatant activation. However, the confidence intervals for the effect sizes for all activation levels for anxiety, effort, and performance expectations included zero, and the hypothesis was only supported for distracting (negative) thoughts. This is also consistent with the overall results for distracting (negative) thoughts, with the most detrimental effects occurring when stereotype threat was activated in a subtle manner. The trend was slightly different for the working memory findings, in that the effect size for moderate activation levels was negative and was the only effect that differed from zero. Additionally, the subtle activation effect size was in the positive rather than negative direction. However, this effect size is from one small sample ($N = 38$). There were no data available to perform the analyses for physiological arousal or performance avoidance. Overall, Hypothesis 15 was not supported; gender-based effect sizes were not larger when stereotype threat was activated via subtle cues.

The gender-based stereotype threat between-group results present a less clear picture, and only the pattern for the distracting (negative) thoughts effect sizes were in the anticipated direction. Additionally, all of the between-group effect sizes had confidence intervals that included zero. Consistent with the within-group results the between-group findings did not

support Hypothesis 15, gender-based effect sizes were not larger when stereotype threat was activated via subtle cues.

Table 13

Gender-Based Stereotype Threat Within-Group Meta-Analytic Results by Stereotype Activation Level

Mediator	Moderator Level	<i>k</i>	<i>N</i>	<i>d</i>	<i>SD_d</i>	δ	<i>SD_δ</i>	%Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Anxiety	Blatant	7	469	0.05	0.18	0.05	0.00	100.00	0.05	0.05	-0.12	0.23
	Moderate	14	822	0.08	0.28	0.09	0.10	88.23	-0.05	0.23	-0.09	0.26
	Subtle	10	383	0.23	0.47	0.24	0.36	49.59	-0.25	0.74	-0.11	0.62
Physiological Arousal	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	2	143	0.27	1.32	0.27	1.32	3.24	-3.74	4.28	-14.04	14.04
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Working Memory	Blatant	5	343	-0.40	0.28	-0.40	0.14	75.50	-0.61	-0.18	-0.79	-0.04
	Moderate	8	407	-0.42	0.79	-0.42	0.74	12.95	-1.47	0.63	-1.24	0.23
	Subtle	1	38	0.04	—	0.04	—	—	—	—	-0.60	0.69
Distracting (Negative) Thoughts	Blatant	7	471	-0.04	0.39	-0.04	0.34	38.63	-0.53	0.45	-0.45	0.36
	Moderate	4	259	0.47	0.30	0.51	0.17	72.29	0.23	0.80	-0.08	1.17
	Subtle	4	182	1.00	0.38	1.12	0.25	69.34	0.72	1.43	0.33	2.84
Effort	Blatant	7	556	0.01	0.27	0.01	0.16	72.40	-0.22	0.24	-0.27	0.29
	Moderate	12	529	0.16	0.37	0.18	0.24	66.89	-0.15	0.52	-0.10	0.82
	Subtle	5	238	0.51	0.52	0.58	0.49	33.96	-0.15	1.65	-0.17	1.33
Performance Avoidance	Blatant	3	168	0.04	0.45	0.05	0.41	35.51	-0.73	0.82	-1.47	1.66
	Moderate	4	283	0.07	0.32	0.08	0.23	57.58	-0.30	0.46	-0.50	0.67
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Performance Expectations	Blatant	9	457	0.04	0.56	0.05	0.54	25.94	-0.70	0.80	-43.00	0.54
	Moderate	5	205	-0.14	0.39	-0.16	0.26	65.54	-0.55	0.23	-0.74	0.38
	Subtle	3	344	-0.27	0.37	-0.30	0.35	26.78	-1.70	0.74	-0.96	0.36

Note: *k* = Number of effect sizes (*d* values); *N* = total sample size; *d* = sample size weighted mean effect size; *SD_d* = sample size weighted standard deviation of *d* values; δ = mean true effect size; *SD_δ* = standard deviation of effect sizes; %Var = percent variance accounted for in observed *d* values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Table 14

Gender-Based Stereotype Threat Between-Group Meta-Analytic Results by Stereotype Activation Level

Mediator	Moderator Level	<i>k</i>	<i>N</i>	<i>d</i>	<i>SD_d</i>	δ	<i>SD_δ</i>	%Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Anxiety	Blatant	1	170	0.66	—	0.72	—	—	—	—	0.35	1.17
	Moderate	5	293	0.23	0.19	0.25	0.00	100.00	0.25	0.25	-0.01	0.52
	Subtle	2	145	0.82	0.18	0.89	0.00	100.00	0.89	0.89	-1.05	14.08
Physiological Arousal	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Working Memory	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	3	99	0.30	1.26	0.30	1.21	7.90	-1.97	2.58	-14.04	14.04
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Distracting (Negative) Thoughts	Blatant	1	49	0.10	—	0.11	—	—	—	—	-0.52	0.78
	Moderate	1	79	0.44	—	0.48	—	—	—	—	-0.01	1.08
	Subtle	1	34	0.75	—	0.83	—	—	—	—	0.03	2.27
Effort	Blatant	3	236	0.10	0.72	0.12	0.78	10.43	-1.35	1.58	-3.56	8.84
	Moderate	4	187	0.27	0.39	0.30	0.29	58.41	-0.16	0.77	-0.40	1.13
	Subtle	5	189	-0.03	0.89	-0.03	0.93	13.84	-1.45	1.39	-1.50	1.39
Performance Avoidance	Blatant	2	100	-0.07	0.24	-0.08	0.00	100.00	-0.08	-0.08	-14.71	14.71
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Performance Expectations	Blatant	2	93	-0.46	0.44	-0.52	0.35	50.95	-1.61	0.57	-14.67	14.67
	Moderate	2	89	-0.58	0.61	-0.66	0.61	26.03	-2.53	1.21	-14.14	14.14
	Subtle	2	320	-0.49	0.50	-0.55	0.53	10.71	-2.19	1.10	-14.04	14.04

Note: *k* = Number of effect sizes (*d* values); *N* = total sample size; *d* = sample size weighted mean effect size; *SD_d* = sample size weighted standard deviation of *d* values; δ = mean true effect size; *SD_δ* = standard deviation of effect sizes; %Var = percent variance accounted for in observed *d* values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Hypothesis 16 predicted that the mediator effect sizes would be largest for racial minorities when stereotype threat was activated using moderate cues. As can be seen in Table 15 and Table 16, there were not sufficient data available across the moderator categories to test Hypothesis 16. The most data were available for subtle stereotype threat activation, but data for blatant and moderate activation levels were not consistently available. Although there were no specific hypotheses for age-based stereotype threat and activation levels, these results are presented in Table 17 and Table 18 for the sake of completeness. Interestingly, similar to the race-based stereotype threat results, data were available for most mediators for subtle activation, but not for moderate and blatant activation. The availability of more data points for the mediators when stereotype threat is activated via subtle cues is noteworthy. This findings and potential implications are discussed in the discussion.

Table 15

Race-Based Stereotype Threat Within-Group Meta-Analytic Results by Stereotype Activation Level

Mediator	Moderator Level	k	N	d	SD_d	δ	SD_δ	% Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Anxiety	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	1	86	0.50	—	0.54	—	—	—	—	0.07	1.12
	Subtle	10	779	0.18	0.27	0.19	0.16	70.56	0.03	0.41	-0.02	0.41
Physiological Arousal	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Working Memory	Blatant	1	69	-0.21	—	-0.21	—	—	—	—	-0.71	0.26
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	1	36	-0.54	—	-0.54	—	—	—	—	-1.43	0.12
Distracting (Negative) Thoughts	Blatant	1	80	0.02	—	0.02	—	—	—	—	-0.46	0.51
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	3	146	0.96	1.04	1.08	1.17	8.66	-1.12	3.29	-2.39	14.04
Effort	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	1	36	-0.05	—	-0.06	—	—	—	—	-0.83	0.70
	Subtle	3	91	0.02	0.85	0.03	0.86	19.46	-1.45	1.39	-14.11	14.11
Performance Avoidance	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	1	198	-0.19	—	-0.21	—	—	—	—	-0.54	0.10
Performance Expectations	Blatant	2	122	-0.75	0.82	-0.85	0.91	10.68	-3.66	1.97	-14.04	14.04
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	3	328	0.12	0.25	0.13	0.18	59.70	-0.21	0.47	-0.59	0.91

Note: k = Number of effect sizes (d values); N = total sample size; d = sample size weighted mean effect size; SD_d = sample size weighted standard deviation of d values; δ = mean true effect size; SD_δ = standard deviation of effect sizes; % Var = percent variance accounted for in observed d values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Table 16

Race-Based Stereotype Threat Between-Group Meta-Analytic Results by Stereotype Activation Level

Mediator	Moderator Level	<i>k</i>	<i>N</i>	<i>d</i>	<i>SD_d</i>	δ	<i>SD_δ</i>	% Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Anxiety	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	1	36	0.53	—	0.57	—	—	—	—	-0.14	1.59
	Subtle	2	104	0.35	0.08	0.38	0.00	100.00	0.38	0.38	-0.42	1.41
Physiological Arousal	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Working Memory	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	1	36	-0.65	—	-0.65	—	—	—	—	-1.62	0.03
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Distracting (Negative) Thoughts	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	1	58	0.32	—	0.35	—	—	—	—	-0.22	1.01
	Subtle	1	45	1.64	—	1.94	—	—	—	—	0.72	14.04
Effort	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	1	48	0.05	—	0.06	—	—	—	—	-0.67	0.74
	Subtle	4	114	0.32	1.53	0.36	1.69	6.42	-2.41	3.13	-5.71	14.07
Performance Avoidance	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Performance Expectations	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	1	58	0.06	—	0.06	—	—	—	—	-0.52	0.67
	Subtle	2	80	-0.79	0.64	-0.90	0.64	28.95	-2.85	1.06	-14.11	14.11

Note: *k* = Number of effect sizes (*d* values); *N* = total sample size; *d* = sample size weighted mean effect size; *SD_d* = sample size weighted standard deviation of *d* values; δ = mean true effect size; *SD_δ* = standard deviation of effect sizes; %Var = percent variance accounted for in observed *d* values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Table 17

Age-Based Stereotype Threat Within-Group Meta-Analytic Results by Stereotype Activation Level

Mediator	Moderator Level	<i>k</i>	<i>N</i>	<i>d</i>	<i>SD_d</i>	δ	<i>SD_δ</i>	% Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Anxiety	Blatant	4	323	0.49	0.38	0.53	0.33	36.61	-0.01	1.06	-0.13	1.39
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	2	107	0.15	0.04	0.16	0.00	100.00	0.16	0.16	-0.21	0.55
Physiological Arousal	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Working Memory	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	2	168	-0.34	0.09	-0.34	0.00	100.00	-0.34	-0.34	-1.42	0.50
	Subtle	1	110	-0.54	—	-0.54	—	—	—	—	-1.00	-0.14
Distracting (Negative) Thoughts	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	3	146	0.04	0.23	0.04	0.00	100.00	0.04	0.04	-0.60	0.70
Effort	Blatant	1	97	0.63	—	0.72	—	—	—	—	0.22	1.37
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	1	65	-0.42	—	-0.48	—	—	—	—	-1.17	0.08
Performance Avoidance	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Performance Expectations	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	2	168	-0.24	0.24	-0.27	0.1	85.21	-0.58	0.05	-14.04	14.04
	Subtle	3	216	-0.52	0.61	-0.59	0.64	15.73	-1.79	0.62	-14.04	1.24

Note: *k* = Number of effect sizes (*d* values); *N* = total sample size; *d* = sample size weighted mean effect size; *SD_d* = sample size weighted standard deviation of *d* values; δ = mean true effect size; *SD_δ* = standard deviation of effect sizes; % Var = percent variance accounted for in observed *d* values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Table 18

Age-Based Stereotype Threat Between-Group Meta-Analytic Results by Stereotype Activation Level

Mediator	Moderator Level	<i>k</i>	<i>N</i>	<i>d</i>	<i>SD_d</i>	δ	<i>SD_δ</i>	%Var	CV ₁₀	CV ₉₀	CI _L	CI _U
Anxiety	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	1	66	0.50	—	0.54	—	—	—	—	0.00	1.22
Physiological Arousal	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	1	103	0.54	—	0.54	—	—	—	—	0.13	1.02
Working Memory	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	1	84	-1.93	—	-1.93	—	—	—	—	-5.93	-0.98
	Subtle	1	109	-0.67	—	-0.67	—	—	—	—	-1.18	-0.26
Distracting (Negative) Thoughts	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	1	66	-0.13	—	-0.14	—	—	—	—	-0.70	0.39
Effort	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	1	66	0.03	—	0.03	—	—	—	—	-0.52	0.60
Performance Avoidance	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	—	—	—	—	—	—	—	—	—	—	—
	Subtle	—	—	—	—	—	—	—	—	—	—	—
Performance Expectations	Blatant	—	—	—	—	—	—	—	—	—	—	—
	Moderate	1	84	-0.20	—	-0.22	—	—	—	—	-0.74	0.25
	Subtle	2	126	-1.44	1.25	-1.72	1.66	5.75	-6.85	3.40	-14.05	14.05

Note: *k* = Number of effect sizes (*d* values); *N* = total sample size; *d* = sample size weighted mean effect size; *SD_d* = sample size weighted standard deviation of *d* values; δ = mean true effect size; *SD_δ* = standard deviation of effect sizes; %Var = percent variance accounted for in observed *d* values due to all corrected artifacts; CV₁₀ and CV₉₀ = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean effect size (δ)

Summary of Results

To aid in presenting a summary of the results, the overall within and between-group effect sizes are presented in Figure 6. This figure plots the effect size for each mediator for each level of analysis to help provide a visual comparison of the magnitude and direction of the various effect sizes. As can be seen in the figure, most mediator effect sizes were in the hypothesized direction and the pattern of effect sizes was similar across the within- and between-group results.

Specifically, stereotype threatened individuals had higher levels of anxiety and physiological arousal, more distracting (negative) thoughts, and demonstrated increased levels of effort than non-stereotype threatened individuals. They also had lower working memory capacity and decreased performance expectations. The performance avoidance effect sizes were the only results that did not follow the hypothesized pattern for both within- and between-group findings, in that the within-group effect was positive and the between-group effect size was negative.

However, both of these effect sizes were small and their confidence intervals included zero.

Generally, the between-group effect sizes were larger in magnitude than the within-group effect sizes. Although many of the between-group effects were larger than the within-group effects, many of these effects were based on a smaller number of data points with overall smaller sample sizes and their confidence intervals included zero.

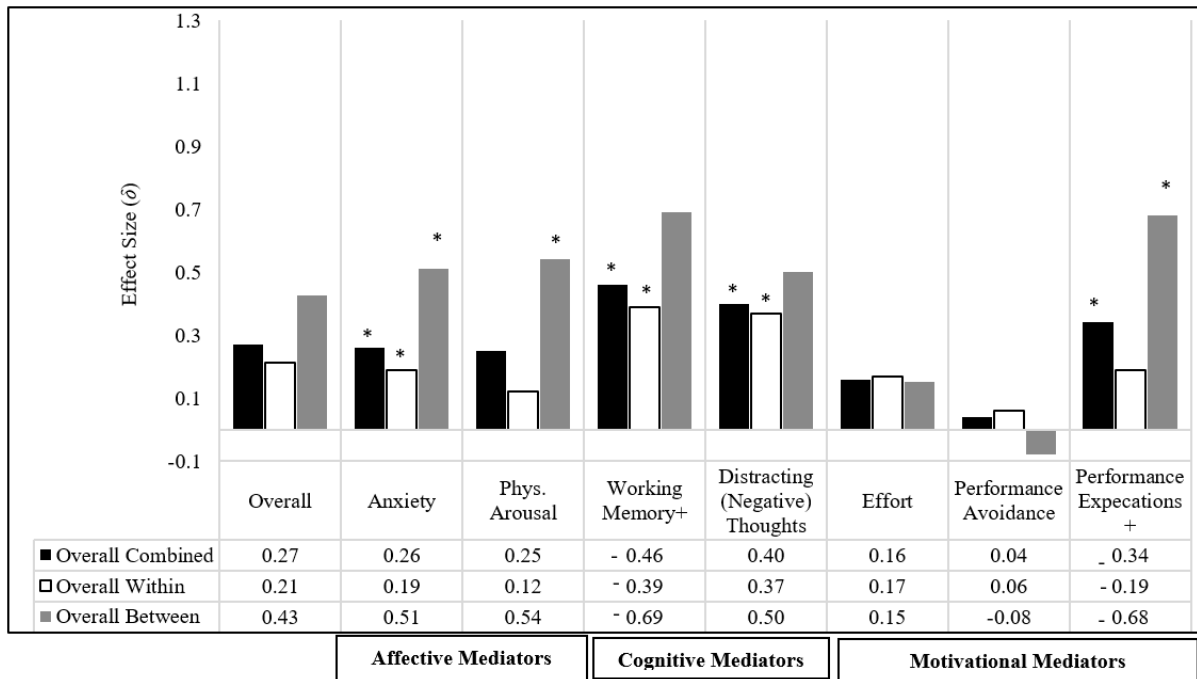


Figure 6. Overall mediator effect sizes *Denotes that confidence intervals do not include zero. +The effect sizes for Working Memory and Performance Expectations were reversed to ease the interpretation.

An overview of the findings for each hypothesis is presented in Table 19. Overall the affective mediator results provided the most consistent evidence for being mediators of the stereotype threat-performance relationship. Anxiety was higher for stereotype threatened individuals and this effect was consistent for both within- and between-group design results. Additionally, the between-group results for physiological arousal also provides some evidence that affective reactions are higher for stereotype threatened individual. However, this finding was not supported in the within-group design results.

Table 19

Summary of Results by Hypothesis

Hypothesis	Results
H1: Anxiety will be higher in stereotype threatened individuals.	Supported – Supported for both within and between-group results.
H2: Physiological arousal will be higher in stereotype threatened individuals.	Not supported – Effect size approached zero for within-group results. Between-group results were based on one data point (k).
H3: The mean effect size for physiological arousal will be larger than the effect size for anxiety for stereotype threatened individuals.	Not Supported – Effect sizes were not larger for physiological arousal than anxiety.
H4: Working memory capacity will be lower in stereotype threatened individuals than individuals not exposed to stereotype threat.	Supported – Supported for within-group results; effect size in hypothesized direction for between-group results, but confidence interval included zero.
H5: Stereotype threatened individuals will have more distracting (negative) than individuals not exposed to stereotype threat.	Supported – Supported for within-group results; effect size in hypothesized direction for between-group results, but confidence interval included zero.
H6: When stereotype threat is activated using subtle cues, stereotype threatened individuals will have more distracting (negative) thoughts as individuals not exposed to stereotype threat.	Supported – Supported for within-group results; effect size in hypothesized direction for between-group results, but confidence intervals included zero.
H7: When stereotype threat is activated using blatant cues, stereotype threatened individuals and individuals not exposed to stereotype threat will have a similar number of distracting (negative) thoughts.	Supported – Supported for both within and between-group results. Between-group results were based on one data point.
H8: Stereotype threatened individuals will exhibit increased effort on difficult tests compared to non-threatened individuals.	Not Supported – Effect sizes were in predicted direction for both within and between-group results; however, confidence intervals included zero.
H9: Stereotype threatened and non-threatened individuals will not differ in levels of effort on easy tests.	Supported – Supported for both within and between-group results. These results were based on one sample.
H10: Studies that use self-report measures of effort will produce smaller mean effect sizes for effort than studies that use more objective (e.g., items completed) measures.	Supported – Supported for both within and between-group results.
H11: Prevention-focus and avoidance orientation will be higher in stereotype threatened individuals compared to non-threatened individuals.	Not Supported – Prevention-focus and avoidance orientation was not higher for stereotype threatened individuals compared to non-threatened individuals. Between-group results were based on only two samples.
H12: Performance expectations will be lower in stereotype threatened individuals compared to non-threatened individuals.	Supported – Supported for between-group results, effect size in hypothesized direction for within-group results, but confidence intervals included zero.
H13: Studies that measure performance expectations prior to the completion of the task will produce larger mean effect sizes than studies that measure performance expectations after task completion.	Not Supported – Effect sizes were in predicted direction for both within and between-group results; however, confidence intervals included zero.

Hypothesis	Results
H14: Mediator effect size differences will be larger for stereotype threatened racial minorities than stereotype threatened women and older adults.	Not Supported –Mixed pattern of results across the different mediators (See Figure 7 and Figure 8)
H15: For women, the mediator effect sizes will be largest when stereotype threat is activated using subtle cues.	Not Supported – Gender-based effect sizes were not larger when stereotype threat was activated with subtle cues
H16: For racial minorities, the mediator effect sizes will be largest when stereotype threat is activated using moderate cues.	Not Tested – Sufficient data were not available across the activation levels for race-based effects.

The cognitive processes mediator findings also provide support for these types of processes being mediators of the stereotype threat-performance relationships. Working memory was lower for stereotype threatened individuals and distracting (negative) thoughts were higher for this group. Additionally, the moderator analyses suggest monitoring processes are activated when subtle cues are used and suppression processes are operating when blatant activation cues are leveraged. However, this support was only evident for the within-group findings.

Lastly, the mixed results for the motivational process mediators present more nuanced evidence for these types of processes being mediators of the stereotype threat-performance relationships. Based on the current meta-analysis results, performance avoidance is likely not a mediator of the stereotype threat-performance relationship. It is unclear if performance expectations are mediators of the stereotype threat-performance relationships because lowered expectations for the stereotype threat group were only observed for the between-group findings. The results for effort suggest that it may also impact the stereotype threat-performance relationship through increased levels of exerted effort, but this effect was only evident when effort was operationalized using objective rather than self-report measures.

The within and between-group results by stereotype threatened group are presented in Figure 7 and Figure 8. The overall effect size for each mediator is also included in the graphs. In

summary, the magnitudes of the effects vary, the findings for each group are largely consistent with the overall results. Anxiety, physiological arousal, and distracting (negative) thoughts were generally higher for stereotype threatened individuals, and working memory and performance expectations were generally lower for the stereotype threatened groups. Little to no stereotype threat effects were observed for effort and performance avoidance. However, these findings should be interpreted cautiously as many of these effects were based on a smaller number of data points with overall smaller sample sizes and their confidence intervals included zero. Additionally, although a number of moderators were explored, the results across many of the findings suggest the presence of additional moderators.

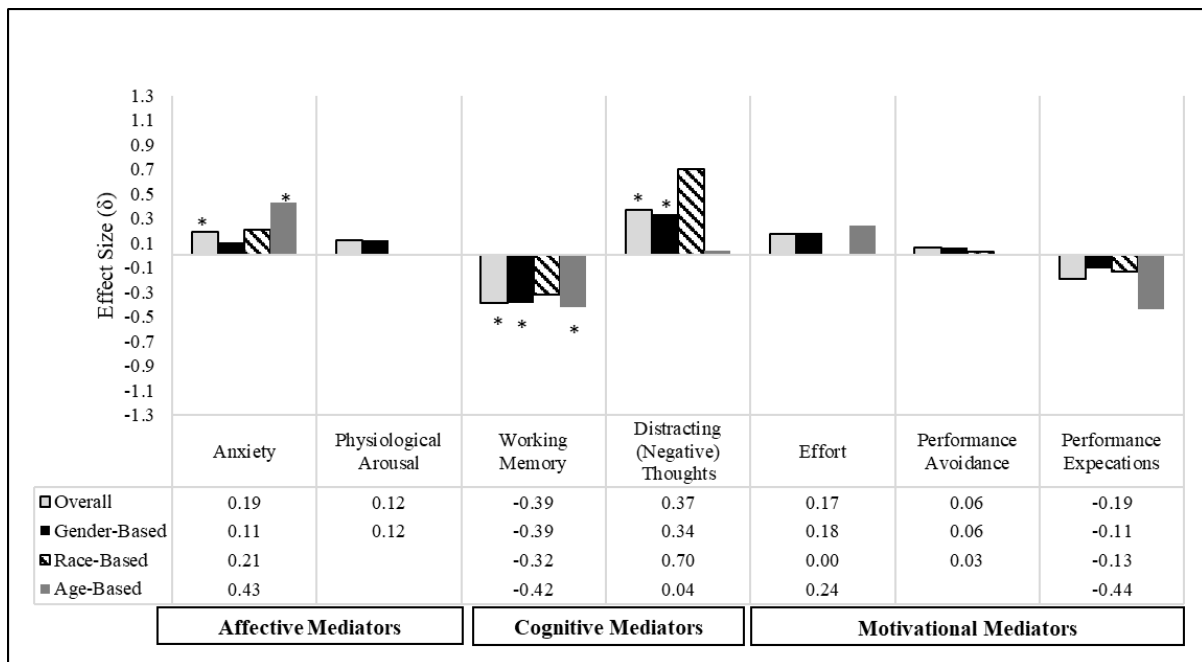


Figure 7. Within-group effect sizes by stereotype group. *Denotes that confidence intervals do not include zero

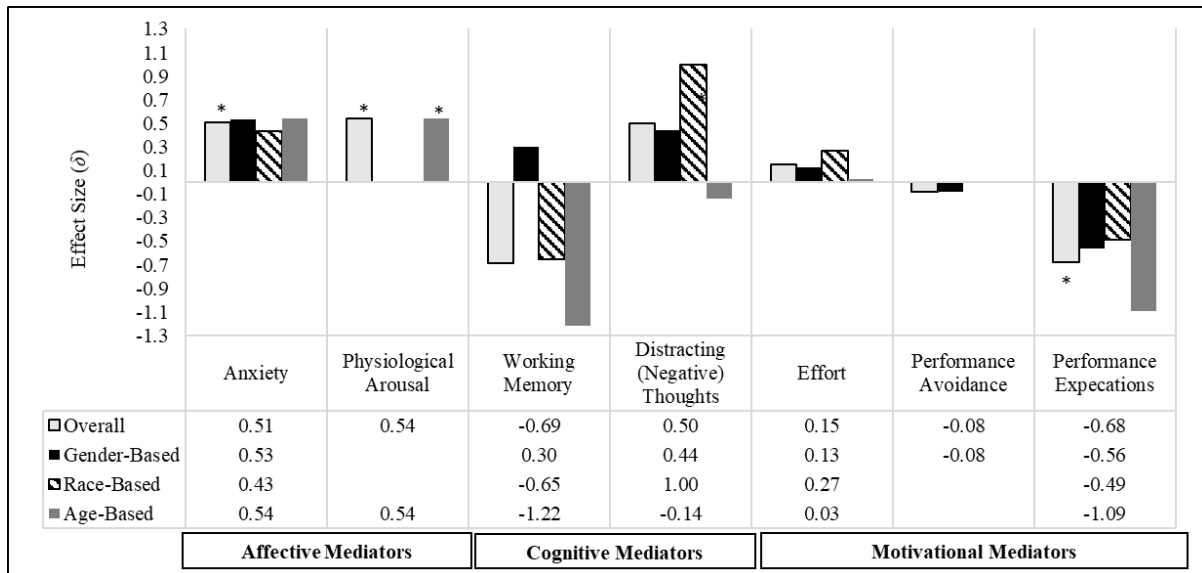


Figure 8. Between-group effect sizes by stereotype group. *Denotes that confidence intervals do not include zero .

DISCUSSION AND CONCLUSION

The goal of the current meta-analysis was to provide a qualitative and quantitative review of the mediators of the stereotype threat-test performance relationship. The current work builds on previous meta-analyses that documented evidence of a stereotype threat-performance relationship (Lamont et al., 2015; Nguyen & Ryan, 2008; Shewach et al., 2019) by investigating those factors that may connect stereotype threat to overall decreased test performance. Specifically, building on models proposed by Smith (2004), Schmader et al. (2008), and Pennington et al. (2016) the dissertation explored the degree to which affective, cognitive, and motivational mediators were impacted by stereotype threat. The purpose for exploring these relationships was to build a better understanding of the mediators of stereotype threat effects and inform a strategy for targeted interventions to combat stereotype threat.

The current meta-analysis presents results for both within-group and between-group designs. Figure 5 presents the different comparison groups that are included in within- and between-group designs. In the current study investigating the impact of stereotype threat across both design types was used to fully understand the nature of stereotype threat's effects. Within-group designs would allow us to test if stereotype threat is operating within a stereotyped group (e.g., women experiencing threat in mathematics testing situations). Between-group results essentially test the Group (stereotyped group [e.g., African Americans] vs. non-stereotyped group [e.g., Whites]) by Condition (stereotype threat vs. non-stereotype threat) interaction that is typically included in stereotype threat studies. These analyses are leveraged to help us better understand if stereotype threat explains group differences (e.g., Black-White cognitive ability test score differences). In the current study it was hypothesized that both within-group and between-group design effect sizes would follow the same hypothesized pattern. That is, if there

were differences in levels of the mediator variables for stereotype threatened groups vs. non-stereotype threatened groups these differences would be observed for both within- and between-group design results. However, it was expected that between-group results would be larger than the within-group results because differences on the variables of interest could be caused by stereotype threat, as well as a number of other factors related to belonging to a certain demographic group (e.g., women, African-American). Therefore, when taken to the between-group level of analyses any observed differences may be attributed to stereotype threat plus any factors related to belonging to each group.

Overall when the results are looked at in aggregate, the combined effect sizes follow the hypothesized pattern with the size of between-group effects being moderate while the within-group effect was smaller (see Figure 9). Additionally, the majority of the effect sizes followed the expected pattern (see Figure 6); stereotype threat impacted many of the proposed mediators and the effects tended to be larger for between-group effects. However, because a number of the effects had confidence intervals that included zero only two of the mediators should be interpreted. These are the main effect results for anxiety and the test operationalization moderator results for effort. For these effects both the within- and between group effects were meaningful (i.e., confidence intervals did not include zero) and as hypothesized the effect sizes followed the expected magnitude pattern with the between-group effects being larger than the within-group effects.

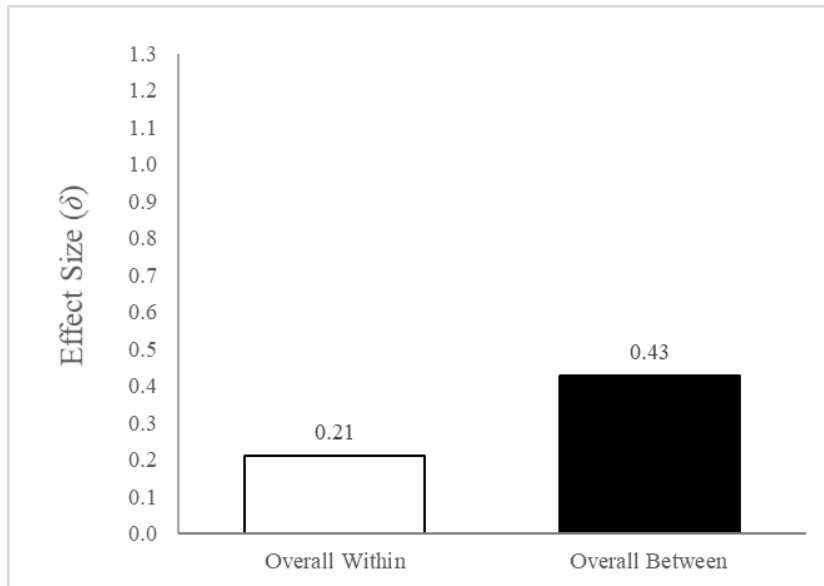


Figure 9. Aggregate within- and between-group design effect sizes. +The effect sizes for Working Memory and Performance Expectations were reversed to ease the interpretation.

However, although four of the other six mediators (working memory, distracting thoughts, physiological arousal, and performance expectations) followed a similar pattern many of these effects should not be interpreted as meaningful because the confidence intervals included zero. It is interesting to note that anxiety and effort had the most data points available to test for these effects, since confidence intervals tend to narrow as the number of data points increases it is not too surprising that these effects were the only meaningful effects. This seems to suggest that the lack of interpretable effects for the other mediators is likely because of second-order sampling error, that is the limited number of samples across many of the mediators limits the ability to detect meaningful effects.

Given the lack of consistent and meaningful effects across the seven mediators it is difficult to provide conclusive guidance on the development of intervention strategies that could target specific mediators. Consequently, the remaining parts of the discussion will first focus on

the implications of the anxiety and efforts results, then the discussion will highlight limitations as well as future research directions given the current results.

Review of Mediator Results

The purpose of the current meta-analysis was to determine the magnitude of a stereotype threat's effects on a number of proposed affective, cognitive, and motivation mediators. The purpose for exploring these effects was to determine the relative impact of stereotype threat on each mediator and subsequently create a framework for prioritizing intervention strategies that would target the negative impact of stereotype threat. However, given the wide-ranging and inconsistent pattern for the within- and between-group design results, the picture of stereotype threat's impact on these mediators remains largely unclear and there is still uncertainty on the direction forward. However, the results for anxiety and effort offer a promising starting point for understanding stereotype threat's mediator landscape. The implications for these results are discussed in further detail below.

The Negative Impact of Anxiety

Targeting anxiety is a possible avenue forward when looking to combat stereotype threat's negative effects. Stereotype threat's effects on anxiety were the most consistent effects in the current meta-analysis. The results of the meta-analyses demonstrate that anxiety was higher for those individuals exposed to stereotype threat, compared to those not exposed to stereotype threat. Stereotype threat's negative effects on anxiety were observed across both within- and between-group design results (see Table 7 and Table 8) and a similar pattern can also be observed when investigating these effects across the various subgroup analyses (see Table 11 and Table 12).

Anxiety's link to performance in a stereotype threat context can be understood through Eysenck et al.'s (2007) attentional control theory. The theory proposes that anxiety may affect performance effectiveness by interfering with processing efficiency. Efficiency is defined as the relationship between performance effectiveness and the effort or resources used to reach effectiveness (Eysenck & Calvo, 1992). Therefore, as anxiety increases more resources and effort are needed and subsequently used to reach performance effectiveness and as result performance efficiency decreases. This decrease in efficiency is likely particularly detrimental in high-stakes testing situations where the tests are difficult and oftentimes accompanied with a time limit.

It is also important to note that the effects of anxiety can be far-reaching and have an impact on a number of other mediators. Anxiety inhibits performance by reducing an individual's attentional control resources as they are focused on and distracted by the source of the anxiety (e.g., stereotype threat) rather than the task. A *stimulus-driven attentional system* drives these shifts in attention. Ideally individuals would be leveraging a *goal-directed attentional system* which is driven by an individual's expectations, knowledge, and goals rather than threat stimuli in the environment (Eysenck et al. 2007). Therefore, as anxiety increases individuals may have less control over their negative thoughts, be more distracted, and be more likely to direct their effort and motivation in a non-effective manner. As a result, the negative effects of anxiety likely have a cascading effect as they lead to negative consequences for other key processes.

The Complex Role of Effort

Consistent with much of the previous work exploring the stereotype threat and effort relationship the results were null across both the within and between-group results (within-group:

$\delta = 0.14$, $CI = -0.05 - 0.33$; between-group: $\delta = 0.07$; $CI = -0.34 - 0.50$), suggesting that effort is neither higher nor lower in stereotype threatened groups. These results are none too surprising given that effort has been proposed to both increase (motivation to not fail and confirm a negative stereotype) and decrease (through self-handicapping behaviors) in stereotype threat scenarios. However, how effort is measured (self-report vs. objective measures) and the nature of the tasks (difficult vs. easy) has further complicated the understanding of the stereotype threat-effort relationship.

Although the results investigating effort levels on tasks of varying difficulty were inconclusive, the results investigating measurement type provides some clarity on understanding effort as a stereotype threat mediator. When effort is measured via self-report measures, there is no difference between stereotype threatened and non-stereotype threatened individuals. However, when effort was measured via objective measures (e.g., counting the number of items individuals completed during a designated test time), there is evidence that stereotype threatened individuals exhibit higher levels of effort. This pattern was consistent across both the within- and between-group design results ($\delta = 0.38, 0.70$). Additionally, in line with the hypothesis, the between-group effect size was larger than the within-group effect.

The effort measurement type results begin to bring some clarity to the stereotype threat-effort relationship. If individuals are in fact exhibiting higher levels of effort in stereotype threat scenarios, as the objective measure results suggest, but are not seeing the payoff of their efforts they may be unwilling to admit on a self-report measure that they were working hard and putting effort towards a given task. Additionally, Jamieson and Harkins (2008) proposed that although individuals may exhibit increased effort in stereotype threat scenarios, they may actually be engaging in behaviors that are counterproductive and could hinder their performance. That is, as

individuals increase their effort on a task they may be less likely to pick-up on cues that they are making mistakes and could struggle to course correct to eliminate their mistakes. As a result, although effort is higher the increased energy is in vain because it is being directed in an inefficient manner.

Practical Implications

Strategies to combat the negative effects of stereotype threat have been explored since Steele and Aronson's (1995) seminal article, when the authors removed the cue that asked participants to provide their race before an exam. The purpose of the present study was to inform the strategies that are used to combat the effects of stereotype threat. Although the results did not provide conclusive results for the evidence of many of the proposed mediators, the results for anxiety suggest that it is a promising avenue to begin addressing stereotype threat effects. Additionally, the findings for effort imply that it may be another promising area to target with interventions.

Combatting Anxiety

Anxiety is likely inherent in any type of evaluation context, and its effects are particularly detrimental in high-stakes testing situations (von der Embse et al., 2017). When addressing general test anxiety interventions that focus on behavioral, cognitive, and skill-based strategies have proven effective (Ergene, 2003; Lovett & Nelson, 2017; von Der Embse, Barterian, & Segool, 2013). Thus, an effective strategy that targets stereotype threat's effects on anxiety will need to incorporate these components as well as incorporating tactics that are specific to the stereotype threat experience. Specifically targeting components of the evaluation environment that trigger the heightened state of arousal and anxiety is key. For example, Steele and Aronson (1995) and Danaher and Crandall (2008) found that by simply not asking questions of race or

gender identification questions prior to a test resulted in a significant reduction in typical stereotype threats effects on test performance. When Steele and Aronson (Study 4) tested the effect of removing the requirement to indicate race before a test in an experimental setting the non-threatened groups performance was more than a standard deviation higher than the threatened group ($d = 1.09$); when Danaher and Crandall replicated this experiment in a real-world test settings they found that changing the timing of asking demographic questions (i.e., moving the question after the test) reduced male-female test score differences by 33%. These results provide a case for the value in removing cues that might heighten individuals' awareness of their group identity. Heightening awareness of a group identity could impact an individual's awareness of being evaluated based on negative stereotypes tied to that identity; however, when a negatively stereotyped group identity is not brought to the forefront of the individual's attention it could result in decreased levels of arousal for these individuals and subsequently have a positive impact on test performance.

Redirecting Effort

The current study revealed that individuals who were exposed to stereotype threat demonstrated higher levels of effort when completing tasks. However, this effect was only seen when effort was measured via objective measures rather than self-report measures. Therefore, although individuals may be exhibiting higher levels of effort, they may not necessarily be cognizant or willing to admit the additional energy they are directing towards a task.

Higher levels of effort may inherently seem like a positive outcome, and this is likely true across a number of scenarios, but past stereotype threat research has demonstrated that higher levels of effort could actually be damaging. Jamieson and Harkins' (2007, 2009) *mere effort theory* hypothesizes that stereotype threatened individuals are motivated to perform well on tasks

so they do not confirm negative stereotypes. That is, they will demonstrate increased levels of effort to quickly respond to problems and quickly move to the next to complete as many problems as possible in a given timeframe. However, as an individual is putting forth increased effort the stereotype threatened individual may lean on their dominant responses and may not adequately adjust when a situation or problem requires a different approach, especially if they are not given enough time to correct their actions. As a result, although these individuals are motivated to succeed and are putting in the effort, stereotype threat negatively impacts the performance of individuals on novel tasks that require a different approach to solving a problem. Consequently, in these types of situations instructing stereotype threatened individuals to “Work smarter, not harder,” is likely good advice. In line with this advice, Jamieson and Harkins (2009) found that providing test-takers a clear outline on the different problem-types on a test and the need to use different strategies to solve the different types of problems eliminated test performance differences between threatened and non-threatened female test-takers. Therefore, one line of defense against the negative implications that are the result of misaligned effort may be to explicitly state in instructions to test-takers what they should expect in the test ahead so that they can appropriately prepare and align the necessary energy.

Limitations

Meta-analytic techniques are touted for their comprehensive approach to addressing and exploring important phenomenon but they are not without their limitations. The current study focused on testing specific hypotheses related to the mediators of stereotype threat; however, a notable limitation is second-order sampling error (Hunter & Schmidt, 2004). Although stereotype threat as a construct has been studied quite extensively, the work investigating the different proposed mediators has not been as comprehensive. For example, although a large

number of studies that could possibly be included in the current meta-analysis were identified, a number of them had to be excluded because they did not measure the necessary mediators or did not include the statistical information to compute effect sizes for the mediators that were of interest for the current study. One observation was that much attention has been placed test performance as the key criterion in stereotype threat research. As a result, although many of the variables of interest (e.g., anxiety, working memory, effort) were mentioned across different stereotype threat studies because they were not the focal point of a study the results for some of the secondary variables (i.e., mediator) were not reported.

The current study could be at risk of second-order sampling error. That is, the studies included in the current meta-analysis may not be a “good” representation of possible studies investigating the relationships between stereotype threat and the different mediators. Although second-order sampling error is likely not going to be entirely eliminated, when the meta-analytic sample is small the likelihood of second-order sampling error increases and negatively impacts the ability to estimate the population effect and variance. Attempts were made to obtain a complete and full set of data points. First, a broad search was used to ensure the dataset was comprehensive. When variables were mentioned in studies but not included in the reports, study authors were contacted. However, the attempts to gather the data were not always successful (e.g., seventy-two authors were contacted regarding missing study data, but only eleven responded to requested with the necessary data).

The tendency to treat stereotype threat as a panacea for all subgroup differences makes it a particularly visible target for skepticism. This visibility highlights the importance of taking publication bias quite seriously (Ryan & Nguyen, 2017). First, attempts were made to be as comprehensive as possible and roughly a third of the included studies were from non-published

research reports (26 of 80 reports, 32.5%). Although substantial effort was put forth to identify all possible studies for the meta-analysis, there is still some risk that studies could have been overlooked, so as a next step, a number of analyses were conducted to test for publication bias and the majority of these results suggest that publication bias is not a major concern in the current meta-analysis (see Table 6).

Although several efforts were made to ensure the dataset was as comprehensive as possible, some of the mediator variables did not have many data points. This may be because the research in this area has not uncovered significant results and as a result have not been published. However, it is more likely that there has simply not been substantial exploration in these areas because of the novelty of some of these variables (i.e., performance avoidance) or the difficulty in capturing these types of measures (i.e., physiological arousal). Additionally, as research in this area grows it will also allow for further exploration of the different moderators that may impact the stereotype threat-mediator relationships. Exploration of these moderator relationships was an additional area where a limited number of data points impacted the ability to draw conclusions and exploring these moderators presents another topic for future research.

The current study tested for a number of different moderators, those specific to particular mediators (e.g., measurement timing for performance expectations), as well as moderators that were expected to impact all the mediators (e.g., stereotype target). Similar to Nguyen and Ryan (2008), hierarchical moderator analyses were conducted to test for group-based and activation effects. A small number of data points limited the ability to test for a number of these effects. Additionally, even after accounting for different moderators the results revealed that additional moderators might exist (because the percent of observed variance being accounted for was less than 75%).

Another potential concern may be the decision to only test the IV-mediator relationship in the current meta-analysis (Step 1 in Figure 3). That is, the purpose of the current meta-analysis was to test for mediators of the stereotype threat-performance relationship. Traditionally, tests for mediation (Baron & Kenny, 1986) include a test of the IV-DV relationship (e.g., stereotype threat to performance, Step 2 in Figure 3), a test of the IV-mediator relationship (e.g., stereotype threat to proposed affective, cognitive, and motivational mediators, Step 1 in Figure 3), a test of the mediator-DV relationship (e.g., proposed mediators to performance, Step 3 in Figure 3), and finally a test of the IV to DV relationship accounting for the proposed mediators (Step 4 in Figure 3). Although this approach for testing for mediation is disputed (MacKinnon & Fairchild, 2009; O'Rourke & MacKinnon, 2018), it does provide a useful model for understanding the different relationships that impact the mediation process.

Using Baron and Kenny's (1986) step-by-step mediation model as a guide is also helpful in understanding the current state of the literature and where there may be gaps in the exploration of different relationships. A number of meta-analyses have established the link between stereotype threat and performance (Lamont et al., 2015; Nguyen & Ryan, 2008; Shewach et al., 2019), and the link between the proposed mediators and performance have also been established (e.g., der Embse, Jester, Roy, & Post, 2017; Hausknecht et al., 2004; Oberauer, Schulze, Wilhelm, & Süß, 2005; von Unsworth & Engle, 2005;). The lack of any reviews that quantified the IV-mediator relationship for the stereotype threat-performance relationship prompted the current meta-analysis. The current meta-analysis only focused on the IV-mediator relationship and did not test the full mediation model therefore there is still a necessity to test the full mediational model including all the steps. However, given the results of the current meta-

analysis and the still unclear picture of stereotype threat's impact on a number of mediators this is likely an opportunity for future research rather than a question to address now.

Future Research

The current study aimed to advance the stereotype threat literature by providing a qualitative and quantitative review of the mediators of stereotype threat effects. Specifically, to date there have only been qualitative reviews of stereotype threat effects and their potential mediators (Pennington et al., 2016; Schmader et al., 2008; Smith, 2004). The current study moves past a qualitative review and provides a preliminary summary of the magnitudes of the proposed mediators (see Figure 6). Although the present study addresses a critical gap in the literature, there are still a number of areas of future research. These areas are discussed below.

The notable lack of a clear pattern of results across the within- and between-group design findings highlights a dearth of research across many of the mediators. In particular, although both performance avoidance and physiological arousal have been proposed as potential mediators of the stereotype threat-performance relationship the research in these areas is somewhat lacking. As noted previously, measures of physiological arousal often require extensive equipment and training that may not be feasible for many researchers. Additionally, this work has largely been confined to laboratory settings. However, recent advances in technology have opened opportunities to leverage physiological measures with less interference and in more applied settings (Ganster, Crain, & Brossoit, 2018). With these advances there are likely opportunities for stereotype threat research to capitalize on these technologies. For example, it may be possible for research participants to wear devices (e.g., FitBit, Apple Watch) during testing scenarios to measure the impact stereotype threat has on heartrate, blood pressure, and a number of other physiological responses. Additionally, sending message through these

devices may be one way to target and mitigate threat (e.g., targeting anxiety through messages to breathe and remain focused).

Targeting how individuals approach a stereotype threatening scenario has been one proposed way to mitigate the deleterious effects of stereotype threat. Indeed, past research has demonstrated that since stereotype threatened individuals are likely to adopt a performance avoidance approach, when testing scenarios are framed in an avoidant manner (e.g., encouraging an approach that focuses on minimizing or avoiding errors rather than promoting finding the correct answer) stereotype threatened individuals work more effectively and have more positive test outcomes (Chalabaev et al., 2012). Additionally, Alter, Aronson, Darley, Rodriguea, & Rube (2009) found when stereotype threat was reframed as challenge where one can achieve rather than a threat that one should avoid, stereotype threat effects were reduced. Despite these positive outcomes the current study did not find evidence for performance avoidance being a mediator of stereotype threat. It is important to note that this finding is based on a fairly small number of data points (within-group $k = 8$, between-group $k = 2$). The smaller number of data points may be due to how performance avoidance is operationalized in studies. Performance avoidance is often included in stereotype threat research as an additional experimental condition (Chalabaev et al., 2012; Finnigan & Corker, 2016) and not tested as an outcome variable. Although performance avoidance focused strategies have proved fruitful it is still unclear if the performance avoidance approach is the true mediator that is operating. Therefore, an area of future research may be to include performance avoidance and approach measures with stereotype threat manipulations to more clearly understand how stereotype threat is impacting these processes.

The impetus for the current study was the recognition of the lack of interventions to address stereotype threat in applied settings, specifically employment settings. Although the

current meta-analysis hoped to inform the creation of targeted interventions, the lack of consistent findings suggest that more work needs to be done in understanding the mediational processes before intervention research can be truly effective. Therefore, research investigating the efficacy of interventions is likely further in the future. It will also be important that these interventions are tested in the appropriate settings. Much of the work investigating the efficacy of interventions have been tested in laboratory or educational settings and not with working adult populations. Recently, Shewach et al. (2019) investigated study setting (laboratory vs. operational samples) as a moderator for the stereotype threat-performance relationship and only identified four operational samples in comparison to 177 laboratory samples. Additionally, in the current meta-analysis study setting was coded and only three samples in the dataset were conducted in an operational setting. This lack of research in operational settings highlights the need to consider organizational demands as stereotype threat research continues to grow. It will also be important that stereotype threat interventions are designed to work and accommodate the demands and expectations of organizational testing settings.

Additionally, as the efficacy of these interventions are explored it will be important to determine their efficacy across different stereotype threatened groups. Nguyen and Ryan (2008) noted the differential impact of stereotype threat on women and ethnic minorities and speculated that these different groups may react to stereotype threat in different manners. There were a limited number of datapoints available for these analyses so Hypothesis 15 and Hypothesis 16 could not be fully explored in the current study, but results do point to potentially different experiences. Specifically, stereotype threat may negatively impact women through its effects on cognitive mediators, while it may be impacting older adults through affective processes. These

results highlight the importance of understanding what interventions strategies will be most impactful to different groups.

Finally, the current study focused on stereotype threat effects on mediators in ability testing scenarios; however, stereotype threat implications are not just limited to the test setting. Stereotype threat effects have been demonstrated for athletic ability (Stone et al., 1999), driving (Jonaisse, Gagnon, & Voloaca, 2013), social, (Bosson et al., 2004) and management skills (Flanagan, 2015) amongst others. Therefore, there is an opportunity to explore mediators when stereotype threat manifest in these settings. Ostensibly, there are likely some mediators that operate across these different settings (e.g., anxiety in driving and cognitive testing settings), additionally there is also a possibility that these other experiences of stereotype threat likely elicit a different set of responses (e.g., Beilock et al., 2006) and future research can explore how these reactions manifest and how these experiences may differ across different operationalizations of stereotype threat.

Conclusion

There is a well-documented history of stereotype threat effects across a number of different settings and stereotyped groups. Although there has been some progress, strategies to mitigate stereotype threat have not been broadly successful. This lack of progress may be due to premature attempts to target stereotype threat without fully understanding the process that underlie these effects. The current dissertation sought to push forward the progress of stereotype threat interventions by providing direction on what processes should be targeted through intervention strategies. However, given the lack of a consistent pattern of results across the within- and between-group findings the understanding of stereotype threat's impact on a variety of mediators remains largely unclear. These results (or lack thereof) point to a need for a more

targeted approach to studying stereotype threat's impact past just its negative impact on test performance. The link between stereotype threat and performance has been determined (although there are still some critics regarding the magnitude of the impact) and it is now time to move forward and seek to build a more comprehensive and detailed understanding of the phenomenon. Continued work in understanding stereotype threat's effect on affective, cognitive, and motivational mediators will advance progress in understanding this attention-grabbing, yet not fully understood topic and hopefully move us further in addressing subgroup differences in high-stakes testing settings.

REFERENCES

References marked with an asterick indicate sample contributed an effect size to the meta-analysis.

- *Abrams, D., Crisp, R. J., Marques, S., Fagg, E., Bedford, L., & Provias, D. (2008). Threat inoculation: experienced and imagined intergenerational contact prevents stereotype threat effects on older people's math performance. *Psychology and Aging, 23*, 934-939.
- *Abrams, D., Eller, A., & Bryant, J. (2006). An age apart: the effects of intergenerational contact and stereotype threat on performance and intergroup bias. *Psychology and Aging, 21*, 691-702.
- Aguinis, H., Werner, S., Lanza Abbott, J., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods, 13*, 515-539.
- *Allen, B., & Friedman, B. H. (2016). Threatening the heart and mind of gender stereotypes: Can imagined contact influence the physiology of stereotype threat? *Psychophysiology, 53*, 105-112.
- Alter, A. L., Aronson, J., Darley, J. M., Rodriguez, C., & Ruble, D. N. (2010). Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. *Journal of Experimental Social Psychology, 46*, 166-171.
- Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A., & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology, 40*, 401-408.

- *Aneyeji, A. I. (2013). *Stereotype threat, impostorism, and Black American women: Examining how clinical and social psychological variables affect achievement* (Unpublished doctoral dissertation). Long Island University, Brooklyn, NY.
- Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: Review and meta-analysis. *Frontiers in psychology*, 6, 900.
- 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.
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational behavior and human decision processes*, 50, 248-287.
- Barber, S. J., & Mather, M. (2013). Stereotype threat can both enhance and impair older adults' memory. *Psychological Science*, 24, 2522-2529.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- *Beaton, A., Tougas, F., Rinfret, N., Huard, N., & Delisle, M.-N.. (2007). Strength in numbers? Women and mathematics. *European Journal of Psychology of Education*, 22, 291-306.
- *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.
- Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin*, 32, 1059-1071.

- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*, 174-181.
- *Berjot, S., Roland-Levy, C., & Girault-Lidvan, N. (2011). Cognitive appraisals of stereotype threat. *Psychological reports, 108*, 585-598.
- Blascovich, J., Spencer, S. J., Quinn, D. M., & Steele, C. M. (2001). African-Americans and high blood pressure: The Role of stereotype threat. *Psychological Science, 12*, 225-229.
- Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implication of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52*, 561-589.
- Bosson, J. K., Haymovitz, E. L., & Pintel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology, 40*, 247-255.
- *Bouazzaoui, B., Follenfant, A., Ric, F., Fay, S., Croizet, J.-C., Atzeni, T., & Tacconnat, L. (2015). Ageing-related stereotypes in memory: When the beliefs come true. *Memory, 24*, 659-668.
- *Boucher, K. L., Rydell, R. J., & Murphy, M. C. (2015). Forecasting the experience of stereotype threat for others. *Journal of Experimental Social Psychology, 58*, 56-62.
- *Brodish, A. B. (2008). *Stereotype threat and achievement goals: An integrative approach*. (Unpublished doctoral dissertation). University of Wisconsin-Madison, Madison, WI.
- Brodish, A. B., & Devine, P. G. (2009). The role of performance–avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of experimental social psychology, 45*, 180-185.

- *Brown, R. P., & Day, E. A. (2006). The difference isn't Black and White: Stereotype threat and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology, 91*, 979-985.
- Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76*, 246-257.
- *Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33*, 267-285.
- *Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 16*, 572-578.
- *Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*, 853-859.
- Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: when the "have-nots" gain but the "haves" gain even more. *American Psychologist, 60*, 149-160.
- Chalabaev, A., Major, B., Sarrazin, P., & Cury, F. (2012). When avoiding failure improves performance: Stereotype threat and the impact of performance goals. *Motivation and Emotion, 36*, 130-142.
- Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal of educational Psychology, 97*, 268-274.

- *Chasteen, A. L., Bhattacharyya, S., Horhota, M., Tam, R., & Hasher, L. (2005). How Feelings of Stereotype Threat Influence Older Adults' Memory Performance. *Experimental Aging Research, 31*, 235-260.
- Cohen, G. L., & Garcia, J. (2005). "I am us": Negative stereotypes as collective threats. *Journal of Personality and Social Psychology, 89*, 566-582.
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*, 1307-1310.
- Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science, 324*, 400-403.
- Cohen, G. L., Purdie-Vaughns, V., & Garcia, J. (2012). An identity threat perspective on intervention. In M. Inzlicht & T. Schmader (Eds.) *Stereotype Threat: Theory, Process, and Application*. (pp.124-140). New York: Oxford University Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Colom, R., Martínez-Molina, A., Shih, P. C., & Santacreu, J. (2010). Intelligence, working memory, and multitasking performance. *Intelligence, 38*, 543-551.
- 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-Klaassen, J. M. (2010). *Stereotype threat's effect on women's achievement in chemistry: The interaction of achievement goal orientation for women in science majors*. (Unpublished doctoral dissertation). University of Nevada - Las Vegas, Las Vegas, NV.

- Cortina, J. M., & Landis, R. S. (2011). The earth is not round ($p = .00$). *Organizational Research Methods, 14*, 332-349.
- *Cotting, D. I. (2003). *Shedding light in the black box of stereotype threat: The role of emotion*. Unpublished doctoral dissertation). City University of New York, New York City, NY.
- Creswell, J.D., Welch, W.T., Taylor, S.E., Sherman, D.K., Greunewald, T.L., & Mann, T. (2005). Affirmation of personal values buffers neuroendocrine and psychological stress responses. *Psychological Science, 16*, 846-851.
- Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological review, 96*, 608-630.
- Croizet, J.-C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588-594.
- Croizet, J. C., Després, G., Gauzins, M. E., Huguet, P., Leyens, J. P., & Méot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30*, 721-731.
- Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology, 89*, 220-230.
- Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*, 1639-1655.
- *Delgado, A. R., & Prieto, G. (2008). Stereotype threat as validity threat: The anxiety-sex-threat interaction. *Intelligence, 36*, 635-640.

- *Desrichard, O., & Köpetz, C. (2005). A threat in the elder: the impact of task-instructions, self-efficacy and performance expectations on memory performance in the elderly. *European Journal of Social Psychology, 35*, 537-552.
- *Dinella, L. M. (2004). *A developmental perspective on stereotype threat and high school mathematics* (Unpublished doctoral dissertation). Arizona State University, Tempe, AZ.
- Downey, S. N., van der Werff, L., Thomas, K. M., & Plaut, V. C. (2015). The role of diversity practices and inclusion in promoting trust and employee engagement. *Journal of Applied Social Psychology, 45*, 35-44.
- Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology, 67*, 382-394.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23.
- Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999). Working memory, short term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.
- Ergene, T. (2003). Effective interventions on test anxiety reduction: A meta-analysis. *School Psychology International, 24*, 313-328.
- Eysenck, M. W. (1992). *Anxiety: The cognitive perspective*. Hove, UK: Lawrence Erlbaum Associates Ltd.
- Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*, 409-434.

- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion, 7*, 336-353.
- Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009). R² effect-size measures for mediation analysis. *Behavior Research Methods, 41*, 486-498.
- Finnigan, K. M., & Corker, K. S. (2016). Do performance avoidance goals moderate the effect of different types of stereotype threat on women's math performance?. *Journal of Research in Personality, 63*, 36-43.
- Flanagan, J. (2015). Gender and the workplace: The impact of stereotype threat on self-assessment of management skills of female business students. *Advancing Women in Leadership, 35*, 166-171.
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of school psychology, 53*, 25-44.
- Foote, W. E. (2000). *The effect of stereotype threat on women's math performance: an examination of mediating and moderating variables*. (Unpublished doctoral dissertation). California State Fullerton, Fullerton, CA.
- *Forbes, C. E., Leitner, J., Duran-Jordan, K., Magerman, A., Schmader, T., & Allen, J. J. (2014). Spontaneous default mode network phase-locking moderates performance perceptions under stereotype threat. *Social Cognitive & Affective Neuroscience, 10*, 994-1002.
- Forbes, C. E., Schmader, T., & Allen, J. J. (2008). The role of devaluing and discounting in performance monitoring: a neurophysiological study of minorities under threat. *Social Cognitive and Affective Neuroscience, 3*, 253-261.

- Ganster, D. C., Crain, T. L., & Brossoit, R. M. (2018). Physiological measurement in the organizational sciences: A review and recommendations for future use. *Annual Review of Organizational Psychology and Organizational Behavior*, 5, 267-293.
- Garcia, J., & Cohen, G. L. (2012). A social psychological approach to educational intervention. In E. Shafir (Ed.), *Behavioral Foundations of Public Policy* (pp. 329-347). Princeton, NJ: Princeton University Press.
- Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology*, 29, 17-28.
- *Good, J. J., Woodzicka, J. A., & Wingfield, L. C. (2010). The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *The Journal of Social Psychology*, 150, 132-147.
- *Grand, J. A., Ryan, A. M., Schmitt, N., & Hmurovic, J. (2011). How far does stereotype threat reach? The potential detriment of face validity in cognitive ability testing. *Human Performance*, 24, 1-28.
- Gresky, D. M., Ten Eyck, L. L., Lord, C. G., & McIntyre, R. B. (2005). Effects of salient multiple identities on women's performance under mathematics stereotype threat. *Sex Roles*, 53, 703-716.
- Grimm, L. R., Markman, A. B., Maddox, W. T., & Baldwin, G. C. (2009). Stereotype threat reinterpreted as a regulatory mismatch. *Journal of Personality and Social Psychology*, 96, 288-304.
- *Hall, W. (2012). *Stereotype threat impairs the feeling of learning*. (Unpublished doctoral dissertation). University of British Columbia, Vancouver, Canada.

- *Hardee Bailey, A. A. (2004). *Effects of stereotype threat on females in math and science fields: An investigation of possible mediators and moderators of the threat-performance relationship*. (Unpublished doctoral dissertation). Georgia Institute of Technology, Atlanta, GA.
- *Harder, J. A. (2000). *The effect of private versus public evaluation on stereotype threat for women in mathematics*. (Unpublished doctoral dissertation). University of Texas- Austin, Austin, Texas.
- Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology*, *91*, 436- 455.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*, 639-683.
- *Hess, T. M., Hinson, J. T., & Hodges, E. A. (2009). Moderators of and mechanisms underlying stereotype threat effects on older adults' memory performance. *Experimental aging research*, *35*, 153-177.
- Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist*, *55*, 1217-1229.
- *Howard, K. E. (2005). *Stereotype threat and anxiety: Examining their connection to working memory capacity in middle school children* (Unpublished doctoral dissertation) University of Southern California, Los Angeles, CA.
- Hunter, J. E. & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings (2nd edition)*. Thousand Oaks, CA: Sage Publications.

- *Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95*, 796-805.
- Inzlicht, M. & Schmader, T. (2012). *Stereotype threat: Theory, process, and application*. New York: Oxford University Press.
- *Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93*, 544-564.
- *Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin, 35*, 1301-1314.
- *Jamieson, J. P., & Harkins, S. G. (2011). The intervening task method: implications for measuring mediation. *Personality & Social Psychology Bulletin, 37*, 652-661.
- Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of Experimental Social Psychology, 46*, 208-212.
- John-Henderson, N. A., Rheinschmidt, M. L., & Mendoza-Denton, R. (2015). Cytokine responses and math performance: The role of stereotype threat and anxiety reappraisals. *Journal of Experimental Social Psychology, 56*, 203-206.
- *Johns, J., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General, 137*, 691-705.

- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*, 175-179.
- Joanisse, M., Gagnon, S., & Voloaca, M. (2013). The impact of stereotype threat on the simulated driving performance of older drivers. *Accident Analysis & Prevention, 50*, 530-538.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189-217.
- *Kang, S. K., & Chasteen, A. L. (2009). The moderating role of age-group identification and perceived threat on stereotype threat among older adults. *International Journal of Aging & Human Development, 69*, 201-220.
- *Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 47*, 193-198.
- *Keller, J., & Bless, H. (2008). When positive and negative expectancies disrupt performance: Regulatory focus as a catalyst. *European Journal of Social Psychology, 38*, 187-212.
- *Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29*, 371-381.
- Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods, 15*, 624-662.

- Kovenklioglu, G., & Greenhaus, J. H. (1978). Causal attributions, expectations, and task performance. *Journal of Applied Psychology, 63*, 698-705.
- Kravitz, D. A. (1995). Attitudes toward affirmative action plans directed at blacks: Effects of Plan and Individual Differences¹. *Journal of Applied Social Psychology, 25*, 2192-2220.
- Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology, 80*, 942-958.
- *Krotenberg, L. A. (2003). *Cognition and emotions in stereotype threat in African American and Hispanic college students*. (Unpublished doctoral dissertation). St. John's University, Queens, NY.
- Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and Aging, 30*, 180-193.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.
- Leary, M. R., & Shepperd, J. A., (1986). Behavioral self-handicaps versus self-reported handicaps: A conceptual note. *Journal of Personality and Social Psychology, 51*, 1265-1268.
- Lewis, N. A., & Sekaquaptewa, D. (2016). Beyond test performance: A broader view of stereotype threat. *Current Opinion in Psychology, 11*, 40-43.
- Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports, 20*, 975-978.

- Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., von Hippel, W., & Bell, A. E. (2009). Interacting with sexist men triggers social identity threat among female engineers. *Journal of Personality and Social Psychology, 96*, 1089-1103.
- Lovett, B. J., & Nelson, J. M. (2017). Test anxiety and the Americans with disabilities act. *Journal of Disability Policy Studies, 28*, 99-108.
- Lu, A., Feng, Y., Tian, H., Yu, Z., Hong, X., & Zheng, D. (2015). Anxiety and mind wandering as independent consequences of stereotype threat. *Social Behavior and Personality, 43*, 537-546.
- Lupien, S. J., & McEwen, B. S. (1997). The acute effects of corticosteroids on cognition: integration of animal and human model studies. *Brain Research Reviews, 24*, 1-27.
- MacKinnon, D. P., & Fairchild, A. J. (2009). Current directions in mediation analysis. *Current Directions in Psychological Science, 18*, 16-20.
- *Mangels, J. A., Good, C., Whiteman, R. C., Maniscalco, B., & Dweck, C. S. (2012). Emotion blocks the path to learning under stereotype threat. *Social Cognitive & Affective Neuroscience, 7*, 230-241.
- Marx, D. M. (2012). Differentiating theories: A comparison of stereotype threat and stereotype priming effects. In M. Inzlicht & T. Schmader (Eds.) *Stereotype Threat: Theory, Process, and Application*. (pp.124-140). New York: Oxford University Press.
- Marx, D. M. & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology, 44*, 645-657.

- Marx, D. M., & Stapel, D. A. (2006). Distinguishing stereotype threat from priming effects: on the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology, 91*, 243-254.
- *Marx, D. M., Stapel, D. A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparisons under threat. *Journal of Personality and Social Psychology, 88*, 432-446.
- *Mazerolle, M., Régner, I., Morisset, P., Rigalleau, F., & Huguet, P. (2012). Stereotype Threat Strengthens Automatic Recall and Undermines Controlled Processes in Older Adults. *Psychological Science, 23*, 723-727.
- McFarland, L. A., Lev-Arey, D. M., & Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Human Performance, 16*, 181-205.
- McIntyre, R. B., Lord, C. G., Gresky, D. M., Ten Eyck, L. L., Frye, G. J., & Bond Jr, C. F. (2005). A social impact trend in the effects of role models on alleviating women's mathematics stereotype threat. *Current Research in Social Psychology, 10*, 116-36.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology, 39*, 83-90.
- *McKay, P. F. (1999). *Stereotype threat and its effect on the cognitive ability test performance of African-Americans: The development of a theoretical model*. (Unpublished doctoral dissertation). University of Akron, Akron, OH.
- *Mrazek, M. D., Chin, J. M., Schmader, T., Hartson, K. A., Smallwood, J., & Schooler, J. W. (2011). Threatened to distraction: Mind-wandering as a consequence of stereotype threat. *Journal of Experimental Social Psychology, 47*, 1243-1248.

- Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*, 247-259.
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*, 879-885.
- *Newman, M. L. (2004). *Testosterone, status, and social stereotypes: Implications for cognitive performance*. (Unpublished doctoral dissertation). University of Texas-Austin, Austin, TX.
- Nguyen, H. D. (2006). *Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence*. (Unpublished doctoral dissertation). Michigan State University, East Lansing, MI.
- *Nguyen, H. H. D., O'Neal, A., & Ryan, A. M. (2003). Relating test-taking attitudes and skills and stereotype threat effects to the racial gap in cognitive ability test performance. *Human Performance*, *16*, 261-293.
- Nguyen, H.D. & Ryan, A.M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314-1334.
- Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H. M. (2005). Working memory and intelligence--their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 61-65
- *O'Brien, L. T., & Hummert, M. L. (2006). Memory performance of late middle-aged adults: contrasting self-stereotyping and stereotype threat accounts of assimilation to age stereotypes. *Social Cognition*, *24*, 338-358.

- *O'Mally, J. (2011). *Implicit measurement of challenge and threat as motivational responses to stereotype threat* (Unpublished doctoral dissertation). University of Alabama, Tuscaloosa, AL.
- O'Rourke, H. P., & MacKinnon, D. P. (2018). Reasons for testing mediation in the absence of an intervention effect: A research imperative in prevention and intervention research. *Journal of Studies on Alcohol and Drugs*, *79*, 171-181.
- *Oswald, D. L., & Harvey, R. D. (2000). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology*, *19*, 338-356.
- *Pearson, B. E. (2006). The effects of stereotype threat on the standardized test scores of Black college students. (Unpublished doctoral dissertation). University of Michigan, Ann Arbor, MI.
- *Prather, K. (2005) Controlling the threat of stereotypes. (Unpublished doctoral dissertation). George Washington University, Washington D.C.
- Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLOS one*, *11*, e0146487.
- Plous, S., & Williams, T. (1995). Racial stereotypes from the days of American slavery: A continuing legacy. *Journal of Applied Social Psychology*, *25*, 795-817.
- Ployhart, R. E., Ziegert, J. C., & McFarland, L. A. (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactions research. *Human Performance*, *16*, 231-259.
- Rendina-Gobioff, G., & Kromrey, J. D. (2006) PUB_BIAS: A SAS macro for detecting publication bias in meta-analysis. http://analytics.ncsu.edu/sesug/2006/PO04_06.PDF

- * Rivero-Conil, S. C. (2009). The effects of cognitive busyness on stereotype threat. (Unpublished doctoral dissertation). Carlos Albizu University, San Juan, Puerto Rico.
- Roberson, L., & Kulik, C. T. (2007). Stereotype threat at work. *The Academy of Management Perspectives, 21*, 24-40.
- Ryan, A. M. (2001). Explaining the black-white test score gap: The role of test perceptions. *Human Performance, 14*, 45-75.
- Ryan, A. M., & Nguyen, H. H. D. (2017). Publication bias and stereotype threat research: A reply to Zigerell. *Journal of Applied Psychology, 102*, 1169-1177.
- Ryan, A. M., & Sackett, P. R. (2013). Stereotype threat in workplace assessments. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.). *APA handbook of testing and assessment in psychology, vol. 1: Test theory and testing and assessment in industrial and organizational psychology*. (pp. 661-673). Washington, DC, US: American Psychological Association.
- *Rydell, R. J., & Boucher, K. L. (2009). Capitalizing on multiple social identities to prevent stereotype threat: The moderating role of self-esteem. *Personality and Social Psychology Bulletin, 36*, 239-250.
- *Rydell, R. J., Van Loo, K. J., & Boucher, K. L. (2014). Stereotype threat and executive functions: Which functions mediate different threat-related outcomes? *Personality and Social Psychology Bulletin, 40*, 377-390.
- Sackett, P. R. (2003). Stereotype threat in applied selection settings: A commentary. *Human Performance, 16*, 295-309.
- Sackett, P. R., & Ryan, A. M. (2012). Concerns about generalizing stereotype threat research to findings in operational high-stakes testing. In M. Inzlicht & T. Schmader (Eds.)

- Stereotype threat: Theory, process, and application* (pp. 249-263). New York: Oxford University Press.
- *Salinas, M. F. (1998). *Stereotype threat: The role of effort withdrawal and apprehension on the intellectual underperformance of Mexican Americans*. (Unpublished doctoral dissertation). University of Texas at Austin, Austin, TX.
- Sawyer, T. P., Jr., & Hollis-Sawyer, L. A. (2005). Predicting stereotype threat, test anxiety, and cognitive ability test performance: An examination of three models. *International Journal of Testing, 5*, 225-246.
- Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. *Journal of Applied Psychology*.
- *Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 440-452.
- *Schmader, T., Forbes, C. E., Zhang, S., & Mendes, W. B. (2009). A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin, 35*, 584-596.
- Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General, 136*, 241-255.
- Schimel, J., Arndt, J., Banko, K. M., & Cook, A. (2004). Not all self-affirmations were created equal: The cognitive and social benefits of affirming the intrinsic (vs. extrinsic) self. *Social Cognition, 22*, 75-99.

- *Schuster, C., Martiny, S. E., & Schmader, T. (2015). Distracted by the unthought-Suppression and reappraisal of mind wandering under stereotype threat. *PloS one*, *10*, 1-13.
- *Scott, G. L. (2012). *The role of media portrayals on the activation of stereotype threat: A study of African American test performance*. (Unpublished doctoral dissertation). City University of New York, New York City., NY.
- *Seibt B, & Forster J. (2004) Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology*, *87*, 38-56.
- *Seitchik, A. E., Jamieson, J., & Harkins, S. G. (2014). Reading between the lines: Subtle stereotype threat cues can motivate performance. *Social Influence*, *9*, 52-68.
- Sekaquaptewa, D., & Thompson, M. (2002). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, *39*, 68-74.
- Shapiro, J. R. (2011). Different groups, different threats: A multi-threat approach to the experience of stereotype threats. *Personality and Social Psychology Bulletin*, *37*, 464-480.
- Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, *11*, 107-130.
- Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of Personality and Social Psychology*, *104*, 277-288.

- Shelton, J. N., Richeson, J. A., Salvatore, J., & Trawalter, S. (2005). Ironic effects of racial bias during interracial interactions. *Psychological Science, 16*, 397-402.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.
- Shih, M., Pittinsky, T. L., & Ho, G. C. (2012). Stereotype boost: Positive outcomes from the activation of positive stereotypes. In M. Inzlicht & T. Schmader (Eds.) *Stereotype threat: Theory, Process, and Application* (pp. 141-156). New York: Oxford University Press.
- *Simmons, C. M. (2010). *Does manipulating stereotype threat condition change performance goal state* (Unpublished doctoral dissertation). University of Oklahoma, Norman, OK.
- *Smeding, A., Dumas, F., Loose, F., & Régner, I. (2013). Order of administration of math and verbal tests: An ecological intervention to reduce stereotype threat on girls' math performance. *Journal of Educational Psychology, 105*, 850-860.
- Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review, 16*, 177-206.
- Smith, J. L. (2006). The interplay among stereotypes, performance avoidance goals, and women's math performance expectancies. *Sex Roles, 54*, 287-296.
- Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology, 99*, 99-141.
- *Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles, 47*, 179-191.

- *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. L. (2005). *Stereotype threat and women's math performance: The possible mediating factors of test anxiety, test motivation and self-efficacy*. (Unpublished doctoral dissertation. Rutgers University, New Brunswick, NJ.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75*, 1191-1197.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist, 52*, 613-629.
- *Steele, C.M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.
- Steele, C. M., Spencer, C. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34*, 379-440.
- *Stein, R. (2001). *Negative age stereotypes and older adults' memory performance: An examination of age stereotype activation and underlying mechanisms*. (Unpublished doctoral dissertation). Georgia Institute of Technology, Atlanta, GA.
- 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*, 665-693.
- *Swift, H. J., Abrams, D., & Marques, S. (2013). Threat or boost? Social comparison affects older people's performance differently depending on task domain. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68*, 23-30.

- *Taillandier-Schmitt, A., Esnard, C., & Mokoukolo, R. (2012). Self-affirmation in occupational training: Effects on the math performance of French women nurses under stereotype threat. *Sex roles, 67*, 43-57.
- Tajfel, H. & Turner, J.C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of Intergroup Relations* (pp. 7-24). Chicago, IL: Nelson-Hall.
- *Thoman, D. B., White, P. H., Yamawaki, N., & Koishi, H. (2008). Variations of gender-math stereotype content affect women's vulnerability to stereotype threat. *Sex Roles, 58*, 702-712.
- Thomas, D. A., & Ely, R. J. (1996). Making differences matter. *Harvard Business Review, 74*, 79-90.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language, 28*, 127-154.
- Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33*, 67-81.
- von Der Embse, N., Barterian, J., & Segool, N. (2013). Test anxiety interventions for children and adolescents: A systematic review of treatment studies from 2000-2010. *Psychology in the Schools, 50*, 57-71.
- von der Embse, N., Jester, D., Roy, D., & Post, J. (2017). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders, 227*, 483-493.
- von Hippel, W., von Hippel, C., Conway, L., Preacher, K. J., Schooler, J. W., & Radvansky, G. A. (2005). Coping with stereotype threat: Denial as an impression management strategy. *Journal of Personality and Social Psychology, 89*, 22-35.

- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456-467.
- Walton, G. M. & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*, 1447-1451.
- Walton, G. M., Murphy, M. C., & Ryan, A. M. (2015). Stereotype threat in organizations: Implications for equity and performance. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 523-550.
- *Wasserberg, M. J. (2009). *Stereotype threat and the standardized testing experiences of African American children at an urban elementary school*. (Unpublished doctoral dissertation). Florida International University, Miami, FL.
- Wegner, D. M., Erber, R., & Zanakos, S. (1993). Ironic processes in the mental control of mood and mood-related thought. *Journal of Personality and Social Psychology, 65*, 1093-1104.
- Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology, 75*, 315-321.
- *Wong, J. T. (2014). Ageist stereotype threat effects on memory and metacognition in older adults. (Unpublished doctoral dissertation). University of Chicago, Chicago, IL.
- *Woodcock, A. (2004). The role of self-efficacy in stereotype threat. (Unpublished doctoral dissertation). California State University - San Marcos, San Marcos, CA.

APPENDIX A

Stereotype Threat Mediators Coding Sheet

Cell	Label	Definition	Values
A	CODED		
B	Coding Number		
C	Meta-Analysis Article Number	Article number in the present study	
D	Meta-Analysis Sample Number	Sample number in the present study	
E	Full Reference	APA reference	
F	Study Year		
G	Study Sample Number	The sample number (i.e., if there is more than one sample for a study, list the number here) and time point (if relevant)	
H	Mediator	List the type of mediator examined	1 = anxiety 2 = physiological arousal 3 = negative thoughts 4 = working memory 5 = effort 6 = prevention/avoidance focus 7 = performance expectations
I	Mediator Measurement Description	Describe how mediator was measured (e.g., with a questionnaire, measured implicitly)	
J	Mediator Scale Description		
K	Mediator Measurement Method Code	Describe how mediator was measured	1 = indirect/implicit (physiological- blood pressure, skin conductance, facial expression, stimuli responsiveness) 2 = self-report 3 = task

Cell	Label	Definition	Values
			4 = other (describe)
L	Mediator Measurement Timing	Describe when mediator when measured	1 = before performance 2 = after performance 3 = concurrent 4 = no performance
M	Mediator Mean	Mean for mediator across groups	
N	Mediator SD	Standard deviation for mediator across groups	
O	Group 1 <i>n</i>	Sample size of group 1 (stereotype threat group)	
P	Group 1 Mediator Mean	Mediator mean for stereotype threatened group	
Q	Group 1 SD	Standard deviation for stereotype threatened group	
R	Group 2 <i>n</i>	Sample size of group 2 (non-stereotype threat group)	
S	Group 2 Mediator Mean	Mediator mean for non-stereotype threatened group	
T	Group 2 SD	Standard deviation for non-stereotype threatened group	
U	Other statistic if effect sizes are not provided	t, F, value	
V	--	Description of previous value	
W	<i>d</i>	Effect Size	
X	Description of effect size		
Y	Mediator Scale Reliability	Mediator scale reliability	
Z	Mediator Type of Reliability	List the type of reliability	

Cell	Label	Definition	Values
AA	Stereotype activation	Describe how stereotype threat was activated (e.g., asked race before test, interacted with only males)	
AB	Stereotype activation code	Level of Stereotype Activation	1= blatant 2 = moderate 3 = subtle
AC	Task Description	Description of task (e.g., Stroop, verbal ability test, mathematics test, golf exercise)	
AD	Target of Stereotype	Description of the stereotype targeted	1 = gender-based 2 = race-based 3 = age-based 4 = Other (describe)
AE	Negatively Stereotype Group	Negatively Stereotype Group	
AF	Control Group	Control Group	
AG	Test Setting	Individual vs. Group Setting	1 = Individual 2 = Group 3 = not listed 4 = Both
AH	Test Purpose	Laboratory vs. Operational (Decision making) setting	1 = Lab 2 = Operational
AI	Test difficulty	Level of test difficulty	1= very difficult 2 = moderately difficult 3 = easy
AJ	Design	List whether study is cross-sectional, longitudinal, experience-sampling, etc.	
AK	Design Group	Description if study is within-subgroup or between-groups	1= within-subgroup 2 = between-groups
AL	Country	Country where study conducted	
AM	N *use the smallest numbers.	Sample size	

Cell	Label	Definition	Values
AN	Sample description	Description of the sample population	
AO	Selection Criteria - Identification Level	Selected for level of identification (Domain or group)	1 = domain ID 2 = group ID 3 = ability level 4 = both ID levels 5 = ability and domain ID level 6 = ability and group ID level
AP	stereotype endorsement/ knowledge for selection		0 = no 1 = stereotype endorsement 2 = stereotype knowledge 3 = both
AQ	n-male	Number of males	
AR	n-female	Number of females	
AS	n-White	Number of whites	
AT	n-minorities	Number of minorities	
AU	Mean Age		
AV	Academic Field - journal title	Field where study published	
AW	Published Study	Is the study published?	
AX	Amber's Notes	Add any comments/notes	
AY	Coders Notes	Add any comments/notes	

APPENDIX B

Table of Effect Sizes for Each Mediator

Study	Mediator	<i>d</i>	Group 1 <i>n</i>	Group 2 <i>n</i>	Design Group
Abrams, Crisp, Marques, Fagg, Bedford, & Provias, (2008) Study 1a	Anxiety	0.66	49	48	Within
Abrams, Crisp, Marques, Fagg, Bedford, & Provias, (2008) Study 1b	Anxiety	0.37	26	25	Within
Abrams, Eller, & Bryant (2006)	Effort	0.63	49	48	Within
Allen (2012)	Physiological Arousal	-0.13	44	41	Within
Allen (2012)	Working Memory	0.01	44	41	Within
Allen (2012)	Distracting (Negative) Thoughts	0.52	44	41	Within
Anyeji (2013) Study 1	Anxiety	-0.04	32	32	Within
Beaton, Tougas, Rinfret, Huard, & Delisle (2007) Study 2	Performance Expectations	-0.06	22	23	Within
Beaton, Tougas, Rinfret, Huard, & Delisle (2007) Study 2	Anxiety	0.80	22	23	Within
Beilock, Rydell, & McConnell (2007) Study 1	Working Memory	-0.61	14	14	Within
Beilock, Rydell, & McConnell (2007) Study 3	Distracting (Negative) Thoughts	-0.57	33	42	Within
Beilock, Rydell, & McConnell (2007) Study 3	Working Memory	-0.64	15	15	Within
Beilock, Rydell, & McConnell (2007) Study 4	Working Memory	-0.70	30	30	Within
Berjot, Roland-Levy, & Girault-Lidvan (2011)	Distracting (Negative) Thoughts	2.09	22	22	Within
Berjot, Roland-Levy, & Girault-Lidvan (2011)	Effort	-0.63	22	22	Within
Berjot, Roland-Levy, & Girault-Lidvan (2011)	Distracting (Negative) Thoughts	1.67	22	23	Between
Berjot, Roland-Levy, & Girault-Lidvan (2011)	Effort	-1.34	22	23	Between
Bouazzaoui, Follenfant, Ric, Fay, Croizet, Atzeni, & Taconnat (2015)	Performance Expectations	-0.46	46	46	Within
Boucher, Rydell, & Murphy (2015)	Performance Expectations	-1.23	30	30	Between
Boucher, Rydell, & Murphy (2015)	Effort	-1.10	30	30	Between
Boucher, Rydell, & Murphy (2015) Study 1	Anxiety	0.76	60	60	Between
Brodish (2008)	Physiological Arousal	-0.41	47	47	Within
Brodish (2008)	Working Memory	-0.40	47	47	Within
Brodish (2008) Study 1	Performance Avoidance	-0.27	47	47	Within
Brodish (2008) Study 2	Anxiety	0.42	47	47	Within
Brodish (2008) Study 2	Performance Avoidance	0.40	47	47	Within
Brown & Day (2006)	Effort	-0.05	19	17	Within
Brown & Day (2006)	Effort	0.05	19	29	Between

Study	Mediator	<i>d</i>	Group 1 <i>n</i>	Group 2 <i>n</i>	Design Group
Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti (2003) Study 1	Performance Expectations	-0.02	32	32	Within
Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti (2003) Study 2	Performance Expectations	-1.46	25	25	Within
Cadinu, Maass, Rosabianca, & Kiesner (2005)	Distracting (Negative) Thoughts	0.67	30	30	Within
Carr & Steele (2009)	Distracting (Negative) Thoughts	0.70	14	14	Within
Carr & Steele (2009)	Distracting (Negative) Thoughts	0.77	14	20	Between
Carr & Steele (2009) Study 1	Effort	1.40	14	14	Within
Carr & Steele (2009) Study 1	Effort	0.84	14	20	Between
Carr & Steele (2009) Study 2	Effort	0.91	13	18	Within
Carr & Steele (2009) Study 2	Effort	0.90	13	17	Between
Chasteen (2005)	Distracting (Negative) Thoughts	0.02	19	19	Within
Conway-Klaassen (2010)	Performance Avoidance	0.66	20	20	Within
Cotting (2003) Study 1	Anxiety	0.29	52	58	Within
Cotting (2003) Study 2	Anxiety	0.01	40	62	Within
Delgado & Prieto (2008) Study 1	Anxiety	0.67	89	81	Between
Delgado & Prieto (2008) Study 1	Anxiety	0.01	89	79	Within
Desrichard & Köpetz (2005)	Performance Expectations	-1.26	30	30	Within
Desrichard & Köpetz (2005)	Performance Expectations	-2.40	30	30	Between
Dinella (2004) Study 1	Anxiety	0.26	62	49	Between
Dinella (2004) Study 1	Anxiety	-0.28	62	60	Within
Fernandez (2007)	Distracting (Negative) Thoughts	-0.01	32	31	Within
Fernandez (2007)	Performance Expectations	0.05	32	31	Within
Fernandez (2007) Study 1	Anxiety	0.16	32	31	Within
Foote (2000) Study 1	Distracting (Negative) Thoughts	0.08	40	40	Within
Foote (2000) Study 1	Effort	0.36	40	40	Within
Foote (2000) Study 1	Anxiety	-0.02	40	40	Within
Foote (2000) Study 1	Distracting (Negative) Thoughts	0.44	40	39	Between
Foote (2000) Study 1	Effort	0.53	40	39	Between
Foote (2000) Study 1	Anxiety	0.29	40	39	Between
Forbes, Leitner, Duran-Jordan, Magerman, Schmader, & Allen (2014)	Distracting (Negative) Thoughts	0.32	33	25	Between
Forbes, Leitner, Duran-Jordan, Magerman, Schmader, & Allen (2014)	Performance Expectations	0.06	33	25	Between
Good, Woodzicka, & Wingfield (2010) Study 1	Anxiety	1.13	18	7	Between

Study	Mediator	<i>d</i>	Group 1 <i>n</i>	Group 2 <i>n</i>	Design Group
Good, Woodzicka, & Wingfield (2010) Study 1a	Anxiety	0.26	18	15	Within
Good, Woodzicka, & Wingfield (2010) Study 1b	Anxiety	0.53	18	16	Within
Grand, Ryan, Schmitt, & Hmurovic (2011)	Effort	-0.14	120	113	Within
Grand, Ryan, Schmitt, & Hmurovic (2011)	Effort	-0.17	120	52	Between
Hall (2012)	Performance Expectations	-0.19	26	26	Within
Hall (2012) Study 1a	Anxiety	-0.20	26	26	Within
Hall (2012) Study 1b	Anxiety	0.34	26	26	Within
Hardee Bailey (2004)	Distracting (Negative) Thoughts	0.04	19	22	Within
Hardee Bailey (2004)	Performance Expectations	-0.69	19	22	Within
Hardee Bailey (2004)	Distracting (Negative) Thoughts	0.11	19	30	Between
Hardee Bailey (2004)	Performance Expectations	-0.99	19	30	Between
Harder (2000) Study 1	Anxiety	0.96	9	10	Within
Hess, Hinson, & Hodges (2009)	Physiological Arousal	0.54	52	51	Between
Howard (2005)	Working Memory	-0.21	36	33	Within
Howard (2005)	Performance Expectations	-0.26	37	35	Within
Inzlicht & Ben-Zeev (2003) Study 1a	Anxiety	-0.34	11	12	Within
Inzlicht & Ben-Zeev (2003) Study 1b	Anxiety	0.02	13	15	Within
Inzlicht & Ben-Zeev (2003) Study 1c	Anxiety	0.05	12	12	Within
Inzlicht & Ben-Zeev (2003) Study 1d	Anxiety	0.27	14	15	Within
Inzlicht & Kang (2010)	Working Memory	0.59	15	13	Between
Jamieson & Harkins (2007) Study 1	Performance Expectations	-0.16	20	20	Within
Jamieson & Harkins (2007) Study 1	Effort	-0.31	20	20	Within
Jamieson & Harkins (2007) Study 1	Performance Expectations	-0.11	20	20	Between
Jamieson & Harkins (2007) Study 1	Effort	0.14	20	20	Between
Jamieson & Harkins (2007) Study 1	Anxiety	-0.04	20	20	Within
Jamieson & Harkins (2007) Study 1	Working Memory	0.98	20	20	Within
Jamieson & Harkins (2007) Study 1	Anxiety	-0.16	20	20	Between
Jamieson & Harkins (2007) Study 1 [BT]	Working Memory	1.28	20	20	Between
Jamieson & Harkins (2007) Study 2	Anxiety	-0.15	18	18	Within
Jamieson & Harkins (2007) Study 2	Performance Expectations	-0.03	18	18	Within
Jamieson & Harkins (2007) Study 2	Effort	0.16	18	18	Within
Jamieson & Harkins (2007) Study 2	Working Memory	0.29	18	18	Within
Jamieson & Harkins (2007) Study 3	Anxiety	0.21	18	18	Within
Jamieson & Harkins (2007) Study 3	Performance Expectations	0.45	18	18	Within

Study	Mediator	<i>d</i>	Group 1 <i>n</i>	Group 2 <i>n</i>	Design Group
Jamieson & Harkins (2007) Study 3	Effort	0.15	18	18	Within
Jamieson & Harkins (2007) Study 3	Working Memory	0.62	18	18	Within
Jamieson & Harkins (2009) Study 1	Effort	0.58	16	16	Within
Jamieson & Harkins (2009) Study 1	Effort	-0.40	16	16	Between
Jamieson & Harkins (2009) Study 1	Anxiety	-0.02	16	16	Within
Jamieson & Harkins (2009) Study 1	Anxiety	0.19	16	16	Between
Jamieson & Harkins (2009) Study 2	Effort	0.19	32	32	Within
Jamieson & Harkins (2011)	Working Memory	-1.13	32	32	Within
Jamieson & Harkins (2011)	Effort	-0.59	32	32	Within
John-Henderson, Rheinschmidt, & Mendoza-Denton (2015)	Physiological Arousal	1.59	25	24	Within
Johns, Inzlicht, & Schmader (2008)	Working Memory	-0.69	40	41	Within
Kang & Chasteen (2009) Study 1	Distracting (Negative) Thoughts	-0.23	20	22	Within
Kang & Chasteen (2009) Study 1	Anxiety	0.19	20	22	Within
Keller & Bless (2008)	Performance Avoidance	0.12	30	30	Within
Keller & Dauenheimer (2003)	Performance Avoidance	0.00	19	16	Within
Keller & Dauenheimer (2003)	Anxiety	0.34	19	16	Within
Keller & Dauenheimer (2003)	Effort	0.51	19	16	Within
Keller (2002)	Performance Expectations	-0.92	16	16	Between
Keller (2002)	Performance Expectations	-0.08	16	21	Within
Keller (2002)	Effort	0.26	16	21	Within
Keller (2002)	Effort	0.07	16	16	Between
Krotenberg (2003)	Distracting (Negative) Thoughts	0.22	42	40	Within
Krotenberg (2003) Study 1a	Anxiety	0.40	25	23	Within
Krotenberg (2003) Study 1b	Anxiety	-0.12	17	17	Within
Lu, Feng, Tian, Yu, Hong, & Zheng (2015)	Distracting (Negative) Thoughts	0.63	38	38	Within
Lu, Feng, Tian, Yu, Hong, & Zheng (2015)	Anxiety	0.54	38	38	Within
Mangels, Good, Whiteman, Maniscalco, & Dweck (2012) Study 1	Anxiety	0.93	16	18	Within
Marx, Stapel, & Muller (2005) Study 1	Performance Expectations	-1.22	16	16	Within
Marx, Stapel, & Muller (2005) Study 1	Distracting (Negative) Thoughts	0.71	16	16	Within
Marx, Stapel, & Muller (2005) Study 2	Distracting (Negative) Thoughts	1.40	39	38	Within
Mazerolle, Régner, Morisset, Rigalleau, & Huguet (2012) Study 1	Working Memory	-0.68	54	55	Between
Mazerolle, Régner, Morisset, Rigalleau, & Huguet (2012) Study 1	Working Memory	-0.54	54	56	Within

Study	Mediator	<i>d</i>	Group 1 <i>n</i>	Group 2 <i>n</i>	Design Group
McIntyre, Lord, Gresky, Ten Eyck, Frye, & Bond Jr (2005)	Performance Expectations	-0.24	44	17	Between
McKay (1999)	Performance Expectations	-0.10	51	52	Within
McKay (1999)	Anxiety	-0.14	51	52	Within
Mrazek, Chin, Schmader, Hartson, Smallwood, & Schooler (2011)	Distracting (Negative) Thoughts	0.64	21	22	Within
Newman (2004)	Performance Expectations	-1.13	26	27	Between
Nguyen, O'Neal, & Ryan (2003) Study 1	Anxiety	0.51	43	43	Within
O'Brien & Hummert (2006) Study 1	Anxiety	-0.17	29	28	Within
O'Malley (2012) Study 1	Performance Avoidance	0.15	15	14	Within
O'Malley (2012) Study 1	Performance Avoidance	0.26	15	6	Between
O'Malley (2012) Study 2	Performance Avoidance	-0.24	50	49	Within
O'Malley (2012) Study 2	Performance Avoidance	-0.16	50	29	Between
Oswald & Harvey (2000)	Distracting (Negative) Thoughts	0.78	22	23	Within
Oswald & Harvey (2000)	Effort	0.42	18	18	Within
Pearson (2006)	Distracting (Negative) Thoughts	0.02	40	40	Within
Prather (2005)	Effort	0.08	55	59	Within
Prather (2005) Study 1	Anxiety	-0.23	55	59	Within
Rivero-Conil (2009) Study 1	Anxiety	-0.07	37	41	Within
Rydell & Boucher (2009)	Working Memory	-0.43	70	70	Within
Rydell, Van Loo, & Boucher (2013)	Distracting (Negative) Thoughts	-0.01	45	45	Within
Rydell, Van Loo, Loo, & Boucher (2013)	Distracting (Negative) Thoughts	-0.40	40	41	Within
Salinas (1998)	Performance Expectations	-0.27	18	9	Within
Salinas (1998)	Effort	0.24	18	9	Within
Salinas (1998)	Anxiety	0.29	18	9	Within
Salinas (1998)	Performance Expectations	-0.17	18	9	Between
Salinas (1998)	Effort	1.02	18	9	Between
Salinas (1998)	Anxiety	0.26	18	9	Between
Schmader & Johns (2003) Study 1	Anxiety	0.52	14	17	Between
Schmader & Johns (2003) Study 1	Working Memory	-1.19	14	17	Between
Schmader & Johns (2003) Study 1	Anxiety	0.02	14	14	Within
Schmader & Johns (2003) Study 1	Working Memory	-1.66	14	14	Within
Schmader & Johns (2003) Study 2	Working Memory	-0.55	18	18	Within
Schmader & Johns (2003) Study 2	Anxiety	0.70	16	17	Within
Schmader & Johns (2003) Study 2	Anxiety	0.54	16	20	Between

Study	Mediator	<i>d</i>	Group 1 <i>n</i>	Group 2 <i>n</i>	Design Group
Schmader & Johns (2003) Study 2	Working Memory	-0.66	18	18	Between
Schmader & Johns (2003) Study 3	Anxiety	0.38	13	15	Within
Schmader & Johns (2003) Study 3	Working Memory	-1.19	14	14	Within
Schmader, Forbes, Zhang, & Mendes (2009) Study 1	Anxiety	0.39	37	40	Between
Schmader, Forbes, Zhang, & Mendes (2009) Study 3	Anxiety	-0.06	61	50	Within
Schuster, Martiny, & Schmader (2015)	Working Memory	0.04	19	19	Within
Scott (2012) Study 1	Anxiety	-0.19	30	30	Within
Seibt & Förster (2004)	Effort	0.95	14	15	Within
Seibt & Förster (2004)	Effort	0.26	14	15	Between
Seitchik, Jamieson, & Harkins (2014) Study 1	Effort	0.42	18	18	Within
Seitchik, Jamieson, & Harkins (2014) Study 1	Effort	0.19	18	18	Between
Simmons (2010)	Performance Expectations	-0.05	48	48	Within
Smeding, Dumas, Loose, & Régner (2013)	Performance Expectations	-0.19	131	136	Within
Smeding, Dumas, Loose, & Régner (2013)	Performance Expectations	-0.32	131	129	Between
Smith & White (2002)	Performance Expectations	-0.35	17	20	Within
Smith & White (2002)	Effort	0.31	17	19	Within
Spencer (2005)	Performance Expectations	-0.36	20	20	Within
Spencer (2005)	Effort	-0.31	20	20	Within
Spencer (2005)	Anxiety	-0.01	20	20	Within
Spencer, Steele, & Quinn (1999) Study 3	Distracting (Negative) Thoughts	0.34	18	18	Within
Spencer, Steele, & Quinn (1999) Study 3	Performance Expectations	-0.80	18	18	Within
Spencer, Steele, & Quinn (1999) Study 3	Anxiety	0.57	18	18	Within
Steele & Aronson (1995) Study 2	Distracting (Negative) Thoughts	1.66	10	10	Within
Steele & Aronson (1995) Study 2	Effort	1.18	10	10	Within
Steele & Aronson (1995) Study 2	Effort	1.60	10	10	Between
Steele & Aronson (1995) Study 4	Effort	1.71	11	11	Between
Stein (2001)	Distracting (Negative) Thoughts	0.22	30	36	Within
Stein (2001)	Performance Expectations	0.06	30	34	Within
Stein (2001)	Effort	-0.42	30	35	Within
Stein (2001)	Anxiety	0.13	30	35	Within
Stein (2001)	Performance Expectations	-0.60	30	36	Between
Stein (2001)	Effort	0.03	30	36	Between
Stein (2001)	Anxiety	0.51	30	36	Between

Study	Mediator	<i>d</i>	Group 1 <i>n</i>	Group 2 <i>n</i>	Design Group
Stein (2001)	Distracting (Negative) Thoughts	-0.13	30	36	Between
Swift, Abrams, & Marques (2013) Study 2	Anxiety	0.72	59	59	Within
Taillandier-Schmitt, Esnard, & Mokoukolo (2012)	Performance Expectations	1.47	25	15	Within
Taillandier-Schmitt, Esnard, & Mokoukolo (2012)	Anxiety	-0.11	25	15	Within
Thoman, White, Yamawaki, & Koishi (2008)	Performance Expectations	0.54	19	25	Within
Thoman, White, Yamawaki, & Koishi (2008)	Effort	0.25	19	25	Within
Thoman, White, Yamawaki, & Koishi (2008)	Anxiety	0.03	19	25	Within
Wasserberg (2009) Study 2	Performance Avoidance	-0.19	101	97	Within
Wasserberg (2009) Study 2	Performance Expectations	0.28	101	97	Within
Wasserberg (2009) Study 2	Anxiety	0.44	101	97	Within
Wong (2014) Study 1	Performance Expectations	-0.07	42	42	Within
Wong (2014) Study 1	Performance Expectations	-0.20	42	42	Between
Wong (2014) Study 1	Working Memory	-1.95	42	42	Between
Wong (2014) Study 1	Working Memory	-0.28	42	42	Within
Wong (2014) Study 2	Performance Expectations	-0.41	42	42	Within
Wong (2014) Study 2	Working Memory	-0.41	42	42	Within
Woodcock (2014)	Effort	-0.07	64	65	Within

APPENDIX C

Study Effect Sizes Forest Plots by Analysis Level

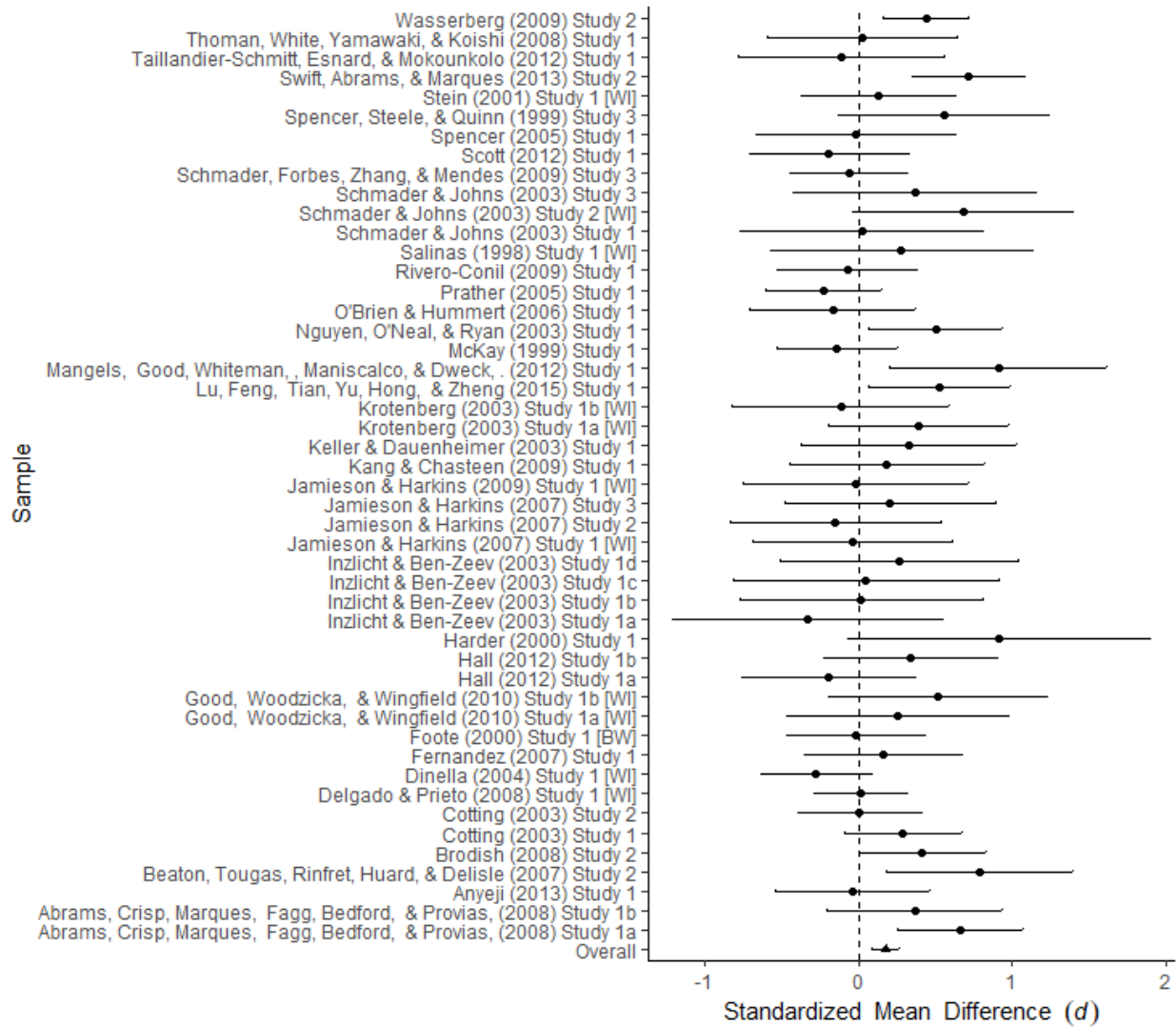


Figure C 1. Forest plot for anxiety within-group effect sizes

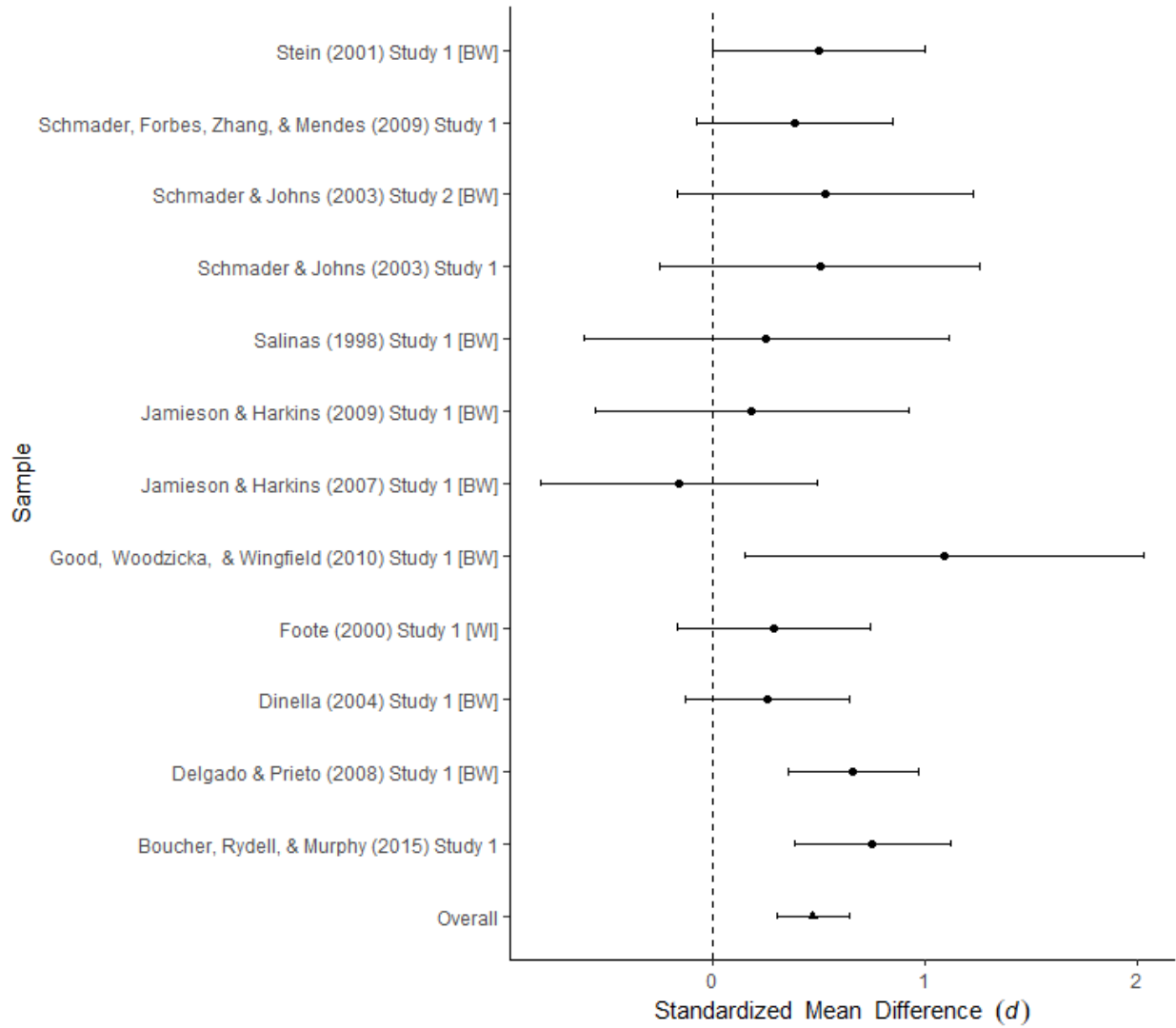


Figure C2. Forest plot for anxiety between-group results

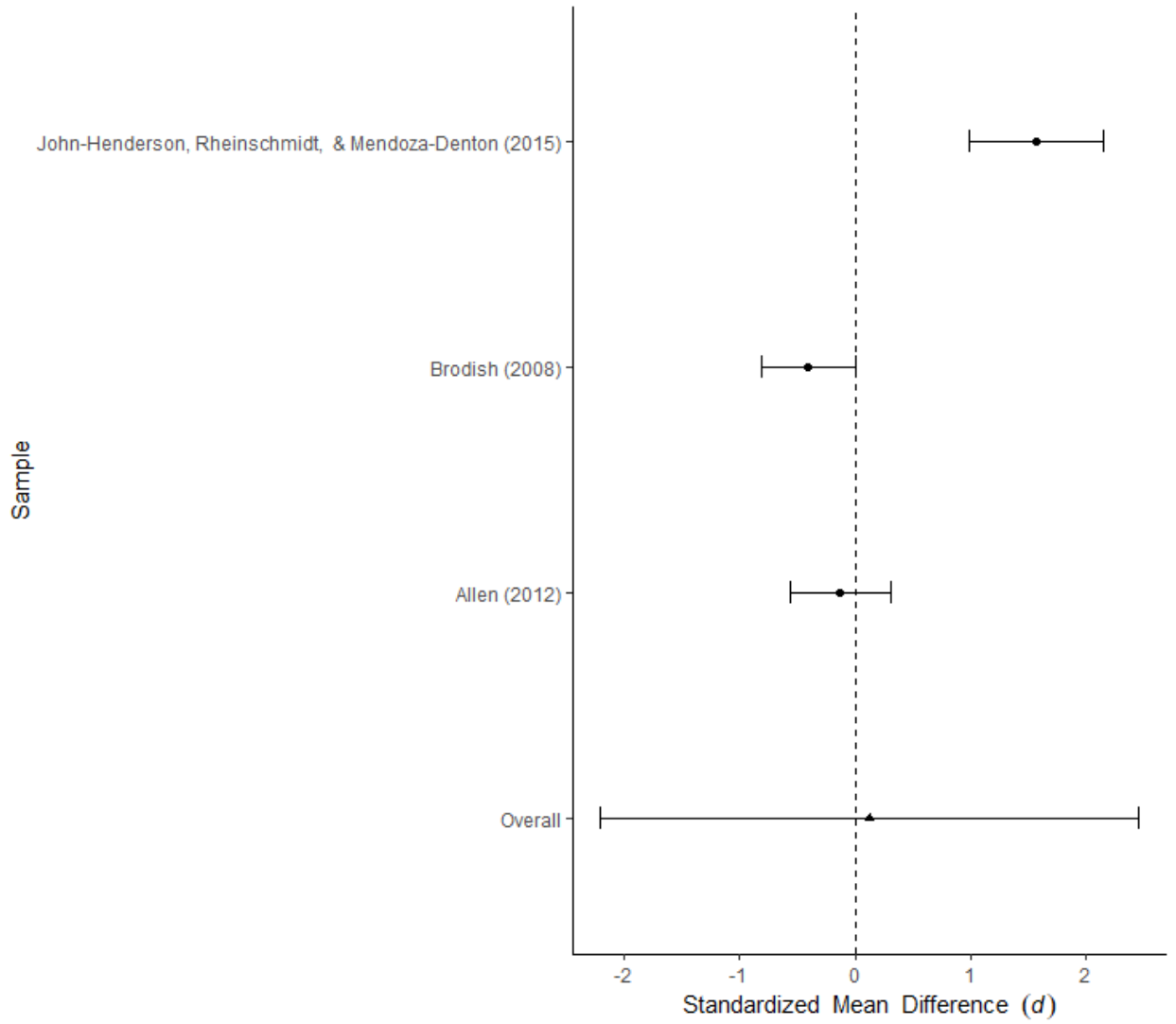


Figure C3. Forest plot for physiological arousal within-group effect sizes

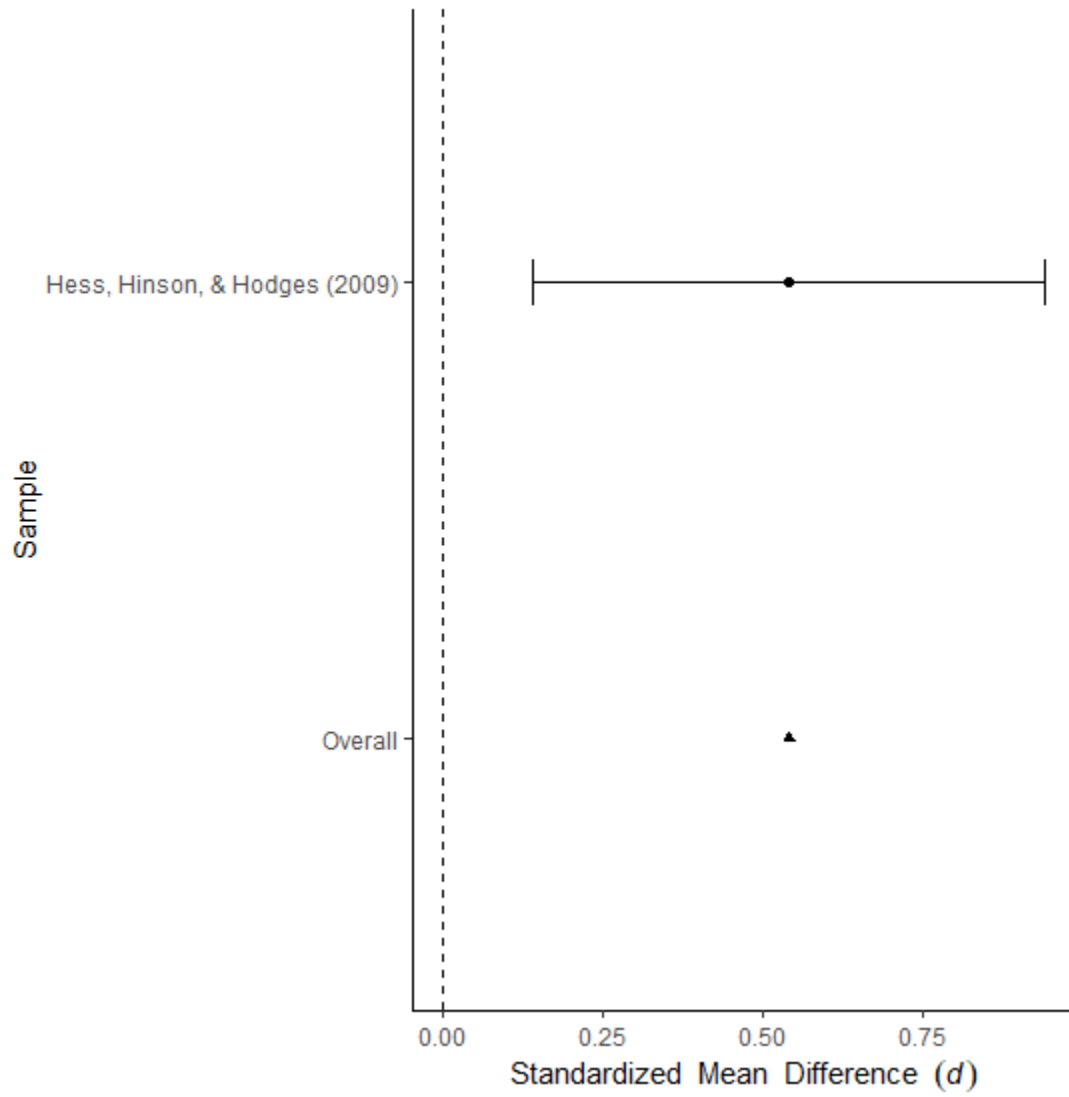


Figure C4 . Forest plot for physiological arousal between-group effect sizes

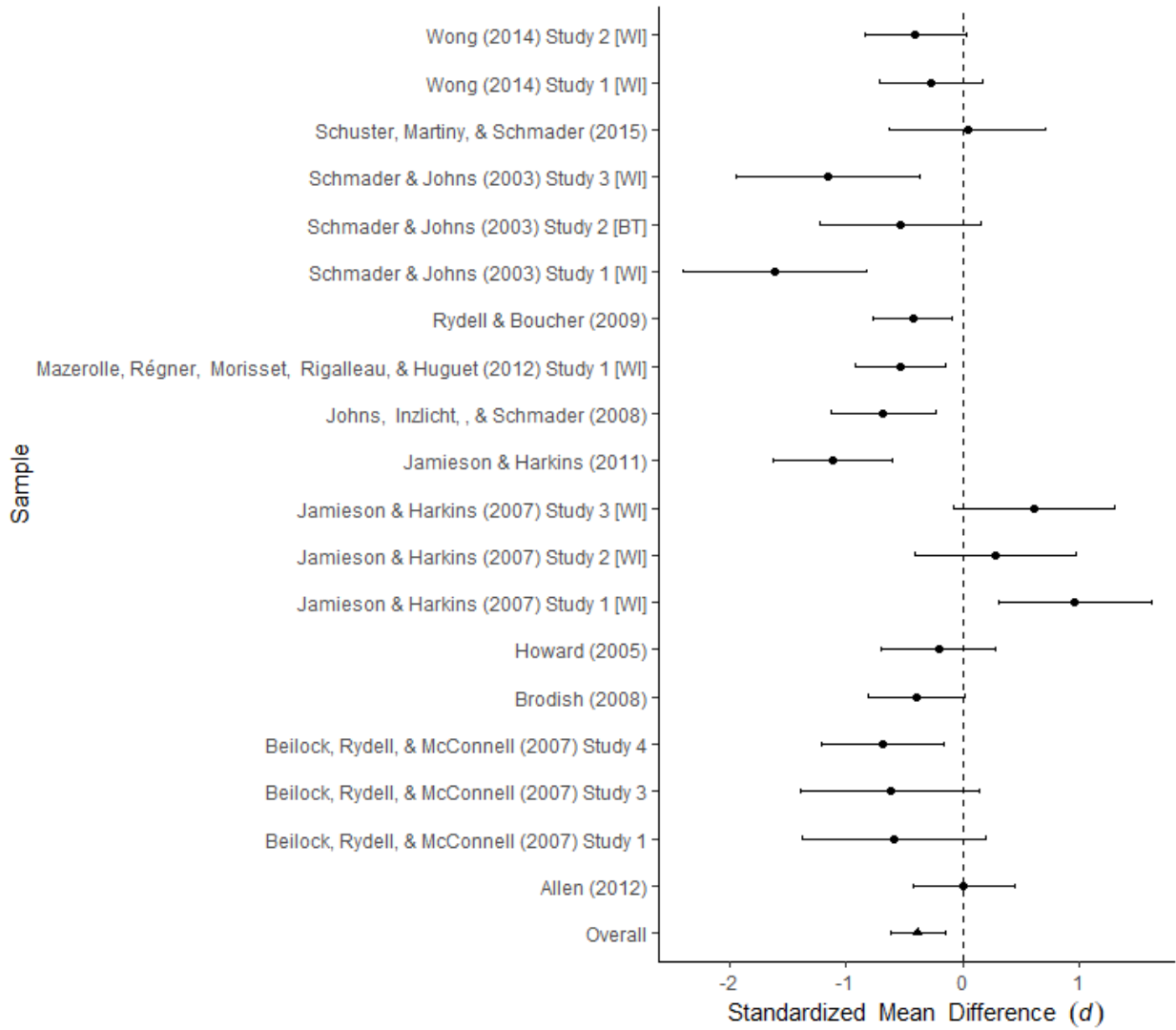


Figure C5. Forest plot for working memory within-group effect sizes

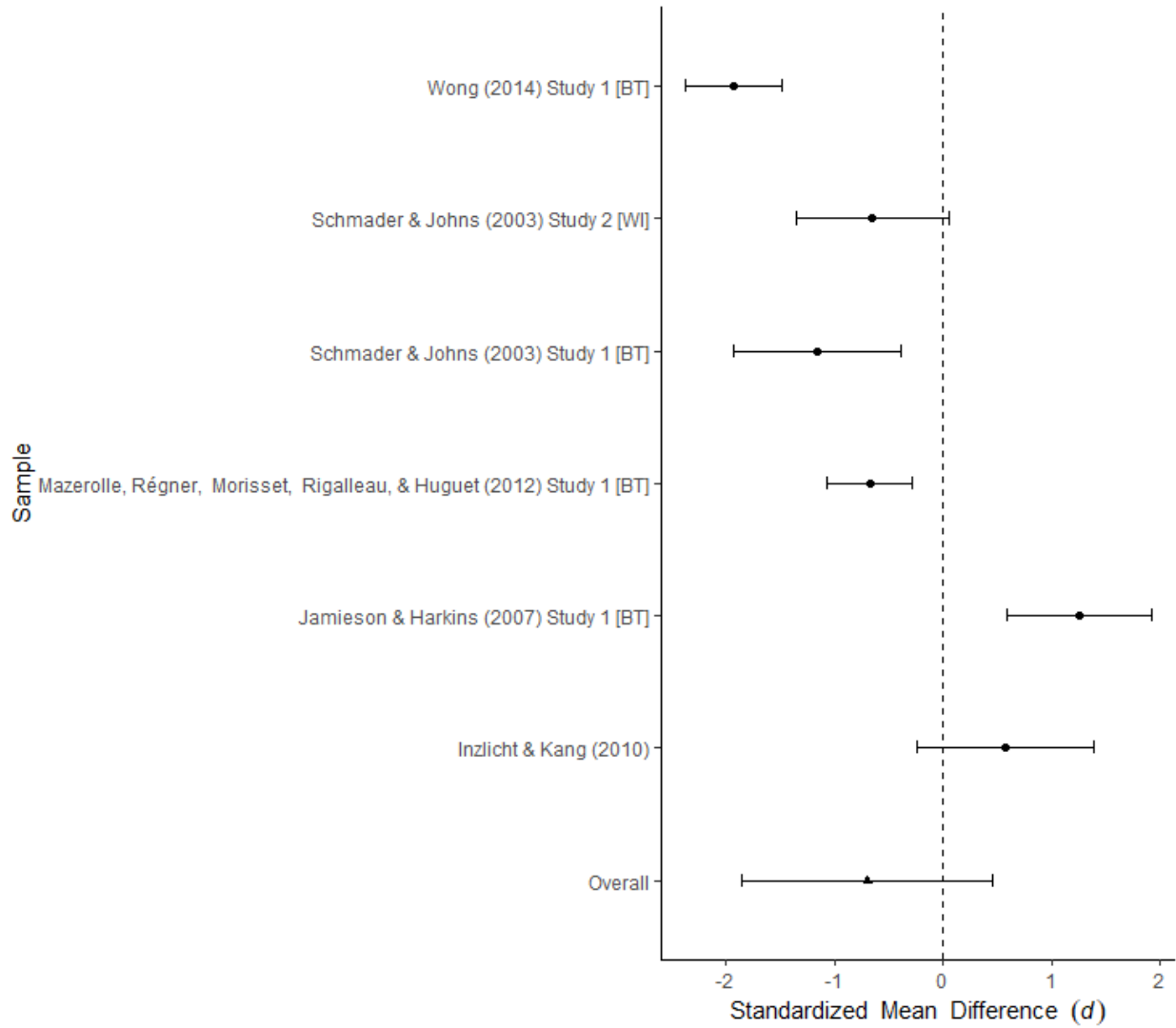


Figure C6. Forest plot for working memory between-group effect sizes

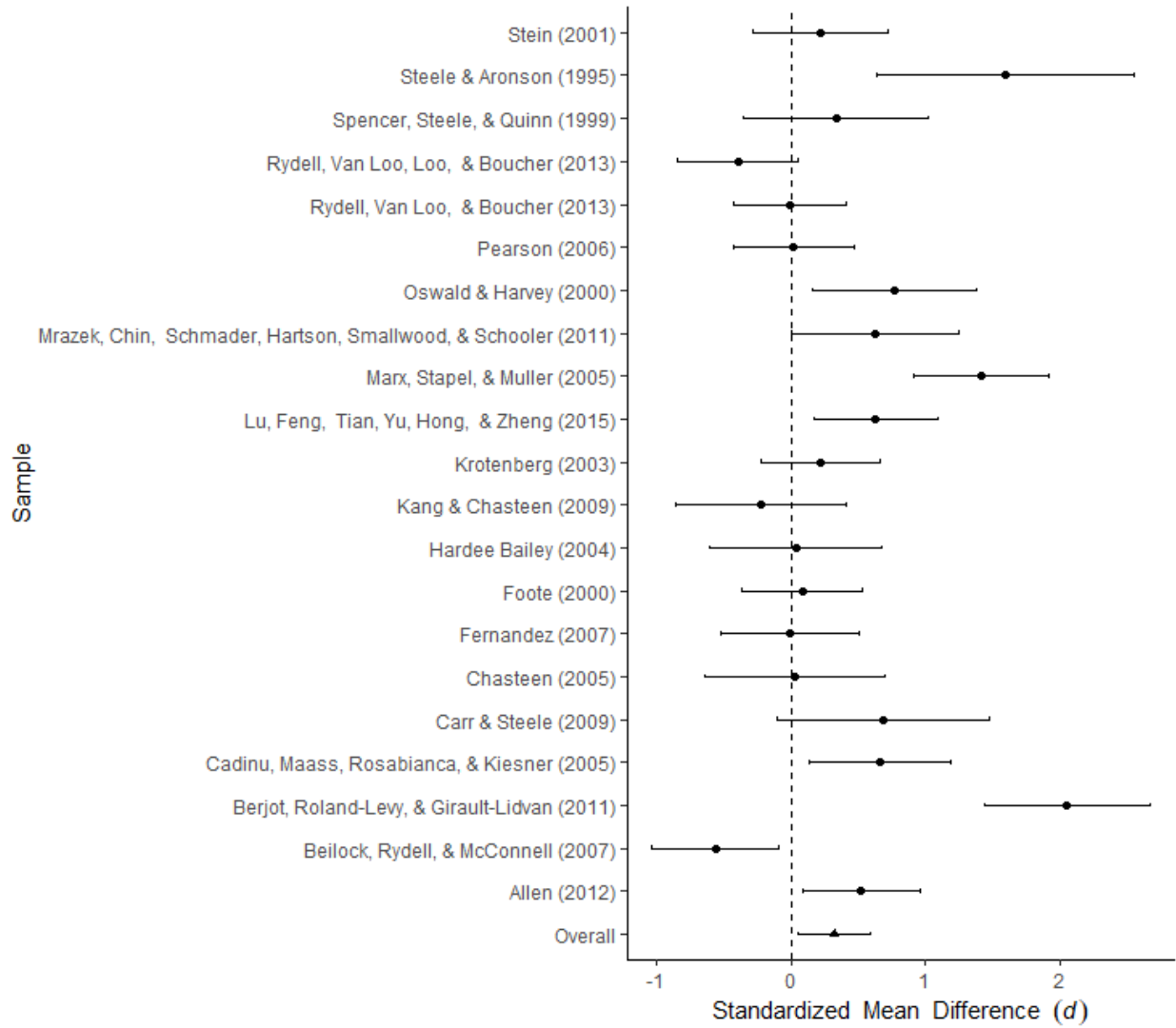


Figure C7. Forest plot for distracting (negative) thoughts within-group effect sizes

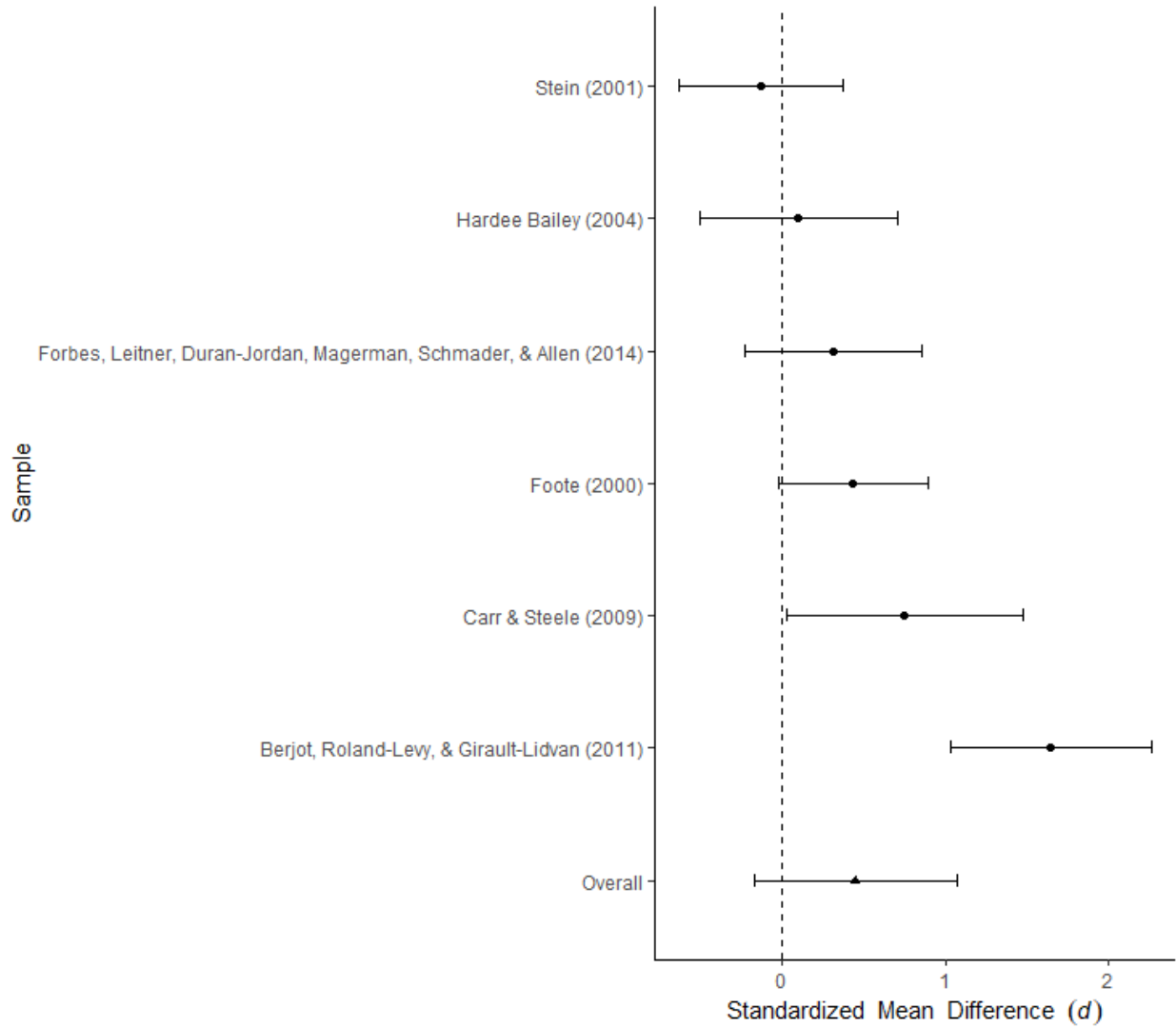


Figure C8. Forest plot for distracting (negative) thoughts between-group effect sizes

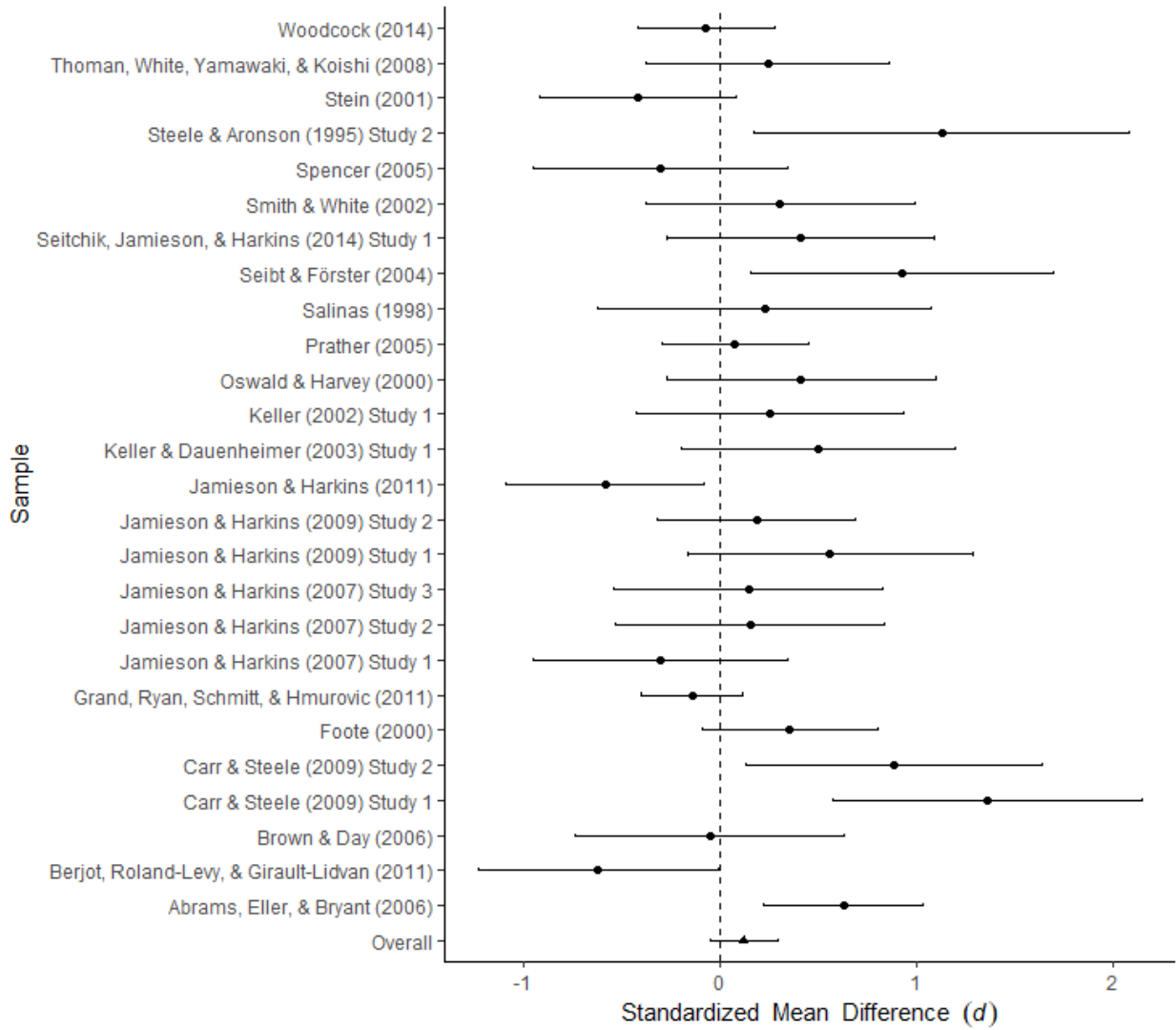


Figure C9. Forest plot for effort within-group effect sizes

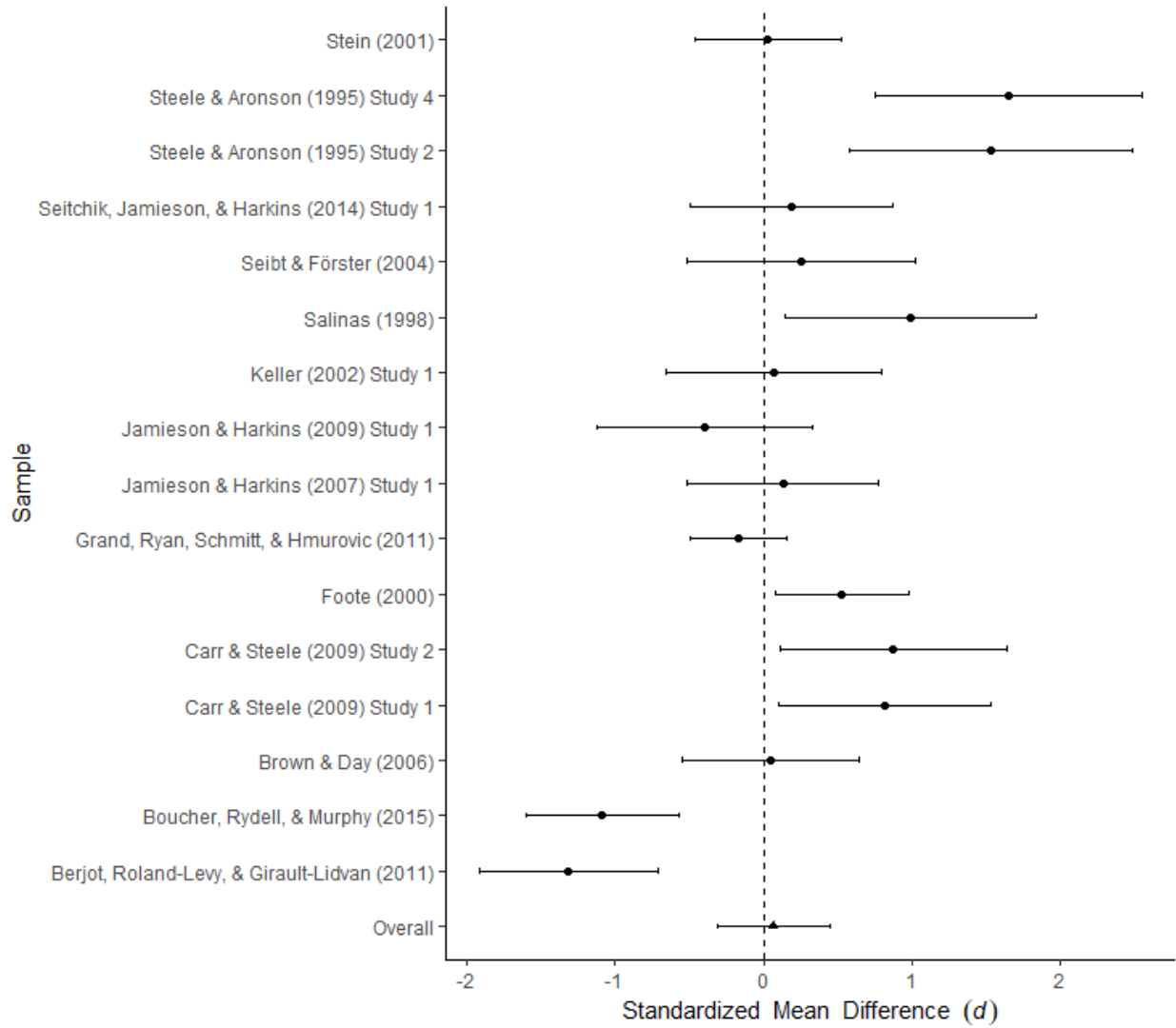


Figure C10. Forest plot for effort between-group effect sizes

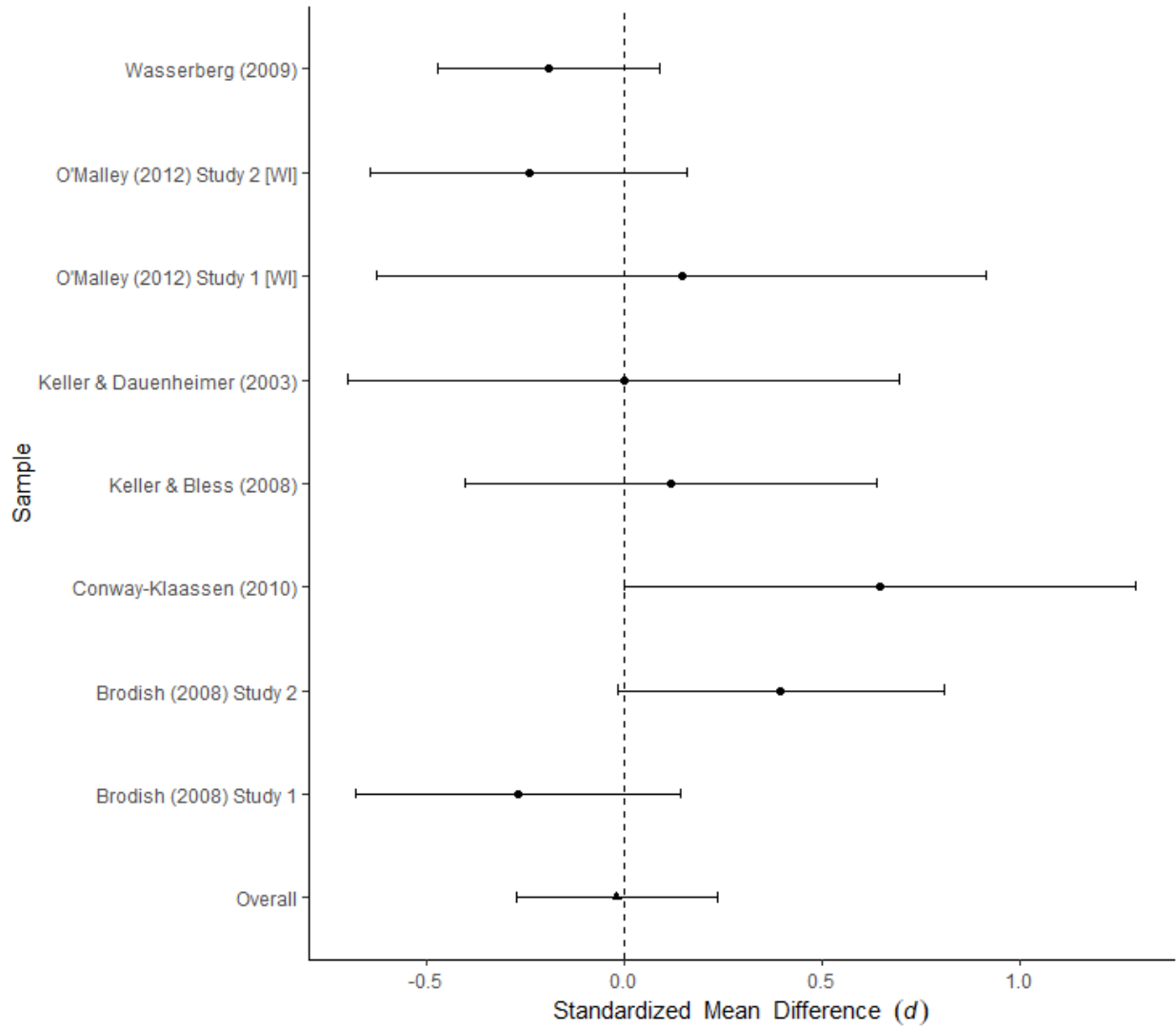


Figure C11. Forest plot for performance avoidance within-group effect sizes

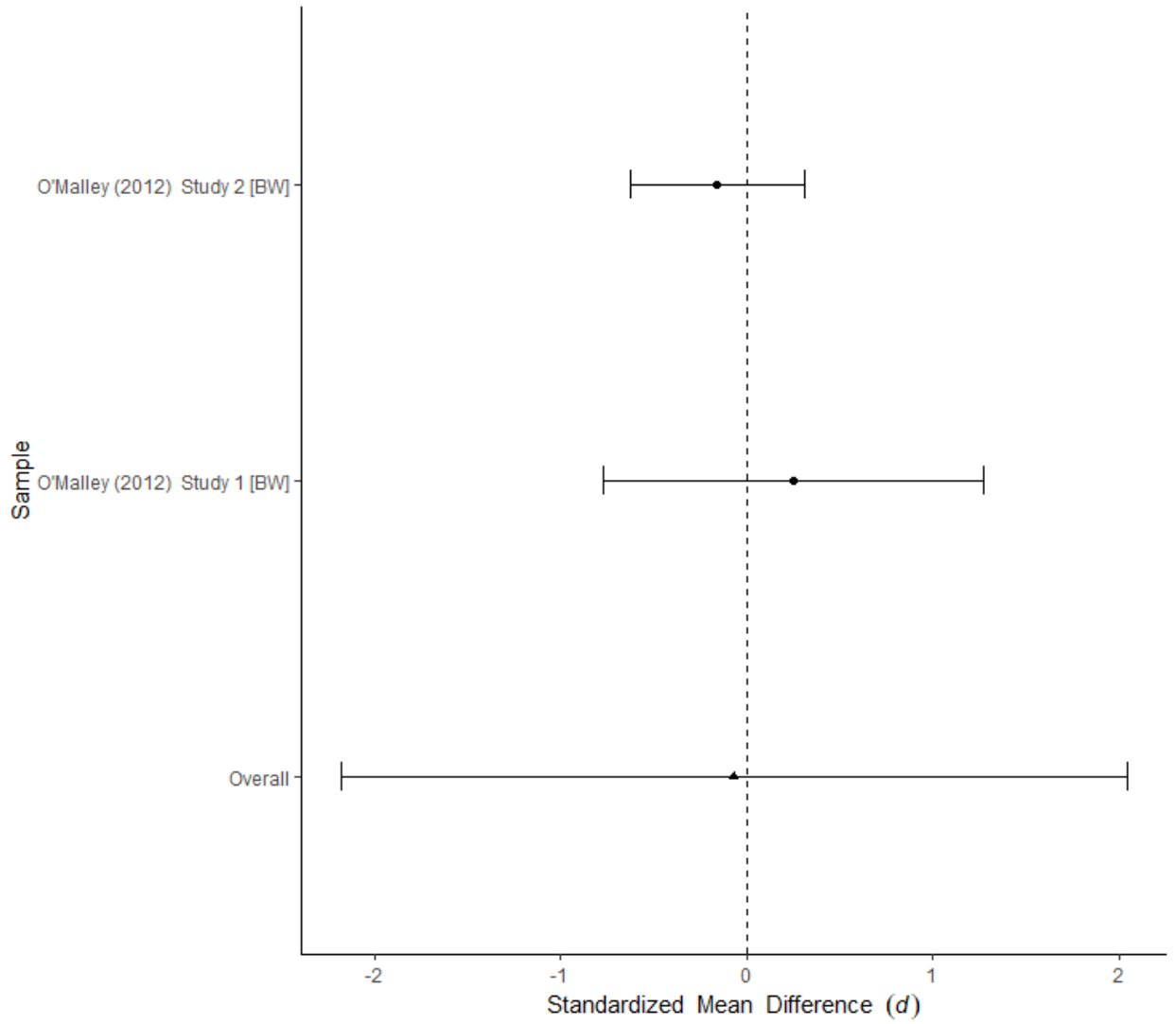


Figure C12. Forest plot for performance avoidance between-group effect sizes

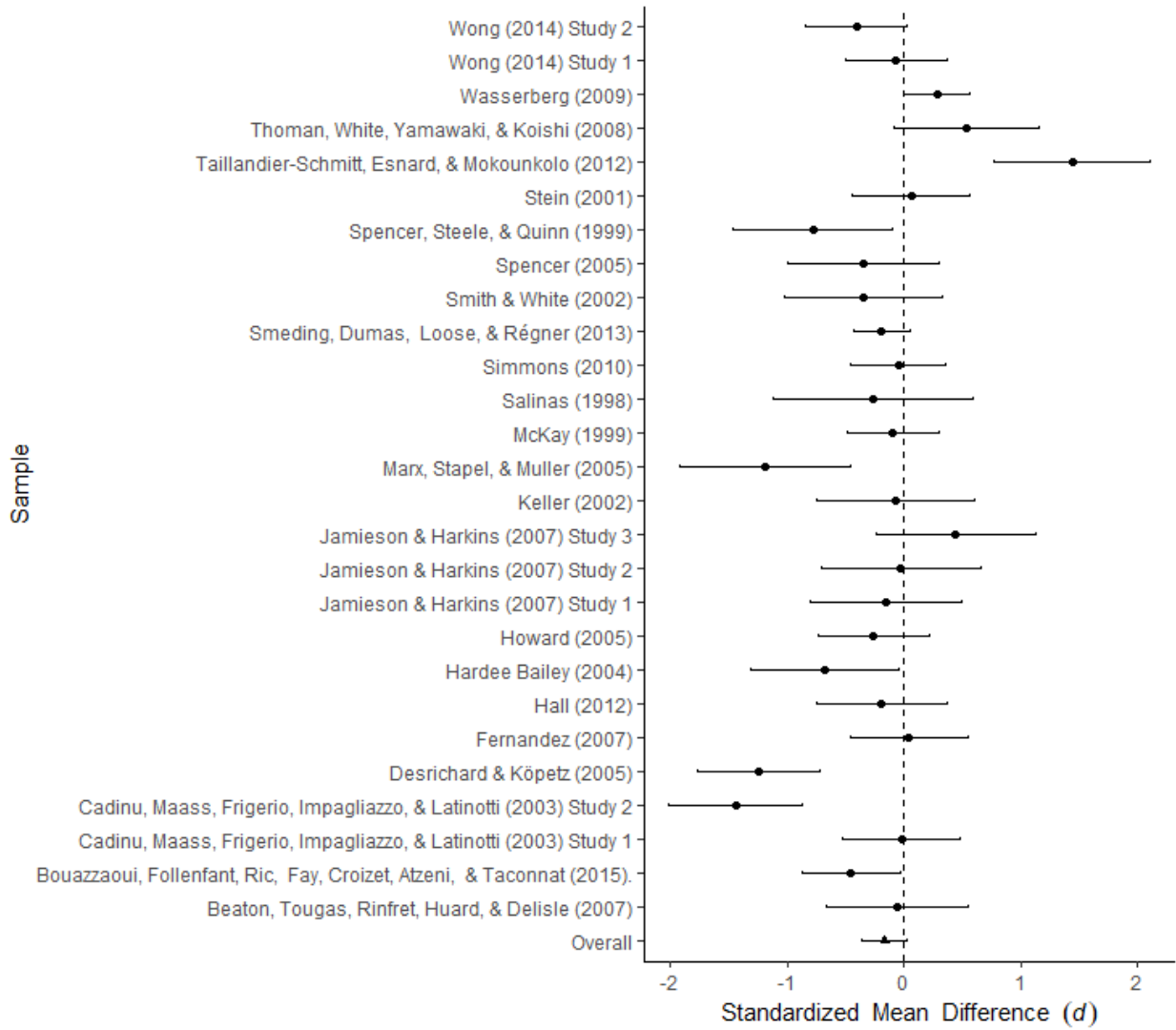


Figure C13. Forest plot for performance expectations within-group effect sizes

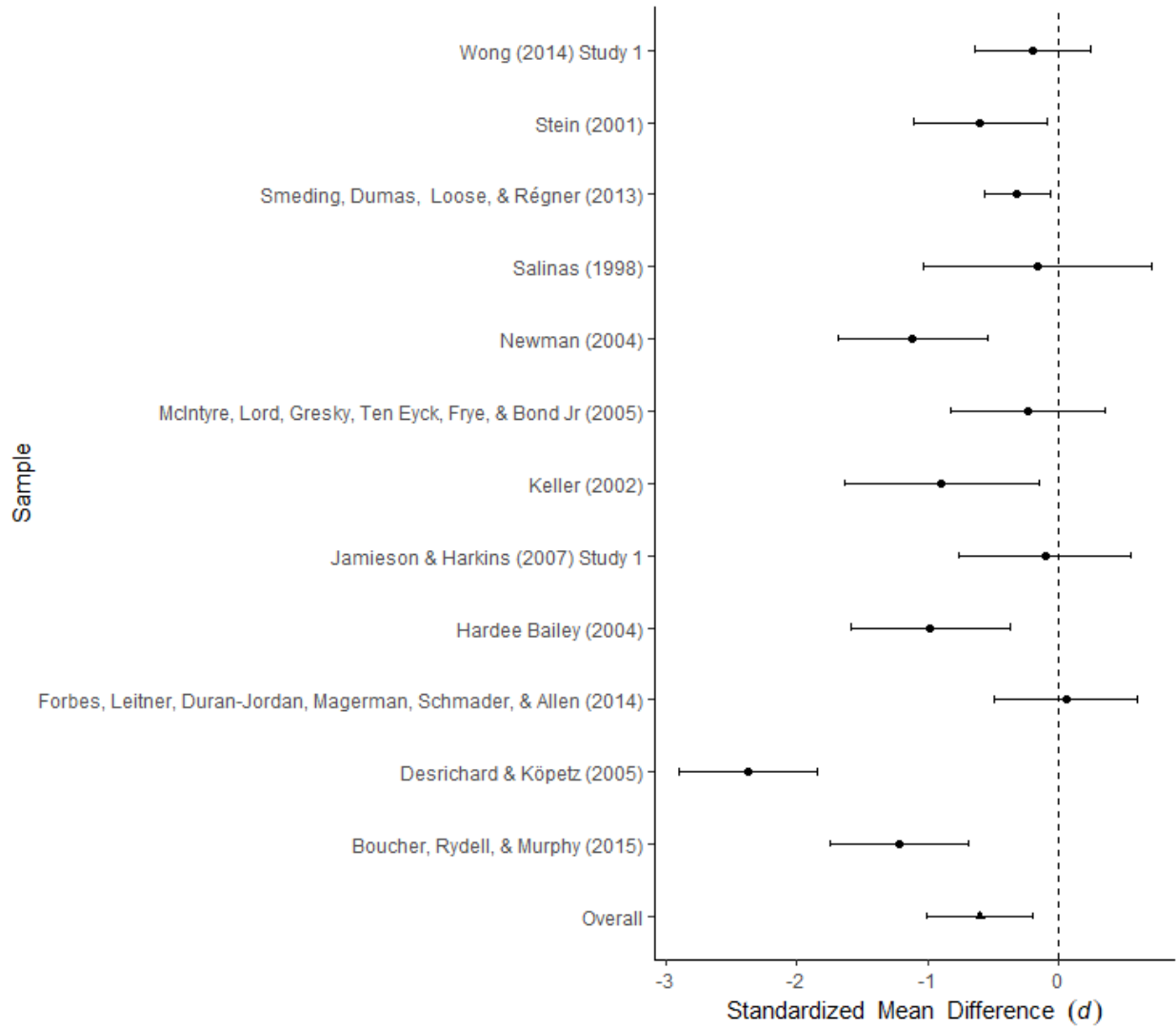


Figure C14. Forest plot for performance expectations between-group effect sizes