2021

The Effects of Exposure to Appropriated Slurs on Stereotype Threat

Charles Mason Guthrie
Walden University

Walden University

College of Social and Behavioral Sciences

This is to certify that the doctoral dissertation by

Charles Mason Guthrie III

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Kimberley Cox, Committee Chairperson, Psychology Faculty
Dr. Anthony Perry, Committee Member, Psychology Faculty
Dr. Elisha Galaif, University Reviewer, Psychology Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2021

Abstract

The Effects of Exposure to Appropriated Slurs on Stereotype Threat

by

Charles Mason Guthrie III

MS, Walden University, 2015

BS, Walden University, 2011

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Social Psychology

Walden University

May 2021

Abstract

Appropriated slurs are refurbished forms of derogatory terms that originate from an outgroup but are adapted by the target group as ingroup terms with specific contextual and application norms. However, the use of appropriated slurs is often debated within the appropriating social group. Within the Black American community, one side of the argument views appropriated slurs as empowering while the other side holds that the underlying slur in its appropriated form causes harm to the integrity of the Black American culture and psyche. Recent survey research supports the perspective that appropriated slurs may have some social benefits; however, social science research has not yet examined how appropriated slurs affect cognition or behaviors. Therefore, the current study addressed this gap by examining how exposure to appropriated slurs affects stereotype activation and academic task performance within the context of the stereotype threat model. In a posttest-only with control group research design, 2x2 ANOVA models were used to compare the mean differences of the dependent variables (stereotype activation and academic performance task) along two independent variables (exposure or no exposure to the selected appropriated slur and racial identity) in 118 Black American adults. The results suggest that appropriated slurs had no effect on the components of stereotype threat. The results have implications for positive social change such that they provide a launching point for further research on the complexity and effects of appropriated slurs.

The Effects of Exposure to Appropriated Slurs on Stereotype Threat

by

Charles Mason Guthrie III

MS, Walden University, 2015

BS, Walden University, 2011

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Social Psychology

Walden University

May 2021

Acknowledgements

This project would not have been possible without the support of several people in my life. I wholly thank my wife and son, Michele and Aidan, for years of patience and understanding the countless times I passed on family activities to focus on completing this research. The continued support from my wife has ensured that I remained focused on my academic goals. Also, I am endlessly grateful to my son for being the inspiration behind the questions that initiated this study—and he continues to inspire me as a social scientist. I also extend my gratitude to Troy and Kina Heflin, for their willingness to volunteer their time to help develop the comedic videos for this research; Troy will always be the funniest person I know. Dr. Kimberly Cox and Anthony Perry also have my deepest gratitude and respect for their guidance and expertise during this project. Their ability to understand and steer my vision for this study was vital to its fruition.

Table of Contents

List of Tables

Chapter 1: Introduction to the Study

A cultural debate within the Black American community concerns the use of appropriated slurs—specifically *nigga*—as an ingroup communication term (Croom, 2013). One view holds that the appropriated slur removes the negativity and derogatory power of the original slur. Ingroup users of the appropriated slur set the rules for its use and meaning within the group. For instance, *nigga* is often used to denote camaraderie or social connectedness between Black Americans (Bianchi, 2014). This appropriated slur can also be modified to express anger, disapproval, or other negative emotions toward other ingroup members. Despite the multicontextual use of the appropriated slur, supporters of its use claim it empowers and uplifts the group (Galinsky et al., 2013; Gaucher et al., 2015). An opposing view holds that any form of the original slur will carry the negative and derogatory connotation of the original slur and such terms should not be used with the group (Allan, 2015; Croom, 2013). This perspective asserts that the derogatory foundations of such a term cannot be removed or altered, and its use is symptomatic of the internalized social prejudices and limit positive ingroup thoughts and behaviors.

Researchers examining the perceptions of appropriated slurs found support for the empowerment perspective (Galinsky et al., 2013); however, such research relied on self-report measures of perceived social status and perceived power, but researchers did not examine cognitive effects (e.g., working memory, attention, etc.) and behavioral outcomes resulting from exposure to appropriated slurs. Exploration of the cognitive and behavioral impacts of appropriated slurs is vital to a deeper understanding of various

social–cognitive–behavioral phenomena such as stereotype threat, which impacts academic and occupational gaps between Black Americans and White Americans in the United States.

In the United States, social inequalities have created a lasting gap in academic achievement for racial minority groups such as Black Americans (Jordt et al., 2017; Musu-Gillette et al., 2017). The National Center for Education Statistics (2014) reported that the average reading and mathematics achievement for children in the United States has increased since the 1970s across all races. However, the data also indicate that, across socioeconomic class, Black American students score lower than their White counterparts on mathematics and reading, and these gaps are found from early education through high school (Musu-Gillette et al., 2017; National Center for Education Statistics, 2014). Subsequently, lower academic performance may lead to less academic achievement and increased disengagement (Wasserberg, 2014). Similarly, Black Americans experience workplace disparities such as lower compensation and fewer opportunities for higher positions compared to their White counterparts (Emerson & Murphy, 2014; Kalokerinos et al., 2014). Black American employees also report less emotional and psychological satisfaction in the workplace due to perceptions of limitation within the organizational culture or structure (Emerson & Murphy, 2014).

Decades of multidisciplinary research has focused on various aspects of both academic and professional achievement gaps between racial minorities and Whites in the United States. However, much of the research initially focused on social group status on a superficial level; for instance, Martin (2009) pointed out that racial variables are typically

used for creating and comparing demographic groups rather than exploring the meaning and experiences related assignment to such groups. That is, education and policy research lacked a deeper understanding of the sociocognitive impacts of social group membership on education and occupational outcomes and experiences. One social psychological model used to examine the impact of race and gender on academic and occupational outcomes is known as *stereotype threat*. Stereotype threat is the psychological discomfort invoked by minimal contextual cues related to salient negative stereotypes that influence implicit cognitive processes and stereotype-related behaviors (Steele & Aronson, 1995).

Contextual framing, cues about one's social group, has been the primary manipulation in stereotype threat research across various stereotyped domains and social groups (McGlone & Pfiester, 2015). Contextual frames may appear in various forms, such as the diagnosticity of a situation, comparisons to other groups, or making relevant stereotypes salient. Yet, in stereotype threat research, scholars have not examined if appropriated slurs act as a contextual frame within the stereotype threat theoretical model. The results of the current study offer much needed insight into the potential contribution appropriated slurs have on the activation of stereotype threat that impedes academic performance among Black Americans. Insights from this study may be used improve intervention programs that target the effects of stereotype threat in academic domains and, thus, may affect positive social change in the persistent academic gap experienced by Black Americans.

In the remainder of this chapter, I briefly summarize the existing literature related to appropriated slurs and stereotype threat as well as identify the overlapping gap in the

research in both areas that was the focus of the current study. Specifically, I highlight these gaps in the problem statement and discuss the need to examine implicit cognitive processes related to appropriated slurs and examine ingroup contextual cues within the stereotype threat model. To that end, I present the purpose of this study and briefly introduce elements of its research design, theoretical framework, scope and delimitations, and limitations. At the conclusion of this chapter, I highlight the significance of the current study within the context of filling a gap in the research and in the context of positive social change.

Background

Slurs and Appropriated Slurs

Research examining slurs and appropriated slurs is often conducted through the lens of communication or linguistics in which definition and functionality of the terms are the common focuses. Such research has been conducted to explore how all words function as descriptive or expressive terms in the context of communication (Croom, 2013). Descriptive terms are typically devoid of emotions or reflections of the speaker's attitudes while attempting to identify individuals or groups by social attributes (e.g., *African American*). Expressive terms are commonly reflective of the speaker's emotional state toward a target and require no descriptive features of the target. Slurs, however, are both descriptive and emotionally charged terms used to identify individuals or groups by social attributes to express negative emotions toward the targeted group (Henry et al., 2014).

Additionally, slurs are typically used to offend, threaten, or derogate the targeted group or individual group member (Henry et al., 2014). The derogatory aspect of slurs suggests that the perpetrator and their group perceives the target group as holding a lower social status and possessing negative attributes; moreover, the use of slurs also indicates that the perpetrator holds negative attitudes or emotions toward the target group. Slurs can produce emotional or psychological distress in target group members, and such words are often viewed as hostile or offensive (Spotorno & Bianchi, 2015). Likewise, slurs may activate the negative stereotypes—cognitive schemata representing beliefs about attributes of the target group—related to the slur when heard by both target group members and others (Burkley et al., 2016; Jeshion, 2013a).

In some cases, slurs are refurbished and adopted by members of the target group as nonoffensive (and even endearing) ingroup terms (Croom, 2013; Galinsky et al., 2003). For example, the Black American culture has appropriated the term *nigger* (with slight variations such as replacing the *–er* ending with an *–a*) as a multiple use ingroup term with specific social and cultural norms regulating its use. However, appropriated slurs are not always accepted universally within the appropriating social group; thus, there are two main attitudes within the Black American community about using the appropriated version of *nigger* as an ingroup term. One side of the argument holds that Black Americans view the appropriated term as an act of empowerment (Croom, 2013)—that is, taking the negative power of the original term and redefining it as a term of group solidarity. Conversely, opponents of the appropriated term within the Black American community view any form of the original term as negative and harmful to the integrity of

the Black American culture and psyche (Rahman, 2011). These opponents assert that the derogatory foundations of a term cannot be removed or altered and use of such terms is a symptom of the internalized social prejudices and can limit positive ingroup thoughts and behaviors.

While the literature points to how slurs in their original form can cognitively and emotionally impact the targets of the slurs (e.g., Burkley et al., 2016; Jeshion, 2013), less is known about how appropriated slurs influence the targets' cognition and behavior. Recently researchers have supported the perspective that appropriated slurs empower target group members and neutralize the negativity of the original slur (Galinsky et al., 2013); however, their research relied on self-report measures of perceived social status and perceived power but did not examine cognitive effects (e.g., working memory, attention, etc.) of appropriated slurs on cognition and behavior.

Stereotype Threat

In their seminal research, Steele and Aronson (1995) observed that cues related to negative stereotypes about academic performance in Black Americans can actually invoke those stereotyped behaviors—a phenomena called *stereotype threat*. Formally defined, stereotype threat is the psychological discomfort invoked by minimal contextual cues related to salient negative stereotypes that influence implicit cognitive processes and stereotype-related behaviors (John-Henderson et al., 2014; Kapitanoff & Pandey, 2017; Lambert et al., 2016; Steele & Aronson, 1995). Decades of research stemming from Steele and Aronson's initial studies has solidified the basic components of stereotype

threat, including contextual framing, stereotype activation, cognitive interference, and behavioral outcomes.

Contextual framing has been established as the catalyst that invokes stereotype threat in individuals. Contextual framing is subtle but relevant cues in the environment (contextual frames) that activate stereotypes and cognitive interference that lead to confirmatory behaviors (e.g., decreased academic performance) in the stereotyped domain. Steele and Aronson (1995) observed that the diagnosticity of a test or asking participants to indicate their race on a demographic form before a test triggers stereotype threat and leads to decreased performance on that test. Other researchers have demonstrated that the race or gender of the proctor or other test-takers could invoke stereotype threat (Kapitanoff & Pandey, 2017). Different forms of media, such as drawings of boys or girls solving (or not solving) math problems correctly, were used by Galdi et al. (2014) in a study examining stereotype threat in young children. Similarly, women who viewed television commercials depicting gender stereotypes performed worse on academic tests (Davies et al., 2002).

Stereotype threat has been implicated in decreased performance in various domains across nearly all social groups (Emerson & Murphy, 2015; Galdi et al., 2014; Thames et al., 2013). Academic performance among racial and gender minorities has dominated the stereotype threat research. Numerous researchers have demonstrated that Black American participants perform worse on academic tests when under stereotype threat compared to Black Americans not under threat and their White test-taker counterparts. Similarly, gender stereotypes also invoke stereotype threat among women

in academic settings, demonstrating how stereotype threat decreases female performance compared to male performance (Galdi et al., 2014). Stereotype threat is not limited to minority groups or academic settings. White male participants under stereotype may perform worse on math tests, empathy behaviors, or athletic tasks when contextual framing invokes stereotype that their group performs worse compared to others (e.g., "White men are inferior athletes compared to Black Americans"; Heidrich & Chiviacowsky, 2015; Stone et al., 1999).

Despite the various contextual frames used in stereotype threat research, no known studies have been conducted to examine appropriated slurs as a contextual cue. Therefore, I addressed this gap in knowledge within the discipline by focusing on how exposure to appropriated slurs affects stereotype activation and behavioral outcomes (e.g., academic performance) and how racial identity may influence such effects.

Problem Statement

Research is needed to address the influence of appropriated slurs on stereotype threat including the activation of stereotypes and stereotyped domain behaviors. There are two gaps in the literature on appropriated slurs and stereotype threat. In the literature related to slurs and appropriated slurs, researchers have demonstrated perceived empowerment and control when people use appropriated slurs in specific social context, supporting the argument that refurbishing negative slurs may have benefits (Galinsky et al., 2013). However, these reserachers used self-reports of both observers and appropriated slur users that lacks measurement related to possible impacts on underlying cognitive processes or resulting behavior outcomes. In the stereotype threat literature,

researchers have focused on a wide range of contextual frames that invoke stereotype threat. For example, classical stereotype threat researchers demonstrated the diagnosticity of a task, identifying one's demographic before a task or the presence outgroup members as sufficient contextual frames to initiate stereotype threat (e.g., John-Henderson et al., 2014; McGlone & Pfiester, 2015; Steele & Aronson, 1995). To date, however, stereotype threat research has not been conducted to examine if appropriated slurs act as a contextual frame within the stereotype threat model.

Research is needed to address these gaps in the literature by examining the influence of appropriated slurs on stereotype threat mechanisms, including the activation of stereotypes and stereotyped domain behaviors. In this study, I address this gap in the literature by focusing on how exposure to appropriated slurs affects stereotype activation and behavioral outcomes such as performance within a stereotyped domain (e.g., academic performance) and how racial identity may influence such effects. This project is unique because it was conducted to address an under-researched area of ingroup use of appropriated slurs, which has become increasingly more common among various minority groups.

Purpose of the Study

The purpose of this study was to quantitatively examine how exposure to appropriated slurs influences stereotype threat. To this end, I employed a quantitative approach modeled after common research paradigms found in the stereotype threat literature. Specifically, in a randomized posttest-only with control group design, I administered two performance tasks and a survey to examine the relationship between

exposure to appropriated slurs and the major outcome components of the stereotype threat paradigm (stereotype activation and stereotyped behavior outcomes) as well as the moderating effect of racial identity among Black American adults. Researchers have often explored racial identity as an attribute that impacts an individual's experience of stereotype threat by either buffering against negative outcomes or exacerbating the threat (Schmader et al., 2015; Shelvin et al., 2014). Therefore, I used appropriated slur exposure as the independent variable, a stereotype activation task and an academic task measure as the dependent variables, and a racial identity measure as a moderating variable. In the next section, I discuss the research purpose and research questions and hypotheses.

Research Questions and Hypotheses

RQ1: What is the difference in academic test performance between Black American adults exposed to appropriated slurs and Black American adults not exposed to appropriated slurs?

$H_0$1: Black American participants exposed to appropriated slurs exhibit similar scores on the academic performance task, as measured by the selected items of the Scholastic Aptitude Test (SAT), compared to Black American participants not exposed to appropriated slurs.

$H_1$1: Black American participants exposed to appropriated slurs exhibit significantly lower scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants not exposed to appropriated slurs.

RQ2: What is the difference in academic task performance between Black American adults who exhibit higher racial identity and Black American adults who exhibit lower racial identity.

$H_0$2: Black American participants who exhibit high racial identity scores, as measured by the Multidimensional Inventory of Black Identity (MIBI) centrality scale, exhibit similar scores on the academic task, as measured by the selected items of the SAT, compared to Black American participants who exhibit low racial identity scores.

$H_1$2: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, exhibit significantly higher scores on the academic task, as measured by the selected items of the SAT, compared to Black American participants who exhibit low racial identity scores.

RQ3: To what extent does racial identity moderate the effects of appropriated slurs on academic test performance?

$H_0$3: Black American participants with high racial identity scores, as measured by the MIBI centrality scale, who are exposed to appropriated slurs exhibit similar scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants with low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

$H_1$3: Black American participants with high racial identity, as measured by the MIBI centrality scale, who are exposed to appropriated slurs exhibit significantly

lower scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants with low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

RQ4: What is the difference in the cognitive activation of negative stereotypes toward one's own racial group between Black American adults exposed to appropriated slur and Black American adults not exposed to appropriated slurs?

$H_0$4: Black American participants exposed to appropriated slurs exhibit similar negative racial stereotype activation, as measured by the word fragment task, compared to Black Americans participants not exposed to appropriated slurs.

$H_1$4: Black American participants exposed to appropriated slurs exhibit significantly more negative stereotype activation, as measured by the word fragment task, compared to Black American participants not exposed to appropriated slurs.

RQ5: What is the difference in the cognitive activation of negative stereotypes toward one's own racial group between Black American adults who exhibit higher racial identity and Black American adults who exhibit lower racial identity?

$H_0$5: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, exhibit similar negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants who exhibit low racial identity scores.

$H_1$5: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, exhibit significantly lower negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants who exhibit low racial identity scores.

RQ6: To what extent does racial identity moderate the effects of exposure to appropriated slurs on the cognitive activation of negative stereotypes toward one's own racial group in Black American adults?

$H_0$6: Black American participants with high racial identity scores, as measured by the MIBI centrality scale, who are exposed to appropriated slurs exhibit similar negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

$H_1$6: Black American participants with high racial identity, as measured by the MIBI centrality scale, who are exposed to appropriated slurs exhibit significantly higher negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants with low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

Theoretical Framework

In this study, I used the stereotype threat model as the central and guiding theoretical framework (Steele & Aronson, 1995). Expanding on concepts from social

identity theory and evaluation apprehension, Steele and Aronson sought to elucidate the sociopsychological phenomenon in which one's actions conform to negative stereotypes about their social group—a phenomenon they would later call *stereotype threat*. Decades of research stemming from Steele and Aronson's initial studies has solidified the basic components of stereotype threat, including contextual framing, stereotype activation, cognitive interference, and behavioral outcomes.

Contextual Framing

Contextual frames are the subtle environmental cues that bring situationally relevant stereotypes into an individual's implicit cognitive processes. In their seminal research, Steele and Aronson's (1995) revealed that contextual frames linked to negative group stereotypes invokes stereotype threat; for example, framing an academic test as either diagnostic or nondiagnostic of intelligence (an attribute that Black Americans are commonly negatively stereotyped) led to lower performance among Black American participants. Subsequent researchers have demonstrated how a wide variety of contextual frames invoke stereotype threat across nearly every social group. Typically, stereotype threat researchers present participants with a contextual frame that initiates activation of the related negative stereotype (often outside the individual's awareness) leading to cognitive interference and undesirable behavioral outcomes.

Stereotype Activation

According to the stereotype threat model, exposure to an adequate contextual frame will implicitly activate the related stereotype in an individual's cognitive processing. Steele and Aronson (1995) first demonstrated this connection by exposing

participants to a contextual frame before administering a word-fragment task that measured race-based stereotype activation. They found that Black American participants in the diagnostic conditions showed greater stereotype activation about their ingroup compared to participants not exposed to the contextual frame. Subsequent researchers confirmed the link between contextual frames and stereotype activation (e.g., Shelvin et al., 2014). Additionally, contemporary researchers have demonstrated that negative stereotype activation leads to negative implicit evaluations about one's group. For example, female children under stereotype threat exhibited greater negative implicit associations toward their gender than female children not under threat (Galdi et al., 2014).

Behavioral Outcomes

Stereotype threat impacts short-term and long-term behavioral outcomes. Short-term behavioral outcomes are the behaviors specific to the nature of the stereotype within the immediate stereotyped situation, such as academic performance when taking a test. Steele and Aronson (1995) first demonstrated how eliciting race-based stereotype threat decreases performance in stereotype relevant behaviors or tasks; in other words, stereotype threat decreased academic performance in Black American participants. Since Steele and Aronson's (1995) seminal research, stereotype threat researchers have demonstrated how a wide variety of behaviors, including athletic ability, interactions with the criminal justice system, driving, health behaviors, and job performance, can be influenced when individuals within a stereotyped group are placed under stereotype

threat (Brochu & Dovidio, 2014; Hakim & Quartiroli, 2016; Lambert et al., 2016; Najdowski et al., 2015).

Researchers using the stereotype threat model have empirically demonstrated that exposure to subtle contextual frames initiates cognitive processes that influence behavioral outcomes across various social groups and behavioral domains; therefore, the stereotype threat model is ideal for this study. Moreover, researchers across decades who have examined stereotype threat offer this study guidance on the necessary components of this model that will facilitate its application, including (a) contextual framing, (b) stereotype activation, and (c) short-term changes in the stereotyped-domain behaviors such as decreased task performance and avoidance of academic activities.

Nature of the Study

I employed a quantitative approach for the current study that is consistent with the stereotype threat research paradigm in which a contextual frame is introduced to participants and the outcome components of the stereotype threat model are measured. To this end, I used a randomized posttest-only with control group research design to examine how the independent variable (appropriated slurs) affected the dependent variables (stereotype activation and academic task performance) among adult Black American participants. I selected this design to avoid priming or sensitizing participants to the academic test. However, to establish a baseline for academic performance in the absence of a pretest measure, I asked participants to rate their academic abilities in secondary school or college by completing a Likert-type items that reads, “I always did well on tests in school” after they complete the academic task (discussed below) in this study. This

approach to establishing a baseline in the absence of a pretest measure has been effectively applied in various stereotype threat studies (see Gonzales et al., 2002 and Steele & Aronson, 1995). Further, I collected data on the variables and sociodemographic information from adult Black American participants via an anonymous online web-based survey tool (i.e., SurveyMonkey). All data were collected from consenting Black American participants using online surveys. I analyzed the collected data using a series of analysis of variance models (described in Chapter 3) to address the presented research questions. In the next paragraphs, I present a brief description of the key variables in the current study—appropriated slurs, stereotype activation, academic task performance, and racial identity.

The appropriated slur (independent variable) of focus was *nigga*, the commonly used adaptation of the original slur (*nigger*) within the Black American community (Croom, 2013). I randomly assigned participants to either the threat condition or the no-threat condition in which they would view similar prerecorded comedic videos either containing or not containing the appropriated slur. In each video, the Black American confederate acting as the comedian delivered the same exact material apart from the appropriated slur. The comedian used the term *nigga* in the stereotype threat condition video and more generic (non-ingroup related) social terms in the no-threat condition video. To ensure the ingroup context of the material was consistent across conditions, *nigga* was replaced with generic terms, such as *friend* or *people*, so that the term conveys ingroup connectedness or references where appropriate. Following Steele and Aronson's (1995) seminal model of stereotype threat research, I measured stereotype activation

using a word fragment task. Word fragment completion tasks are based on cognitive priming concepts that hold a target word is recognized with more ease if it is preceded by a related cue (Heyman et al., 2016). In Steele and Aronson's (1995) seminal research on stereotype threat, they applied a simple word fragment task as a measure of stereotype activation; they presented participants as the lexical access processing (LAP) task—the cognitive retrieval of words. Steele and Aronson built their word fragment task based on similar methods used by Gilbert and Hixon (1991), who noted that word fragment tasks demonstrate the cognitive activation of recently primed or self-generated cognitive constructs.

I used randomly selected items from the math sections of the SAT to measure academic task performance. The SAT, published and maintained by the College Board and Educational Testing Service group, is an instrument intended to measure an individual's verbal and mathematical abilities and is traditionally used to partially inform admissibility into institutions of higher education. The SAT was adapted from early intelligence assessments developed in the 1920s by a committee of psychologists for military recruitment (Gregory, 2007). The first official administration of the SAT was in 1926; over several decades this assessment became increasingly popular and remains a standard instrument.

To measure racial identity, I used the MIBI (Sellers et al., 1997b). The MIBI was developed by Sellers et al. (1997a) to examine the constructs of the multidimensional model of racial identity (MMRI; Sellers et al., 1998), which states that racial groups such as Black Americans may have several social identities, including race, that influence their

cognition and behaviors in various social contexts. The MIBI attempts to capture three of the consistent dimensions of the MMRI—centrality, ideology, and regard—with a 27-item, 7-point Likert-type survey. Overall, higher total scores indicate stronger Black American racial identity.

Definitions

Appropriated slur: A slur that originated as a derogatory and evaluative term targeting a specific social group but was adopted by the target group as a nonoffensive ingroup term (Croom, 2013).

Stereotype threat: The social–psychological process through which an individual's cognition and behavior is influenced by stereotypes related to contextually relevant behaviors (Steele & Aronson, 1995).

Racial identity: The MMRI's definition of racial identity is "the significance and qualitative meaning that individuals attribute to their membership" in their racial group (Sellers et al., 1998, p. 23).

Scholastic Aptitude Test (SAT): An instrument intended to measure an individual's verbal and mathematical abilities, traditionally used to partially inform admissibility into institutions of higher education (Gregory, 2007). Selected items from the SAT were used as a measure of academic task performance.

Assumptions

It was assumed that participants would make reasonable efforts to complete and correctly answer the tasks and surveys in this study; further, it was assumed that participants would understand the questions presented to them and would provide

accurate and honest responses. While stereotype threat is a person-situation phenomenon, there are several individual and environmental factors that may also impact performance. Participants in this study interacted with materials and surveys using their personal computers in a setting of their choice. Various computer operating systems and hardware configurations and user interfaces may have interfered with participants' understanding and responses. Also, participants may have chosen to participate in the study in environments, such as internet cafes, libraries, or their homes, that had distractions or interruptions. Therefore, it was further assumed that the conditions outside of a controlled laboratory environment could mediate participants' level of engagement and performance.

Scope and Delimitations

While appropriated slurs are not limited to a single social group, I focused on Black American participants, age 18 or older, due to the use of the appropriated slur in the Black American community and its well-known ingroup debate over the use of the term (Allan, 2015; Baldwin et al., 2015; Croom, 2015). Further, I focused on the immediate effects of stereotype threat (i.e., test performance and stereotype activation); the long-term impacts and behavioral outcomes such as disidentification and domain avoidance were not examined.

Exposure to appropriated slurs may occur in various social settings and through various channels. For example, an appropriated slur may be used among friends during a social gathering, by musicians in songs, or by actors in television or films. Social context and the speakers of appropriated slurs will likely have a mediational relationship on the

effects of these terms. I focused on Black American adults due to the widespread use of that group's exposure to appropriated slurs and the well-known debate about the use of such terms. Further, the research activities were conducted through a single medium (web-based interface) with participants on their personal computers to reduce the resources needed for the study and to reach a wider base of possible participants.

Limitations

A possible limitation of the current study was the validity of the instruments within the context of the study. The instruments may not have accurately measured stereotype activation, academic task performance, or racial identity due to other factors not measured or observed. Black Americans, for instance, do not all share the same perspectives, values, backgrounds, and social environments, which may lead to different outcomes when exposed to appropriated slurs. Further, participants may belong to a generational cohort that perceives appropriated slurs differently than individuals from another cohort. Further, nearly every minority group in the United States has adopted an appropriated slur for ingroup communication; however, other groups (e.g., lesbians, Hispanics, women) may have different emotional and cognitive reactions to appropriated slurs compared to Black Americans based on historic social experiences.

Another possible limitation to the current study was the demand characteristics: the participants' awareness of the researcher's intent (McCambridge et al., 2012). Such an awareness could have led participants to modify their behaviors and could adulterate the study's findings. In the context of this study, for instance, participants who became aware of the intent to measure the effects of appropriated slurs on cognition may have

attempted to overcome automatic processes that would happen in a real-world situation, thus influencing the study's outcomes. I used two methods to address potential demand characteristics. First, I used deception in the initial recruitment and activities of the study (discussed in Chapter 3). Deceptive methods, however, were disclosed to participants during the debriefing process. Second, I presented a manipulation check question—What do you think this study is attempting to investigate?—after the study activities but before debriefing. The data collected from participants who correctly guessed the true purpose of the study were discarded.

The results of this study may have been limited by the sampling methods and recruitment criteria, limiting generalization of the study's findings to a wider population beyond the study sample and across other groups that use various appropriated slurs. Specifically, selection bias and attrition resulting from the recruitment pool for the this study may have limited the validity of the study. For instance, participants recruited from a preregistered participant pool may share similar attributes or differ in some attribute (e.g., racial identity or academic ability) from individuals who did not volunteer through such participant pools. Similarly, web-based survey methods may present another limitation to the current study; for example, participants with access to web-based technology may be markedly different in the attributes of interest in the study compared to individuals who do not have easy access to the internet. These limitations reflect the scope of the current study and will be discussed in relation to the generalizability of the findings.

Significance of the Study

With this study, I addressed an identified gap in the literature by focusing on how exposure to appropriated slurs influences cognition and stereotyped behaviors in the context of stereotype threat among Black Americans. This study was unique because I addressed an under-researched area of cognitive and behavioral impacts of ingroup use of appropriated slurs that have become increasingly more common among various minority groups (Bianchi, 2014; Croom, 2013). Moreover, of the current literature, this study was the first conducted to examine appropriated slurs as contextual cues in the stereotype threat model.

Overall, the results of the current study provide some insight on the debate (beneficial or harmful to the ingroup) on the value of appropriated slurs. In the context of stereotype threat, the results contribute to how appropriated slurs perform as contextual cues within the stereotype threat model. Broadly, the results of this study begin to contribute to a better understanding of interpersonal communication behaviors such as appropriated slurs. Such understanding, through future research, may lead to more insightful navigation of group identity and its impacts on thoughts and behavior.

In the context of stereotype threat, the results of this study provide much needed insight into the potential contribution appropriated slurs have on the activation of stereotype threat that impedes academic performance among Black Americans. Insights from this project may inspire further researchers to help improve intervention programs that target the effects of stereotype threat in academic domains, which may effect positive social change in the persistent academic gap experienced by Black Americans and other

minority groups. Specifically, the results may contribute to the quality of stereotype threat interventions, such as affirmation-based interventions that aim to reduce the effects of stereotype threat by highlighting the positive aspects of an individual's social identity and how those aspects contribute to their success in a domain where they commonly face negative stereotypes. For example, values affirmation interventions for stereotype threat ask individuals to identify important and valuable characteristics about themselves, which increases self-worth in an environment that threatens their social identity (Covarrubias et al., 2016).

Despite some positive results, the outcomes of affirmation-based interventions for stereotype threat have been inconsistent (Jordt et al., 2017). Possible factors in the variability of affirmation-based interventions may be social and background (e.g., school, community) contexts that contribute to the internalization of negative stereotypes. A better understanding of the ingroup processes and evaluations, such as the use of appropriated slurs and how that impacts an individual's self-evaluation and cognitive process, may elucidate some of these factors responsible for the variability in affirmation-based interventions and inform improved approaches. In turn, more effective stereotype threat interventions will contribute to addressing the persisting educational gaps between minorities and nonminorities in the United States. Furthermore, insights gained from the current study may guide additional research in other social groups (e.g., women, LGBT, Hispanics) that use appropriated slurs.

Summary

A philosophical debate about the utility and effects of appropriated slurs has initiated empirical research on their use; however, such research has been limited to self-reports and lacks cognitive and behavioral observations. Given the prevalence of appropriated slurs in minority cultures in the United States, such ingrained terms are likely to have cognitive and behavioral effects on their social groups. Stereotype threat research has demonstrated that various contextual frames can impact performance behaviors such as academic task performance, but contextual frames sustained by ingroups (i.e., appropriated slurs) have not appeared in the stereotype threat literature. Therefore, research is needed to address these gaps in the literature by examining the influence of appropriated slurs on stereotype threat, including the activation of stereotypes and stereotyped domain behaviors. In the current study, I addressed these gaps in the literature by focusing on how exposure to appropriated slurs affects stereotype activation and behavioral outcomes such as performance within a stereotyped domain (e.g., academic performance).

In Chapter 2, I expand on the aspects of stereotype threat and appropriated slurs. First, I will present seminal and contemporary literature on the stereotype threat model—highlighting the fundamental components of the model—that served as the theoretical framework for this study. Then, I will present a detailed review on appropriated slurs and discuss how these terms served as an independent variable in the study. Further, other key variables including stereotype activation and racial identity will be discussed.

In Chapter 3, I present the rationale and methodological design of the study, including operational definitions of the key variables, the data collection method and related instruments, data analysis, and potential threats to the validity of the study. Further, the target population and sampling approach as well as the recruitment, participation, and ethical procedures will be discussed.

Chapter 2: Literature Review

Introduction

Appropriated slurs are slurs that have been refurbished by members of the target group as nonoffensive, positive, and endearing ingroup terms (Croom, 2013). For example, the Black American community has appropriated *nigger* as a multiple-use ingroup term with specific social and cultural norms regulating the use of the word. However, the appropriation of this slur is not universally accepted within the Black community. Some group members view appropriated slurs as an act of empowerment that neutralizes the negative impact of the original slur, whereas others view it as harmful to the integrity of the group's culture and collective psyche (Croom, 2016). Appropriated slurs are deeply ingrained and widely used in U.S. culture; however, the effects of exposure to such terms on ingroup members are unclear. I sought to examine whether appropriated slurs mediate the negative impacts of the original slur or compound the impacts of the original slur.

Recent researchers have supported the perspective that appropriated slurs empower target group members and neutralize the negativity of the original slur (Galinsky et al., 2013); however, such researchers relied on self-report measures of perceived social status and perceived power but did not examine cognitive effects (e.g., working memory, attention) of appropriated slurs that may lead to behavioral outcomes, such as outcomes examined within the context of stereotype threat. Therefore, research is needed to address this gap by examining the impact of exposure to appropriated slurs on stereotype threat outcomes such as perceived threats of being evaluated based on negative

stereotypes and decreased performance in the stereotyped domain (e.g., academic performance). My goal with this study was to examine if exposure to appropriated slurs activates stereotype threat and influences stereotyped domain performance.

In this chapter, I provide an overview of the literature search strategy I used to review the foundational and extant literature. The stereotype threat model will be outlined and discussed as the main theoretical framework for the current study. Moreover, a literature review of the body of evidence related to the key variables will be presented. I will conclude with a discussion about the gaps in the literature related to the cognitive and behavioral effects of exposure to appropriated slurs.

Literature Search Strategy

For the comprehensive literature search strategy, I used a set of key terms used in isolation and in deliberate combinations as search criteria using multiple search engines filtered to select peer-reviewed journals, books, and selected periodicals from multiple databases including Social Sciences Citation Index, PsycINFO, PsycARTICLES, PsycEXTRA, Science Direct, MEDLINE, ERIC, and Academic Search Complete. I placed key terms into two categories: social psychological terms and language terms. Social psychological terms included *stereotype threat, social identity threat, self-labeling, communication,* and *ingroup communication*. Language terms included *slurs, racial slurs, appropriated slur, reappropriated slur, nigger,* and *nigga*. I used each term in both categories individually as search criteria; moreover, I combined each social psychological term with a language term and used as search criteria. For example, the key variable stereotype threat as well as the combined search terms stereotype threat AND slurs were

used as search criteria. I filtered each search iteration with a date range of 2014–2019; however, some selected literature outside that date range was included if deemed seminal (e.g., Steele & Aronson, 1995) or substantially vital to the context of the current study.

Theoretical Framework

I used the stereotype threat model as the central theoretical framework for this study. In the following theoretical review, I outline the major components of stereotype threat. Moreover, I cite secondary theories and concepts, including social identity theory, stereotype activation theory, and implicit associations, to demonstrate the potential relationship between stereotype threat and key variables.

Stereotype Threat

Expanding on concepts from social identity theory and evaluation apprehension, Steele and Aronson (1995) sought to elucidate the sociopsychological phenomenon in which one's actions conform to negative stereotypes about their social group, a phenomenon later called *stereotype threat*. Specifically, across four studies Steele and Aronson examined how the threat of confirming a negative stereotype about one's group impacted the intellectual test performance of Black Americans. In following sections, I present a review of Steele and Aronson's (1995) seminal studies before discussing the basic components of the stereotype threat model—contextual framing, stereotype activation, and cognitive processes and the behavioral outcomes—that stemmed from their initial work and decades of subsequent research. Further, theoretical frameworks such as the stereotype activation theory and implicit associations that are intertwined and

support the components of the stereotype threat model are presented throughout the discussion.

Seminal Research

In their initial study, Steele and Aronson (1995) predicted that Black American participants exposed to stereotype-relevant cues would exhibit decreased test performance (i.e., number of correct answers) and accuracy (proportion of correct answers to attempted items), lower perceived performance on the test, and lower self-perceived academic competence and self-worth. The researchers asked Black American and White participants to take the same 30-minute academic test derived from the Graduate Record Examination (GRE) after being assigned to one of three conditions. In the threat condition, participants were informed that the test was diagnostic of intellectual ability, a contextual cue that the researchers posited would invoke concerns among the Black American participants about negative stereotypes related to intelligence. Participants in the no-threat condition were informed the test was a problem-solving activity to avoid invoking concerns about intelligence stereotypes. Further, participants in a second no-threat condition were also informed that the problem-solving activity would be challenging; this condition was added to explore if framing the test as a challenge would increase the participants' motivation and performance beyond being nondiagnostic.

The results of test performance using SAT scores as a covariate indicated significant main effects for the threat condition, $F(2, 107) = 4.74$, $p < .02$, with participants in the no-threat challenge condition exhibiting higher scores than participants

in the other two conditions (Steele & Aronson, 1995). A significant main effect for the race condition, $F(1, 107) = 5.22$. $p < .03$, indicated Black American participants exhibited lower scores than their White counterparts. The race by condition interaction, however, was not significant ($p < .19$), which Steele and Arson (1995) reasoned to be the result of White participants in the non-threat condition having marginally higher scores than Black American participants in the same condition. An additional analysis adjusting for such issues reached marginal significance, $F(1, 107) = 3.27$, $p < .08$. Planned contrasts, however, indicated significant group differences that supported Steele and Aronson's predictions. Black American participants in the threat condition scored significantly lower than White participants in the threat condition, $t(107) = 2.64$, $p < .01$, as well as Black American participants in the no-threat condition, $t(107) = 2.88$, $p < .01$, and the no-threat challenge condition, $t(107) = 2.63$, $p < .01$. Overall, no significant effects were found for test accuracy; however, planned contrasts showed that Black American participants in the threat condition were significantly less accurate than their White counterparts, $t(107) = 2.13$, $p < .05$, and Black participants in the no-threat condition, $t(107) = 2.64$, $p < .01$.

Steele and Aronson (1995) found evidence to support their predictions that Black Americans participants under stereotype threat reliably scored lower and less accurately than their White counterparts and Black American participants in the no-threat conditions. However, the overall effects of race by threat condition were only marginally significant and required further examination. Further, Steele and Aronson's (1995) prediction that stereotype threat outcomes are the result of interfering thoughts in the

stereotype-relevant situation was not supported. Perceptions of test performance showed significant condition effects for both estimates of correct answers, $F(2, 106) = 7.91$, $p < .001$, and self-comparison to other test-takers, $F(2, 107) = 3.17$, $p < .05$. Planned contrasts showed that self-perceptions of test performance were lower in Black American participants in the threat condition ($M = 4.89$) when compared to the self-perceptions of Black American participants in the no-threat condition, $t(107) = 2.81$, $p < .01$, and the no-threat challenge condition, $t(107) = 2.40$, $p < .02$. While these results suggest that Black American participants are more self-evaluative about their academic performance under stereotype threat, no significant differences were found for the academic competence and perceived self-worth measures that extend beyond the immediate testing situation.

In their initial study, Steele and Aronson (1995) developed the framework from which stereotype threat research would be conducted for decades. They established that contextual cues related to negative stereotypes impair stereotyped domain behavior (i.e., a diagnostic test decreasing the test performance of Black American participants); however, the underlying mechanisms connecting contextual cues and behavioral outcomes were not yet clear. In subsequent studies, the researchers sought to better examine the interaction between race and stereotype threat as well as the cognitive mechanisms underlying stereotype threat.

Steele and Aronson (1995) conducted a second study to examine if stereotype threat outcomes (e.g., academic test performance) is mediated by anxiety over conforming to the negative stereotypes. They asked Black American and White students ($N = 40$) to complete an academic test and measure of anxiety (Speilberger Anxiety

Inventory, [STAI]) after being assigned to a threat or no-threat condition. Overall, the procedure replicated their previous study apart from slight modifications to the academic test, the exclusion of a no-threat challenge condition, and the inclusion of the STAI after the academic test.

The results on test performance showed a significant main effect of race, $F(1, 35) = 10.84$, $p < .01$, and race by threat condition interaction, $F(1, 35) = 8.07$, $p < .01$; however, there were no significant main effects for threat condition. In addition to significant main effects of the race and threat condition interaction, the planned contrasts confirmed Steele and Aronson's (1995) initial findings. That is, Black American participants in the threat condition significantly scored lower on the academic test than Black American participants in the no-threat condition, $t(35) = 2.38$, $p < .02$, and lower than White participants in the threat condition, $t(35) = 3.75$, $p < .001$, and no-threat condition, $t(35) = 2.34$, $p < .02$.

When examining accuracy, Steele and Aronson (1995) found an overall significant effect of the race by threat condition interaction, $F(1, 35) = 4.18$, $p < .05$. However, the planned contrasted were only marginally significant showing Black American participants in the threat condition underperforming compared to White participants in the same condition and Black American participants in the no-threat condition. Further, they compared the number of test items completed and the speed at which participants completed items. The main effects for completed items and speed for the race by threat condition interaction were not significant but the planned contrasts

showed that on average Black American participants in the threat condition completed fewer test items and took longer to complete the first five items on the test.

Across these two initial studies, Steele and Aronson (1995) found evidence to support that stereotype threat affects test performance in threat-targeted participants (Black Americans) like other forms of evaluative pressure by impairing speed and accuracy. However, they were still unable to provide evidence that stereotype threat creates anxiety specifically stemming from the concerns about confirming negative stereotypes—self-reported effort, cognitive interference, and self-reported anxiety measures collected during their second study did not exhibit significant main effects. To address this issue, the researchers conducted a third study to incorporate stereotype activation.

After establishing that stereotype threat affects academic test performance for Black American participants, Steele and Aronson (1995) conducted a third study in which they specifically examined if stereotypes become cognitively activated under stereotype threat. Black American and White participants (N = 68) were assigned to either the threat, no-threat, or control condition before taking the primary dependent measures of stereotype activation, stereotype avoidance, performance apprehension, and self-handicapping. The academic test, however, which was a primary outcome in the previous study was not administered to any participants. Steele and Aronson (1995) reasoned that the stereotype activation and avoidance measures may invoke stereotype threat in Black American participants across all the research conditions and diminish the effects of the condition manipulations. Like their previous two studies, participants in the

threat and no-threat conditions were presented with test instructions that were either diagnostic or nondiagnostic of their intellectual ability; moreover, participants in the control condition were not presented with test instructions before dependent measures.

Steele and Aronson (1995) tested the assumption that racial stereotypes and anxiety related to confirming those stereotypes will be more cognitively active under stereotype threat. To measure stereotype activation, a word-fragment completion task known to measure the activation of cognitive structure (e.g., stereotypes) was administered. The 80-items task presented participants with incomplete word fragments with at least two blank space placeholders for participants to complete the work. Twelve of the word fragments only had one possible solution related to racial constructs or Black Americans stereotypes and seven of the word fragments reflected self-doubts about ability. The results for stereotype activation indicated significant main effects for threat condition, $F(2, 61) = 5.90$, $p < .005$, and race, $F(2, 61) = 13.77$, $p < .001$, qualified by the race by threat condition interaction, $F(2, 61) = 3.30$, $p < .05$. Black American participants in the threat condition completed more race-related word fragment ($M = 3.7$) than Black American participants in the no-threat condition ($M = 2.1$), $t(61) = 3.53$, $p < .001$. The self-doubt activation measure also produced significant main effects for threat condition, $F(2, 61) = 3.53$, $p < .001$, and race by condition interaction, $F(2, 61) = 3.34$, $p < .05$. Black American participants in the threat condition completed more self-doubt word fragments than Black American participants in the no-threat condition, $t(61) = 3.52$, $p < .001$. Moreover, Black American participants in the threat condition significantly ($p < .05$) completed more race-related and self-doubt word fragments than participants in all

other groups. These results suggest that stereotype threat does cognitively activate negative stereotypes and thoughts of self-doubt.

Additionally, Steele and Aronson (1995) tested the assumption that when racial stereotypes are cognitively activated under stereotype threat that individuals will attempt to alleviate anxiety about conforming to those stereotypes by disassociating from those stereotypes. The stereotype avoidance measure asked participants to rate their preferences on various activities as well as self-perceived attributes related and not related to Black American stereotypes. The self-handicapping measure simply asked participants to rate various how much external factors (e.g., sleep, stress, fairness of standardized tests) could influence their test performance. Significant main effects for condition, $F(2, 61) = 4.73$, $p < .02$, and the race by condition interaction, $F(2, 61) = 4.14$, $p < .03$, were also found for the stereotype avoidance measure. Black American participants in the threat condition tended to rate their interests and self-descriptions less conforming to Black American stereotypes than Black Americans in the no-threat condition, $t(61) = 3.61$, $p < .001$, or White participants ($p < .05$). Relatedly, an analysis the demographic questions revealed that Black Participants in the threat condition were less likely to indicate their race compared to all other groups. They also examined the degree of test apprehension invoked by stereotype threats operationalized as self-handicapping their expected performance before the test. The self-handicapping measure showed significant main effects for threat condition, race, and the race by condition interaction across the sleep, focus, and fairness of tests subscales but not for stress. Black American participants in the threat condition tended to self-handicap more than participants in all other groups.

Steele and Aronson's (1995) third study found evidence to support their predictions that stereotype threat can invoke performance anxiety related to confirming negative stereotypes. Specifically, they demonstrated that Black Americans under stereotype threat exhibited cognitive activation of Black American stereotypes, self-doubt, and stereotype avoidance. However, while the evidence showed that threat condition manipulation disrupted test performance (study one and two) and cognitively activated stereotypes (study 3), Steele and Aronson's initial three studies did not demonstrate that the activation of stereotype and the related anxiety mediated decreased test performance.

Steele and Aronson (1995) conducted a fourth study to examine if cognitively activating stereotypes through other contextual cues (rather than test diagnosticity) would result in similar outcomes (i.e., decreased test performance). Black American and White participants ($N = 47$) were assigned to the threat or no-threat condition. In the threat condition participants were asked to indicate their race before taking the test and no demographic items were presented to the participants in the no-threat condition. Across both threat conditions, the participants were presented with the non-diagnostic description of the test used in Steele and Aronson's (1995) previous studies.

The results indicated a significant main effect for the threat condition, $F(1, 39) = 7.82$, $p < .01$, on test performance. Planned contrasts showed that Black American participants in the threat condition scored lower than Black American participants in the no-threat condition, $t(39) = 2.43$, $p < .02$, White participants in the threat condition, $t(39) = 2.87$, $p < .01$, and White participants in the no-threat condition but not significantly.

Analysis of test accuracy showed Black American participants in the threat condition did exhibit less accuracy compared to participants in other groups. However, only the difference between Black Americans in the threat and no-threat condition reach marginally significance while all other comparisons were found to be non-significant. Significant main effects were found for the threat condition, $F(1, 39) = 12.13$, $p < .01$, for completed number of items. Black American participants in the threat condition completed less items ($M = 11.58$) compared to White participants in the same condition ($M = 20.15$), $t(39) = 3.38$, $p < .001$. Black American and White participants performed equally in the no-threat condition.

A MANOVA analysis of the stereotype threat measure showed a significant race effect, $F(9, 31) = 8.80$, $p < .01$ (Steele & Aronson, 1995). Black American participants reported that evaluations of their academic ability are related to their race. Black American participants also distanced themselves from Black American stereotypes more than White Participants; for example, Black American participants reported valuing athletics less than their White counterparts, $F(1, 39) = 4.11$, $p < .05$. Moreover, Black American participants who perceived their test performance was lower tended to devalue Black American stereotypical activities more than Whites. This relationship between perceived test performance and stereotype avoidance was strongest ($r = .70$) among Black American participants in the threat condition.

Basic components of stereotype threat. Steele and Aronson's (1995) seminal studies outlined the basic components of the stereotype threat model and laid the foundation for decades of subsequent transdisciplinary research across nearly all social

groups within various situational contexts. Even though such subsequent research has explored new complex factors and aspects of stereotype threat, the basic components demonstrated by Steele and Aronson have been common throughout various lines of research. These components include contextual framing, stereotype activation, cognitive interference, and behavior outcomes (e.g., task performance and domain avoidance and will discussed below.

Contextual Framing

Steele and Aronson's (1995) seminal research revealed that negative stereotypes need not be explicitly presented to invoke stereotype threat but only threats in the *air*. That is, in specific situations in which negative stereotypes could become relevant, environmental cues may bring those stereotypes into the cognitive process of those targeted by the stereotypes. This contextual framing has been the primary manipulation in stereotype threat research across various stereotyped domains and social groups (McGlone & Pfiester, 2015). Contextual frames may appear in various forms such as the diagnosticity of a situation, comparisons to other groups, or making relevant stereotypes salient through images or subtle references.

Steele and Aronson (1995) first showed that diagnostic contextual frames were associated with stereotype threat when they informed participants that "personal factors" were being measured by an academic test (p. 799). Since Steele and Aronson, a plethora of stereotype threat research has focused diagnosticity contextual framing. John-Henderson, et al. (2014), for example, demonstrated that manipulating the diagnosticity of an academic test not only decreased test performance but increase inflammation—a

biological process in which the immune system increases the production of inflammatory cytokines to fight infection or injury. In one their studies, John-Henderson et al. (2014) assigned 90 male and female undergraduate students to a stereotype threat condition or a no-threat condition. Using the classical stereotype threat paradigm, the participants in the stereotype threat condition were told they were about to take a test that was diagnostic of their intellectual ability while participants in the no-threat condition were told the same test was a problem-solving task (non-diagnostic). Prior to the test, biological measurements related to inflammation were recorded and a questionnaire about the participants SES background (from childhood and current) was administered. The inflammation measures were taken again after participants completed the test.

Simultaneous regression models examining all the two-way terms was conducted to examine how the influence of childhood SES or current SES impacted any of the outcome variables while controlling for the effect of either SES background (John-Henderson et al., 2014). A main effect of diagnostic condition on the inflammation measures was found, $b = .60$, $t(80) = 3.66$, $p < .001$, demonstrating that diagnosticity invoked the expected biological stress. Moreover, reported early childhood SES experiences and diagnosticity interactions were significant for the posttest biological measures, $b = –.61$, $t(80) = –2.85$, $p < .01$, and simple slopes indicated a negative relationship between inflammation and childhood SES for participants in the diagnostic condition. However, no significant effect was found for current SES. Similar to previous stereotype threat research, the researchers found a significant relationship between threat conditions and test performance, $b = –.93$, $t(83) = –3.88$, $p < .001$. Specifically,

participants in the stereotype threat condition (diagnosticity) exhibited lower test performance compared to participants in the no-threat condition. In contrast to the stress measures, early childhood SES experiences revealed no significant relationship with test performance while the relationship between current SES and test performance was significant, $b = .65$, $t(83) = 3.66$, $p < .001$.

The diagnostic contextual frame has been a common manipulation in stereotype threat research examining academic outcomes; however, diagnostic contextual framing has also been extended to other stereotyped domains such as athletic skill, automobile driving ability, and communication (Lambert et al., 2016; McGlone & Pfiester, 2015; Stone et al., 1999). For instance, McGlone and Pfiester (2015) demonstrated how manipulating the diagnosticity of interpersonal skills invokes stereotype threat and threat influences oral communication and conflict resolution. The researchers assigned 209 male and female participants into one of three conditions: leadership, relationship, or the control condition. Participants in the leadership condition were told that the purpose of the study was to investigate leadership ability and participants in the relationship condition were informed that the study was concerned with the ability to maintain close relationships. No diagnostic information was presented to participants in the control condition. All participants were asked to read fictional vignettes about interpersonal conflicts in a professional setting and make a 90-second audio recording of their recommendations for resolving the situation and complete self-report anxiety surveys. McGlone and Pfiester (2015) reasoned that contextually framing the leadership condition would activate stereotypically male criteria (e.g., leaders are assertive) and the

relationship condition would activate stereotypically female criteria (e.g., close relationships require empathy and caring) leading to stereotype threat in females and males, respectively. That is, they predicted that female participants in the leadership condition and male participants in the relationship condition would experience stereotype threat measured by oral communication presentation (disfluency and tentative language), the number of resolution recommendations, and self-report anxiety measures.

Their results revealed that females in the leadership condition produced more disfluency ($M = 7.29$, $SD = 2.46$) compared to their female counterparts in the relationship ($M = 4.45$, $SD = 1.79$) and control conditions ($M = 4.82$, $SD = 1.80$), $F(1, 185) = 9.79$, $p < 0.05$ (McGlone & Pfiester, 2015). Similarly, females in the leadership condition produced more tentative language ($M = 1.85$, $SD = 1.79$) than females in the other conditions, $F(1, 185) = 8.57$, $p < .05$. Comparable patterns were found for males in the relationship condition—these participants produced more disfluency ($M = 8.11$, $SD = 2.67$) than males in the leadership ($M = 5.82$, $SD = 1.90$) and control conditions ($M = 6.23$, $SD = 2.09$), $F(1, 185) = 8.45$, $p < .05$. Males in the relationship condition also showed more tentative language ($M = 1.55$, $SD = 0.50$) when compared to males in the leadership ($M = 1.15$, $SD = 0.40$) or control conditions ($M = 1.08$, $SD = 0.43$), $F(1, 185) = 6.25$, $p < .05$. However, there were no statistically significant differences in the number resolution recommendations made by female or male participants across groups.

Holding to the concept that diagnostic contextual framing assesses an individual's personal factors in a stereotype relevant situation, McGlone and Pfiester (2015) demonstrated how diagnosticity impacts interpersonal communication performance.

Building on Steele and Aronson's (1995) original research, studies such as John-Henderson et al. (2014) and McGlone and Pfiester (2015) demonstrate how diagnostic contextual framing can invoke stereotype threat across different social groups and affect various outcomes. Contextual frames can also by established by the presence of or comparison to outgroup members in the relevant situation. The mere presence of outgroup members that are stereotypically superior in a domain may invoke stereotype threat in targeted group member while domain exemplars or roles from the targeted ingroup may alleviate the effects of stereotype threat (Marx & Roman, 2002; Stout et al., 2010). More recently, Kapitanoff and Pandey (2017), examined the connection between the gender of college-level statistics instructors and stereotype threat among their female students.

Kapitanoff and Pandey (2017) administered a questionnaire of stereotype endorsement (specifically about females being worse at math than males) and measures of test and math anxiety to 451 participants enrolled across 11 basic level statistics courses taught by male and female instructors. Further, as an indicator of stereotype threat, the researchers calculated an underperformance index that reflected the participants' previous grade point average (GPA) and exam scores collected at the beginning and end of the semester. Overall, their results showed a main effect for instructor gender on the performance indicator, $F(1, 181)=7.413$, $p = .001$ (Kapitanoff & Pandey, 2017). Specifically, female participants that were taught by male instructors showed increases in underperformance ($M = -0.07$) over the course of the semester while

female participants taught by female instructors showed decreases in underperformance (M = 0.01).

Moreover, female participants that endorsed the math stereotype reported greater amounts of anxiety about math courses than female participants that did not endorse the stereotype or gave no opinion (Kapitanoff & Pandey, 2017). Interestingly, women who endorsed the math stereotype and had a female teacher performed worse at the beginning of the course but significantly improved by the end. In these cases, the authors posited that upward social comparison and the perceptions of lower competence compared to their teacher cause an initial anxiety and diminished performance. However, female instructors may have acted as role models and resources for those students over the course of the semester leading to the decrease in underperformance.

Contextual frames simply drawing attention to an individual's group memberships may also be sufficient to invoke stereotype threat. Steele and Aronson (1995) first noticed the connection between identifying one's race and stereotype threat when Black American participants who were under stereotype threat refused to complete the racial identification item on a post-test demographics form. Riciputi and Erdal (2017) investigated how making group identity salient invokes stereotype threat in student athletes. The researchers administered a demographics form and a 10-item math test to 60 student-athletes after assigning them to either the stereotype threat condition or the control condition. In the stereotype threat condition, participants were asked to report their gender, cohort year, and any school-related activities in which they recently or

currently participate. The last of the listed school-related activities included varsity sports. Participants in the control condition were only asked their age and cohort year.

A 2x2 factorial design (stereotype threat condition by gender) using the number of items attempted and items correct on the math test as the dependent variable indicated significant main effects (Riciputi & Erdal, 2017). Participants in the stereotype threat condition attempted fewer test items ($M = 7.56$) than participants in the control condition ($M = 8.90$), $F(1, 54) = 4.64$, $p < .05$. Similarly, participants in the stereotype threat condition correctly answered fewer items ($M = 3.93$) than participants in the control condition ($M = 5.37$); however, these results were marginally significant, $F(1, 54) = 3.61$, $p = .06$. These results support Riciputi and Erdal's hypothesis that merely indicating one's social identity can invoke stereotype threat. Interestingly, there was no effect of indicating gender suggesting the different social identities may become more salient than others depending on the contextual frame.

Stereotype threat may also be invoked by introducing contextual frames that make a specific stereotype salient. For instance, Galdi, et al. (2014) conducted a study in which they asked male and female children to color pictures that either depicted a boy successfully solving a math problem (and a girl not), a girl successfully solving a math problem (and a boy not), or a nature landscape. After the participants colored their assigned picture, the completed an age-appropriate math test among other measures (implicit associations and stereotype endorsement). The researchers predicted that coloring the stereotype-congruent picture (i.e., boys are better at math than girls) would

invoke stereotype threat by making the math stereotype salient, which would be expressed in lower math test scores.

An ANOVA revealed a significant gender and condition interaction, $F(2, 234) = 4.69$, $p = 0.1$; moreover, separate analyses by gender supported the researchers prediction (Galdi et al., 2014). The ANOVA on math scores with only female participants produced a significant main effect on condition, $F(2, 117) = 3.66$, $p < .03$. Specifically, female participants that colored the stereotype-congruent picture that was expected to make math stereotypes salient (stereotype threat condition) performed significantly worse than female participants who colored the non-stereotype-congruent picture or the landscape. Moreover, male participants showed no significant difference in math scores across the conditions.

Making a stereotype salient does not necessarily rely on explicit images or narratives as demonstrated by the Galdi et al. (2014) study but can be accomplished through distal or indirect connections. For example, Pacilli et al. (2016) examined how sexualized images such as those depicted in the fashion industry invoke stereotype threat and decreased math test performance. Across two studies, Pacilli et al. (2016) presented female and male participants between the ages eight and ten with sexualized and non-sexualized fashion images and assessed their performance on a math test. In their first study, the researchers assigned female and male participants to a stereotype threat condition in which participants viewed sexualized images of female children or a control condition in which participants viewed nonsexualized images of female children.

The results indicated a significant main effect for gender, $F(1, 75) = 6.27$, $p < .05$, with males scoring higher ($M = 10.04$) than females on the math test ($M = 8.82$) but no significant main effect was found for condition, $F(1, 75) = 3.32$, $p = .073$ (Pacilli et al., 2016). However, there was a significant main effect for the interaction between gender and condition, $F(1, 75) = 9.09$, $p < .05$. Female participants in the stereotype threat condition scored lower on the math ($M = 7.64$) than their control condition counterparts ($M = 10.00$), $F(1, 75) = 13.61$, $p < 0.05$, and no significant differences in males across conditions were found.

In a second study, Pacilli et al. (2016) utilized a similar format to their first study except participants were exposed to images (sexualized and non-sexualized) of children matching their own gender and a working memory assessment was administered to the participants in addition to the previously used math test. Comparable to the first study, a significant main effect of gender, $F(1, 98) = 10.12$, $p < .05$, was found with female participants scoring lower ($M = 8.99$) than male participants ($M = 9.99$) on the math test. In contrast to the first study, a significant main effect for condition was found, $F(1, 98) = = 46.21$, $p < .05$, showing participants in the threat condition scored lower ($M = 8.43$) than participants in the control condition ($M = 10.57$), and no significant interaction between gender and condition was found in the second study. Examining working memory as a dependent variable, Pacilli et al. found a significant main effect of condition, $F(1, 98) = 8.58$, $p < .05$, that revealed participants presented with same-gender sexualized images (stereotype threat condition) exhibited less working memory capacity than participants in the control condition. Gender alone and the interaction between

gender and condition did not produce significant effects in the working memory models. The Pacilli et al. studies, taken together, show that exposure to stereotyped media images can subtly make gender stereotypes salient and create a contextual frame that leads to stereotype threat.

Stereotype threat researcher have clearly demonstrated that manipulating a contextual frame in a stereotype relevant situation can invoke stereotype threat. However, the path between contextual framing and stereotype threat is not direct but routed through cognitive processes such as stereotype activation and automatic associations, which will be discussed in the following sections.

Stereotype Activation and Implicit Association

Steele and Aronson (1995) found that minimal environmental or situational cues (contextual framing) were sufficient to invoke group related stereotypes in members of that group; for instance, Black Americans asked to indicate their race before a test exhibited cognitive activation of racial stereotypes more than Black Americans not asked about their race. Further, contemporary research has demonstrated that stereotype activation under stereotype threat leads to more negative evaluations of one's own ingroup. This component of the stereotype threat model sits on the foundation of two theoretical concepts, stereotype activation theory and implicit association, which will be briefly discussed here.

Stereotype activation theory (SAT) posits that making a stereotype cognitively accessible within relevant situation can influence behaviors and attitudes even when individual do not subscribe or agree with those stereotypes (Gupta et al., 2013).

Moreover, SAT holds that behavioral reactions and attitudes are dependent on the way stereotypes are activated—explicitly or implicitly (Gupta et al., 2013; Neguţ & Sârbescu, 2014). Explicit activation involves directly linking some attribute to the stereotyped group such as describing Black Americans as unintelligence or lazy. Stereotypes that are explicitly activated are more likely to illicit counterstereotypical behavior and opposition to the stereotype (Gupta et al., 2013). Conversely, implicit activation indirectly invokes stereotypical characteristics with a behavioral domain or task; for example, describing effective medical nurses as empathic and caring (stereotypical female attributes). Stereotypes that are implicitly activated are more likely to illicit stereotype-congruent behaviors and assimilation of the stereotype—even if the individual does not believe the stereotype to be true. The effects of implicitly activated stereotypes established by the SAT is mirrored by the stereotype threat model's contextual framing in which subtle situational cues activate relevant stereotypes leading to stereotype-congruent behaviors.

Further, SAT makes the fundamental assumptions that stereotype activation relies on stereotypes that are well-known and ingrained in society and that targets of stereotype have a basic knowledge of the stereotypes (Neguţ & Sârbescu, 2014; Shelvin et al., 2014; Steele & Aronson, 1995). That is, stereotypes must be common knowledge and accessible to the members of the target group. Societally ingrained stereotypes are often established and perpetuated through social interactions and cultural mediums including films, music, and media. For instance, film and television in the United States traditionally portrayed women as weak and emotion compared to men and Hispanics and Black Americans as criminals (e.g., drug dealers and gang members) or as less intelligent

than Whites (Neguţ & Sârbescu, 2014; Schmader et al., 2015). If a stereotype is well known and ingrained into society, it can be reasoned that individuals within that society have some basic knowledge of that stereotype.

The activation of stereotypes is also assumed to rely on the individual's awareness of the stereotype and its content (e.g., Black Americans are unintelligent). SAT's assumptions about stereotype prevalence and knowledge are extended to the stereotype threat model. Shelvin et al. (2014) examined how stereotype awareness, among other factors, impacted stereotype activation and stereotype threat. In an initial session, 186 Black American participants between ages 10 and 12 were asked to complete various forms and assessments including a Stereotype Awareness Task in which they were instructed to list all the stereotypes they knew about their racial group. In a follow-up session (two weeks later), the researchers randomly assigned participants into a stereotype threat condition or a no-threat condition and administered the Test of Adolescent Language (TOAL). In the threat condition, a White proctor told participants that the TOAL was an assessment of their intelligence and scores between racial groups would be compared. Participants in the no-threat condition were told that the test was to determine the quality of the test items.

Nearly all (94%) of the participants reported knowledge of at least one racial stereotype about their group and on average participants listed approximately five stereotypes (Shelvin et al., 2014). The stereotypes listed by participants were categorized into four main groups: unintelligence, economic status (poor), unattractiveness, and athletic ability (good athletics). The unintelligent stereotype was the most common

stereotype reported. Older participants (age 12 and 11) produced significantly more stereotypes than their younger counterparts, $F(2, 171) = 3.34$, $p = .038$. Twelve-year-old participants produced an average nearly two more stereotypes than eleven-years participants and 10-year-old participants. A 2x2 ANCOVA examining the relationship between the stereotype threat condition and whether participants were aware of the intelligence stereotype show an overall main effect of stereotype threat, $F(1, 143) = 6.60$, $p = .011$, which was qualified by a significant stereotype awareness interaction, $F(1, 143) = 5.54$, $p = .02$. Further analysis revealed that only participants who reported knowledge of stereotypes that Black Americans are unintelligent experienced stereotype threat effects, $F(1, 143) = 10.1$, $p = .002$. That is, participants who were aware of the stereotype had significantly lower TOAL scores in threat condition than those in the no-threat condition. Moreover, participants who did not list the intelligence stereotype in the initial session showed no differences in TOAL scores between the threat conditions.

While stereotype awareness is required in the stereotype threat model, an individual does not necessarily need to believe the stereotype to be true but must perceive that others believe the stereotype. That is, the expectation that others such as researchers, teachers, or other outgroup members endorse a stereotype about one's group influences stereotype activation and stereotype threat more than the targeted individual's beliefs (Picho & Schmader, 2018). Further, chronic, or prolonged exposure to stereotypes is not required to activate stereotypes and influence stereotype threat in targeted individuals.

The activation and influence of stereotypes in the stereotype threat model suggests implicit cognitive processes—residual influences of past experiences or

knowledge that affect current behavior outside the awareness of the individual (Greenwald et al., 1998). Stereotype activation under stereotype threat invokes the implicit associations with the negative aspects of the underlying stereotype—even if the individual does not subscribe to the stereotype. That is, even if individuals who hold counter-stereotype beliefs fall under stereotype threat, those counter-stereotypic associations may not necessarily be accessible in their thoughts while the stereotypic associations remain. Galdi et al. (2014) hypothesized that contextual framing negative gender stereotypes related to math performance would activate implicit associations even when the targets do not endorse the stereotype. As previously discussed, the researchers assigned 240 male and female children to either a stereotype-congruent (stereotype threat), stereotype-non-congruent (no stereotype threat), or a control condition before asking them to complete the Child-IAT (Implicit Association Test) and an age-appropriate math test—lower IAT scores suggest a weaker implicit association between concepts such as stereotypes and their targeted group. In addition to the results demonstrating declined performance on a math test for female participants in the stereotype threat condition, analysis of implicit associations indicated that female participants under stereotype threat showed more negative associations toward their own ingroup (females).

A two-way ANOVA revealed a significant interaction between gender and condition, $F(2, 234) = 6.35$, $p < 0.05$ (Galdi et al., 2014). Further analyses of IAT score for male participants showed no significant effects of condition for male participants; that is, males participants in all three conditions did not produce significantly difference IAT

scores. Analyses of IAT for female participants, however, significant effects for condition, $F(2, 117) = 7.775$, $p < .01$. Specifically, female participants in the stereotype threat condition scored lower on the IAT than females in the no threat and control condition. Further, female participants in the no threat and control condition did not exhibit significantly different IAT scores.

Interestingly, when Galdi and colleagues (2014) conducted one-sample t-tests to test if IAT scores for participants in a specific condition were difference from zero, they found significant results for females in the stereotype threat ($p < .05$) and the control condition ($p < .05$) but not the no stereotype threat condition. This finding supports the researchers' assertion that implicit associations invoked by stereotype threat are potentially malleable. That is, exposing stereotype targets to incongruent stereotypes (e.g., girls are good at math) can avoid negative stereotype associations and reduce the effects of stereotype threat.

Similarly, across two studies, Schmader et al. (2015) operationalized stereotype threat, in part, as the "activation of negative implicit associations to the ingroup" using the IAT as a measure (p. 56) to examine how exposure to Latino stereotypes in film impact Latino viewers. In their first study, the researchers asked 111 Mexican American participants to watch a portion of a movie in either the realistic drama, unrealistic comedy, or a control group (no video presented) before providing evaluations of the movie clips, self-reports on group identity, self-esteem, and affective responses to the clips, and taking the IAT. In the realistic drama condition, Latino characters were a part of the main storyline and used legitimately in the develop of story (no stereotype threat).

In contrast, Latino characters is the unrealistic comedy served as a comedic device in which the character was objectified or degraded in a stereotyped method (stereotype threat condition). All participants view the movie clip alone in an isolated space. Perceptions of the video were measured using a 5-point scale asking participants to evaluate the stereotypically of the featured characters and how negativity of those stereotypes as well as the realistic and engagement attributes of the clip.

When asked about the stereotypical nature of the clips, participants in both film conditions perceived the clips were stereotypic of Latinos but participants who viewed the unrealistic comedy perceived the stereotypical portrayals of Latinos as more negative (M = 4.37, SD = .69) than the participants who viewed the realistic drama (M = 3.32, SD = .81), $t(74) = 6.10$, $p < .001$ (Schmader et al., 2015). Overall, the analysis of implicit ingroup attitudes revealed no significant effects among the film conditions; however, a hierarchal regression analysis integrating group identity and group pride revealed significant relationships between film condition and group identity. Specifically, a significant condition and identity importance interaction was found, $b = .45$, $p < .05$, and qualified by simple slope analyses the revealed a relationship between group identity and lower positive implicit associations toward the ingroup for participants that watched the unrealistic comedy, $b = .43$, $p < .05$, but not for participants who viewed the realistic drama. These results suggest that individuals who place value on their group identity will exhibit negative implicit associations toward their own ingroup when exposed to negative stereotypical displays.

In a second study, Schmader et al. (2015) examined how the presence of others (both ingroup and outgroup members) influence reactions to stereotypical portrayals. Eighty-five participants were asked to watch the unrealistic comedy movie clip from the first study and complete similar measures used in the first study. Unlike the first study, participants in the second study were told that a second person would be watching the clip at the same time from another and they could see that person on a video screen. The second person, however, was a prerecorded video of confederates, either Mexican American or White, that appeared to laugh at certain times during the movie clip. The analyses of IAT scored showed a main effect of race in which Mexican American participants exhibited more negative implicit associations toward their ingroup when paired with a Mexican American confederate than with a White confederate, $F(1, 84) = 5.34$, $p < .05$. Moreover, like the first study, the researchers found a significant interaction between ethnicity and group identity, $b = .24$, $p = .05$, in which the importance of one's group identity predicted more negative implicit associations toward their ingroup. Specifically, Mexican American participants who held their group identify higher exhibited more negative implicit attitudes toward Latinos when they were paired with a Mexican American confederate, $b = .53$, $p < .05$.

Cognitive Interference

Despite the plethora of research on stereotype threat, relatively little about the underlying cognitive processes and mechanisms between contextual framing and declines in task performance (Lambert et al., 2016). One of the commonly implicated mechanisms in the stereotype threat research is working memory. Working memory refers to an

executive cognitive function mediated by neural activity centered in the prefrontal cortex and used to process short-term tasks or goals such as performing immediate tasks; moreover, working memory capacity is limited such that cognitive resources dedicated to using working memory can be depleted by distraction or large cognitive loads. In the context of stereotype threat, individuals placed under threat (through contextual framing that cognitively activate relevant negative stereotypes) experience three potential processes—emotional regulation, self-monitoring, and physiological stress reactions—that limit working memory resources needed to successfully support the stereotyped task (Forbes & Leitner, 2014; Lambert et al., 2016; Popham & Hess, 2015).

Forbes and Leitner (2014) demonstrated how stereotype threat pulls neural resources away from focus on the stereotyped behavior. The researchers assigned 40 female participants to either a stereotype threat condition or a no-threat condition. In the stereotype threat condition participants were told they were taking a test of mathematics intelligence (diagnostic contextual frame) and asked participants to indicate their gender on a demographics form (stereotyped group salience contextual frame). In contrast, participants in the no-threat condition were informed they would be completing a problem-solving exercise and were not asked to identify their gender. Participants in both conditions completed the same math task while continuous EEG activity was recorded. Participants in the stereotype threat condition showed more increased neural activity in information and attention processing areas of the brain and performed worse on the math task than participants in the no-threat condition. Further, neural activity indicating the redirection of cognitive resources needed to perform was observed nearly instantaneously

after the introduction of the contextual frame. Their results suggest stereotype threat has an immediate impact on cognitive resources needed to perform tasks such as math test.

While working memory interference is a commonly cited mechanism in the stereotype threat model, personal factors such as emotional regulation may influence the relationship between stereotype threat and working memory outcomes. For example, to examine how the underlying mechanisms of stereotype threat differ across ages groups, Popham and Hess (2015) measured the working memory of young and older participants when placed under similar stereotype threat. The researchers asked 124 (63 over age 65 and 64 between 18 and 23) to complete several self-report measures including group identification, emotional regulation, and task difficulty as well was speed, accuracy, and working memory tasks. Older and younger participants were tested in separate sessions and assigned to either a negative stereotype or positive stereotype condition. In the positive stereotype condition, older participants were told they would be taking assessment in which older adults are expected to perform better. In contrast, older participants in the negative stereotype condition were told that younger adults are typically better at the assessments. Likewise, younger participants in the positive stereotype condition were told people who shared their same major in college do well on the assessments while younger participants in the negative stereotype condition were told the opposite.

As expected, the stereotype threat manipulation produced significant effects on the speed and accuracy performance of both younger and older participants (Popham & Hess, 2015). However, the impact of stereotype threat was greater for older participants

such as older participants in the negative stereotype condition were slower but made fewer mistakes. The researchers suggested that this result reflects older participants tendency to apply strategic attentiveness (accuracy over speed) when under stereotype threat. When working memory was analyzed, young participants in the negative stereotype condition showed lower working memory when compared to their positive stereotype condition counterparts, $F(1, 58) = 6.92$, $p < .01$. In contrast, no significant effect of stereotype condition was found for older participants. Further analyses found that older participants exhibit significantly higher emotional regulation when compared to younger participant, $F(1, 123) = 19.47$, $p < .001$, and significant effects of emotional regulation on stereotype condition. Younger participants who displayed lower emotional regulation produced lower working memory scores when under stereotype threat, $F(1, 58) = 11.08$, $p < .05$, when comparted to younger participants who displayed higher emotional regulation in the same condition, $F(1, 57) = 0.85$, $p = .77$. Older participants did not produce significant effects when emotional regulation was considered. Analyses that examined emotional regulation and speed and accuracy only found marginal significance for young participants and no effects for the older participants. While Popham and Hess, 2015 demonstrated that working memory can be impacted by stereotype threat, they also found that such outcomes may be moderated by personal factors such as emotional regulation ability.

The directionality of working memory in the stereotype threat model is unclear and working memory may moderate the effects of stereotype threat. That is, individuals with higher working memory capacity may be less susceptible to stereotype threat than

individuals with low working memory capacity. Lambert et al. (2016) investigated the moderation of working memory in older adults under stereotype threat by randomly assigning 39 participants (age 62 and older) to either a stereotype threat or no threat condition before administering a driving simulation. Participants in the stereotype threat condition were informed that the intent of the study was to examine the stereotype that older people are bad drivers; moreover, participants in this condition were presented with materials containing negative examples of elderly drivers (e.g., crash statistics of older drivers). Participants in the no threat condition were only informed that driving data was being collected. In a first session, participants were exposed to the stereotype threat manipulation and asked to complete the driving simulation course in which their brake reaction time and following distance was recorded. In a second session, participants were administered a working memory capacity assessment.

The results of a hierarchal regression model and part-partial correlations indicated significant effects of working memory and stereotype threat on driving performance (Lambert et al., 2016). Specifically, a significant negative relationship between working memory capacity and brake reaction time was found in the stereotype threat condition, $r(18) = -.62$, $p < .01$, but not in the no threat condition. Similar results were found when analyzing following distance performance—a significant negative relationship between working memory capacity and following distance appeared in the stereotype threat group, $r(18) = -.53$, $p < .05$, but not for the no threat condition. Taken together, these results suggest that participants with lower working memory capacity prior to being placed under stereotype threat performed worse when compared to participants with higher

working memory capacity. Moreover, while this finding further supports that working memory plays a vital role in the stereotype threat model, it reflects a deeper complexity of that role in which preexisting levels of working memory capacity may determine the extent of the decline of working memory ability while under stereotype threat.

Behavioral Outcomes

Stereotype threat impacts short- and long-term behavioral outcomes. Short-term behavioral outcomes are the behaviors specific to the nature of the stereotype within the immediate stereotyped situation such as academic performance when taking a test. Steele and Aronson (1995) first demonstrated how eliciting race-based stereotype threat decreases performance in stereotype-relevant behaviors or tasks. Since the initial Steele and Aronson studies, academic or intellectual performance has been among the most studied behavioral outcomes in the stereotype threat literature. The previous sections provided several examples of behavioral outcomes influenced by stereotype threat; for example, John-Henderson et al. (2014), found that invoking SES-based stereotype threat decreased academic test performance in college students who recently experience low SES. Riciputi and Erdal (2017) showed that student athletes performed worse on a math test when their *jock* identity was made salient. Galdi et al. (2014) and Pacilli et al. (2016) demonstrated that stereotype related images decrease math performance in female children.

Stereotype threat, however, is not limited to the academic performance domain. In fact, stereotype threat may impact behavioral outcomes or task performance in any stereotyped domain. Older drivers, for instance, may experience stereotype threat when

presented with stereotypes that elderly people are poor drivers (Lambert et al., 2016). White males and females may be influenced by stereotype threat in athlete performance domains (Hakim & Quartiroli, 2016; Stone et al., 1997). The behaviors of Black Americans under stereotype threat may be impacted in criminal justice interactions (Najdowski et al., 2015). Health-related behaviors among stereotype threaten obese people may also be affected (Brochu & Dovidio, 2014).

Long-term behavioral outcomes affected by stereotype threat include disengagement and or domain avoidance. Repeated experiences with stereotype threat can condition individual to avoid similar evaluative situations by disengaging with the stereotyped domain. For instance, Black American students, under stereotype threat, may avoid academic challenges or pursuits (Steele & Aronson, 1995; Tyler et al., 2016). Disengagement also involves self-handicapping approaches. Stereotype threatened individuals may make excuses about their performance such as “I didn’t prepare” for the task. This self-handicapping allows individuals to rationalized underperformance in the domain while avoiding being labeled as unable. Disengagement from a stereotyped domain does not allow the individual to improve in the domain and undermines long-term outcomes (Flanagan, 2015; Silverman & Cohen, 2014).

The stereotype threat model has empirically demonstrated exposure to subtle contextual frames initiates cognitive processes that influence behavioral outcomes across various social groups and behavioral domains; therefore, the stereotype threat model is ideal for the current study. Moreover, decades of empirical research examining stereotype threat offers the current study guidance on the necessary components of this model that

will facilitate its application, including, but not limited to (a) contextual framing, (b) stereotype activation, (c) potential cognitive processes such as decreases in short-term memory and inattention and (d) short- and long-term changes in the stereotyped-domain behaviors such as decreased task performance and avoidance of academic activities.

Literature Review Related to Key Variables

In this study, I used the basic stereotype threat model serves as the theoretical framework to understand the effects of appropriated slurs on ingroup members. That is, in the context of this study, appropriated slurs will act as the contextual framing cue that potentially activates the negative stereotypes connected to the root slur (i.e., the original slur before being adapted by the target group) reducing cognitive resources needed to perform stereotype-related task. In the following sections, I will discuss the key variables including appropriated slurs, stereotype activation, and stereotype-related task performance in the context of the current study.

Appropriated Slurs

Research on appropriated slurs has been mostly addressed by researchers in the communication, linguistics, and sociology disciplines—often focusing on the linguistic function or intentions of such terms. Further, the psychological investigation of appropriated slurs has been limited to qualitative studies and self-report data with no known examination of the cognitive-behavior effects of appropriated slurs. However, psychological science does provide empirical evidence on the cognitive effects of slurs (in their original form) that informed the current study. Therefore, I begin the following section with a discussion on slurs and their use before discussing appropriated slurs.

Slurs

In natural language, words may have various functions that support effective communication including being descriptive or expressive (Croom, 2016). Descriptive terms are used to neutrally identify characteristics or categorical features of an object or target. In the context of social groups, for example, descriptive terms such as *Black American* or *African American* are used to indicate membership to a specific racial group without attached evaluative or emotional meaning. Expressive terms are used to convey emotion toward an object or target but may not be descriptive; for example, expressive phrases such as [target] is *fucker* or *that goddamn car won't start* demonstrate emotions such as disgust or frustration without being directly descriptive of the target. As a component of language, slurs function as both descriptive and expressive terms that identity some characteristic of the target (e.g., race) as well as demonstrate negative emotion or evaluation toward the target (Archer, 2015; Croom, 2015). Moreover, the specific referential and expressive nature of slurs distinguish them from other pejorative terms because they are linked to underlying stereotypes of the target. For example, the slur *nigger* calls upon the underlying stereotypes attributing negative attributes such as unintelligence and laziness to Black Americans. The link between slurs and their underlying stereotypes is also tightly fastened to the evaluation of the target group such that slurs not only express the speaker's negative evaluation of the target but can implicitly influence others to negatively evaluate the target group (Fasoli et al., 2016; Soral et al., 2018). The stereotype-based nature of slurs allows slur speakers to attempt social control or oppression of the target in which the slur acts as a symbolic term that

communicates the inferiority of the target and implies their exclusion or restriction from resources or opportunity. Slurs can produce emotional or psychological distress in target group members, and such words are often viewed as hostile or offensive (Spotorno & Bianchi, 2015). Moreover, slurs may activate the negative stereotypes—cognitive schemata representing beliefs about the attributes of the target group—related to the slur when heard by both target group members and others (Burkley et al., 2016; Jeshion, 2013a).

Slurs, however, are seemingly contextually flexible and can be adapted to an array of motivation; that is, slurs may be applied in various situations with quite different goals and outcomes. For example, an angry speaker may use a slur to display frustration and contempt toward someone in traffic while the same slur with slight variations maybe be used between companions during friendly banter. Archer (2015) put forth the Facework Scale to demonstrate how language can operate along a continuum of motivation and contextual applications. The Facework Scale assumes that face-enhancing and face-aggravating language fall at opposite ends of the same continuum that represents the speaker's evaluation and intent toward the target and the target's understanding of the speaker's evaluation and intent. On the face-aggravating end of the scale, slurs are intentionally employed by a speaker to derogate and attack the target with clear intention as in the example of the angry motorist using a slur toward another commuter. On the face-enhancing end of the scale, Archer suggests that slurs can be used within specific context and under clear social rules to demonstrate friendship and positive social connection—such as two friends using slurs for friendly banter between each other.

Appropriating Slurs

The appropriation of slurs is a complex socio-psychological process through which the ingroup develops the rules of how a slur is adopted and its appropriated use (Bianchi, 2014). Slur appropriation may be viewed as an ingroup taking control of the oppressive power and stigmatization connected with the original slur; for example, Black Americans have appropriated the original slur *nigger* as a multi-use ingroup term with specific social and cultural norms regulating its use. In part, these norms include word ending variations such as —*a* replacing the original –*er* and the limiting the use of the word to ingroup members (Croom, 2013; Gaucher et al., 2015). The appropriated slur *nigga* morphed into a contextually flexible ingroup term through which proponents express ingroup solidarity and empowerment. However, Black American opponents of this appropriated slur hold that the original negative connotations can never be removed from the word and even its appropriated forms are harmful to the psyche of the Black American community. In other words, some hold that appropriating slurs does not empower one's group but sustains the harmful effects of the original slur.

Some theoretical and empirical work supports the possibility that appropriating slurs diminish the effects of the original slur and empower the targeted group. Archer's (2015) Facework Scale, for instance, proposes that slurs can be contextually flexibly depending on the social situation and the intentions of the person using the slur. This contextual flexibility of slurs underlies slur appropriation—the refurbishing and adoption of slurs by target group members such that the appropriated slurs become non-offensive and even endearing ingroup terms with specific social rules and applications (Croom,

2013; Galinsky et al., 2013). Seminal research by Galinsky, Hugenberg, Groom, and Bodenhausen (2003) offered a three-level (individual, intragroup, and intergroup levels) model of appropriation that considers the conditions and consequences of slur appropriation. At level one, a member of the targeted group decides independently of the target group to self-label using the slur. At level two, the targeted group collectively decides to self-label using the slur. The conditions of level two appropriation include group cohesion, collective self-esteem, and current social status mobility. Level three of the model represents intergroup acceptance of the appropriated slur. That is, the targeted group as well as the slur's originating group has accepted the slurs new connotations and devaluation of the previous meanings. The conditions of level three require the target group to have a sufficiently increased or increasing social status, ambivalent attitudes toward the targeted group from other groups, and successful social competition.

Reflecting parts of the appropriation model and Facework Scale, recent empirical research has examined the contextual flexibility and perceptions of using appropriated slurs (Galinsky et al., 2013; Gaucher et al., 2015). Across ten experiments, Galinsky et al. (2013) tested their model of appropriation by measuring perceptions of power related to self-labeling behaviors (i.e., using an appropriated slur). Users of appropriated slurs reported perceptions of more social power. Similarly, outgroup observers perceived appropriated slur users and their stereotyped group as being more socially powerful. Overall, these results support the model of appropriation previously discussed and the premise that self-labeling with an appropriated slur may be linked to the perceptions of social power. Similarly, Gaucher et al. (2015) tested if appropriated slurs (specifically

slut) can produce empowerment and reduce stereotype endorsement related to rape myths in women. They assigned female participants to either a supportive or typical contextual condition in which a male or female character in the vignette shouted the slur. The supportive contextual condition vignette was set in a social justice march about gender inequality and the typical contextual vignette was nondescript in its setting. After reading the vignette, participants completed measurements of emotional reactions, empowerment, self-objectification, rape myth endorsements, and how negative they perceived the slur.

Compared to participants in the typical context condition, participants who read the supportive context vignette reported higher emotional responses related to self-assurance and lower responses related to fear as well as greater feeling of empowerment (Gaucher et al., 2015). Moreover, participants in the supportive context condition were less likely to endorse common rape myths. However, participants in both contextual conditions indicated that the appropriated slur was a negative term. Interestingly, the gender of the character uttering the slur did not affect outcome measure in the supportive context condition but increased more negative responses in the typical condition when the character was a male.

In a second study, Gaucher et al. (2015) replicated their previous study but added a second social justice condition not related to gender inequality and a control condition that did not require participants to read a vignette. In contrast to their initial findings, woman in both social justice conditions show lower feeling of empowerment compared to the control group. However, participants in the social justice condition were significantly less likely to endorse rape myth stereotypes. Moreover, empowerment did

not differ between the social justice conditions suggesting that the presence of the appropriated slur did not produce negative reactions in the participants. There were no significant differences in emotional reactions in the second study. The results of these two studies suggest that exposure to an appropriated slur (e.g., *slut*) in a supportive environment reduced stereotype endorsement related to rape myths such as the culpability of the victim. Moreover, feelings of empowerment were decreased when social justice was salient, but the presence of the appropriated slur did not decrease empowerment in this context—suggesting that appropriated slurs do not impact positive social environments.

While studies such as Galinsky et al. (2013) and Gaucher et al. (2015) support the perspective that appropriated slurs empower target group members and neutralize the negativity of the original slur, such research relied on self-report measures of perceived power and social status but did not examine the cognitive effects (e.g., stereotype activation, working memory, attention, etc.) and behavioral outcomes of appropriated slurs.

Stereotype Activation and Implicit Association

Stereotype activation is the processes through which cognitive schemata are pulled into the reach of current cognitive processing and use to inform judgements or actions (Wang et al., 2017). The use of cognitive schemata (stereotypes) contributes to cognitive efficiency and quicker actions and may influence a wide range of attitudes and behaviors; moreover, stereotype activation is relatively automatic or unintentional processes that remains outside of the awareness of those under activation. Neuroscience

research has also supported the automatic nature of stereotype activation by demonstrating neural reactions to stereotype priming and reaction to stereotyped targets (Wang et al., 2016). Stereotypes can be activated through various subtle or direct stimuli that contribute to the way stereotypes manifest through behavior.

Stereotype activation research has broadly explored how the activation of stereotypes influence non-target (i.e., those whose group not targeted by a stereotype) attitudes and behaviors. For instance, individuals exposed to ethnic branding or advertisement exhibited strengthened implicit associations (e.g., Native Americans are warlike) but weakened explicit stereotypes toward the characterized group especially for individuals expected to have higher levels of mental flexibility (Angle et al., 2017). Similarly, individuals primed with alcohol advertisements show higher levels of implicit bias toward Black Americans; however, this implicit stereotype activation was not replicated in subsequent studies due to the potential waning effects of the nature of some types of stereotype priming (alcohol advertisements) cues (Stepanova et al., 2018). Negative implicit stereotype activation also influences how individuals evaluate music from various genres. For example, individuals primed with negative stereotypes about hip-hop or heavy metal music before listening to those types of music were more likely to negatively evaluate the music than individuals not primed (Neguţ & Sârbescu, 2014). While there may have been underlying social groups attached to music genre stereotypes, the evaluation of those connected social groups was not directly studied.

The demonstration of how stereotype activation influences attitudes and behaviors has also been extended to stereotype activation of stereotypes relevant to one's self

concept. Several lines of research have examined how stereotype activation of stereotypes related to social groups to which a person belongs (or assigned) influence that person. Women primed with male-dominate stereotypes, for example, tend to evaluate stereotype-typical male careers such as entrepreneurism and science, technical, engineering, and mathematic fields as less attainable for themselves (Gupta et al., 2013; Schuster & Martiny, 2017). Additionally, individuals primed with stereotypes of higher performing occupations (technical specialist or athlete) tended to perform better on related tasks than individuals primed with lower performing occupations (Wang et al., 2017). Women who referenced stereotypical characteristics as reasons for low performance activate stereotypes in others and, in turn, increased negative evaluations and stereotype endorsement of women's abilities (Burkley et al., 2016).

An organic overlap between stereotype activation research and stereotype threat research is evident as both lines of study examine how stereotypes influence behaviors. Steele and Aronson's (1995) pioneering work in stereotype threat found that individuals under stereotype threat completed more race-related word fragments suggesting that contextual framing that invokes situationally relevant stereotypes. Such findings suggest that contextual framing implicitly activates the relevant stereotypes (i.e., race or gender stereotypes) in individuals that belong to the targeted group. Contemporary research continues to examine the role of stereotype activation in the context of the stereotype threat model. For example, female children exposed to contextual frames related to math performance exhibited negative implicit stereotype associations between females and math as well as decreased task performance (Galdi et al., 2014). Similarly, Mexican

Americans who watched films with stereotypical portrayals of their racial group produced negative implicit stereotype associations toward Mexican Americans (Schmader et al., 2015).

While stereotype activation seems to be the link between contextual framing and behavioral outcomes, several factors such as social identities and stereotype valence moderate this link and impact behavioral outcomes. Exhibiting higher implicit associations towards one's own social group when exposed to contextual frames may depend on how strongly a person identifies with that group. Group identification is the degree to which an individual self-subscribes to a specific social group and the self-worth or values they place on membership in that group (McKinley et al., 2014). While early research implicated group identity as a buffer against stereotype threat, relatively recent studies posited that racial identity may buffer or increase the effects of stereotype threat depending on interpretation and valence of group membership.

Schmader et al. (2015), for instance, found that Mexican Americans whose ethnic identity was stronger tended to demonstrate more negative implicit stereotype associations toward their own group when under stereotype threat. Domain identification—the value upon which a person places on his or her connection to a specific performance area (e.g., math, sports, music, etc.) has also been implicated as a moderator of stereotype threat. Typically, individuals who highly identify with a stereotyped domain are more susceptible to stereotype threat. That is, individuals who place more self-worth or value on a specific domain (e.g., academics, mathematics, sports) are more likely to be threatened by relevant negative stereotypes related to that

domain. For example, when female college students highly identify as a mathematic major, they are more impacted by stereotype threat and exhibit lower math test scores compared to female student who do not identify as mathematicians (Deviyanti, 2015).

Stereotype valence (positive or negative context) seems to impact the automatic associations when stereotypes are activated. When individuals with higher levels of mental flexibility were exposed to positive ethnic branding images, their positive implicit associations were strengthened (Angle et al., 2017). In classical stereotype threat research, the subtle suggestion of negative stereotypes invokes stereotype threat and hinders stereotype-related performance. It stands to reason that if contextual framing includes a positive stereotype then stereotype threat could be alleviated or potentially reversed creating improvements in stereotyped domain performance. Interestingly, attempting to control the valence (positive or negative context) of stereotypes may help increase performance but only when an individual strongly identifies with the relevant domain (Saad et al., 2015). When examining the impact of instructor gender on college-level math students, Kapitanoff and Pandey (2017) found that female students who endorsed gender-math stereotypes (i.e., women cannot do math) initially performed worse when assigned a female instructor when compared to female students who did not endorse the gender-math stereotypes. That is, when confronted with a positive role model or stereotype (women can be successful in math careers) some women perform worse rather than better.

Kapitanoff and Pandey (2017) speculated that this observation may have be caused by an upward social comparison which caused performance anxiety; however, a

similar phenomenon was observed in a study on domain identification in Chinese American women and math performance (Saad et al., 2015). Saad et al. assigned 119 Chinse American females to either a positive stereotype condition or a control condition before administering a math test and domain (math) identity questions. In the positive stereotype condition, the researchers attempted to subtly activate positive stereotypes about Asian Americans by asking participants to describe their family lineage and traditions. Participants in the positive stereotype condition and who identified more with the math domain performed better on the math test than participants in the control condition; further, participants in the positive stereotype condition who identified less with the math domain performance worse than those in the control condition.

Racial Identity

The stereotype threat model implicates stereotype activation as the link between contextual framing and behavioral outcomes; that is, an environmental cue related to a negative stereotype will activate that stereotype in an individual's cognitive processes leading to decreased performance in the stereotyped domain. However, the relationship between these components of the stereotype threat model may be moderated by other factors such as racial identity. Racial identity is the extent to which an individual self-subscribes to a specific racial group and the self-worth or personal value they place on membership in that group (McKinley et al., 2014).

Some early research suggested that group identity (racial, gender, etc.) acts as buffer against stereotype threat (for example, Davis, Aronson, & Salinas, 2006). More recent research, however, reveals that racial identity may either buffer against or

exacerbate the effects of stereotype threat depending on the nature of such group membership. During their examination of how stereotype threat varies based on the influences of school and home contexts, Massey and Owens (2014) expected that light skinned compared to dark skinned Black Americans to be more susceptible to the effects of stereotype threat due to lower confidence in their racial identities. Across 28 universities, they administered a survey to Black American Freshmen (N = 918) about various aspects of their background including high school integration, parental race and ethnicity, and skin tone (individual variables). Further, they assessed each of the universities in the studies for diversity and inclusion characteristics such as minority representation, selectivity, and affirmative action policies and behaviors (contextual variables). Focusing on GPA over time as the major dependent variable, the researchers conducted a latent variable analysis on the individual and contextual variables for each participant. Their results supported their predictions that light skinned students were more vulnerable to stereotype threat compared to darker skinned students. While these results assume skin tone as a proxy for racial identity, Massey and Owens (2014) did not directly measure the participants' racial identity.

With a more explicit measure of racial identity, Schmader et al. (2015) examined how various aspects of social identity such as centrality and pride moderate the effects of stereotype threat. The researchers differentiated these two concepts by noting that centrality is the "importance of social identity might signal the perceptual and affective salience of identity relevant cues" and pride is the level positive attitude toward one's ingroup (p. 57). Mexican American participants were randomly assigned to one of three

conditions in which they either watched a realistic drama or an unrealistic (i.e., exaggerated stereotypic behaviors) comedy with Mexican American actors, or no video at all. Participants then completed surveys about their perceptions of the video (e.g., engaging and entertaining), racial/ethnic identification, self-esteem measures, affective responses, and the implicit attitudes test toward Latinos.

Schmader and colleagues' (2015) initially conducted a series of one-way ANOVAs for each dependent variable of interest. Their results showed no group differences in affect between participants who viewed the dramatic video compared to those who viewed the comedic video but participants who viewed either video showed significantly fewer positive emotions and more shame and anger when compared to the control group. No significant results were found between conditions when they looked at self-esteem or implicit group attitudes; however, when the researchers tested the moderating effects of racial identity (centrality and pride), they found several clear interactions in participants that viewed either video.

Schmader et al. (2015) conducted a series of two-step hierarchal regression analyses in which video type (drama or comedy), racial centrality, and group pride were loaded at step one and the interactions of video type and racial centrality and video type and group pride were loaded in step two. Notably, their results revealed a significant interaction between video type and centrality that predicted less positive affect in the comedic (with blatant stereotypes) video condition; moreover, a similar pattern was reveal in the implicit associations model which indicated centrality predicted less positive implicit attitudes towards the participants' ingroup in the comedic video condition. That

is, participants with stronger connections to their racial identity were more likely to experience negative emotions and negative attitudes toward their own group when exposed to explicit stereotypic contextual cues.

Similar relationships between stereotype threat and racial identity were found when Shelvin et al. (2014) examined the influence of racial identity as a moderator of stereotype threat among Black American children between the ages 10 and 12 using an adaptation of the MIBI—the MIBI-t (Multidimensional Inventory of Black Identity-Teen Version). The researchers identified six distinct clusters— ethnically diffuse, high connected, idealized, low connected, assimilationist, and marginalized—of racial identity among their participants (N = 186) using the MIBI-t results. The examination of the interaction between stereotype threat and racial identity (clustered groups) in a 2x6 ANOVA (threat condition x MIBI-t cluster) with an academic test as the dependent variable revealed a main effect of racial identity, $F(5, 140) = 4.31$, $p = .001$, no main effect for stereotype threat condition, $F(1, 140) = 2.66$, $p = .11$, on academic performance, and a significant interaction between stereotype threat and cluster profile, $F(5, 140) = 3.05$, $p = .01$. Only, two MIBI-t clusters showed a significant result from a simple effects analysis—the ethnically diffuse profile, $F(1, 140) = 4.97$, $p = .03$, and the high connected profile, $F(1, 140) = 8.81$, $p = .004$.

The ethnically diffuse cluster was described as individuals who have neither a strong connection to Black American culture nor the nonminority racial culture; moreover, this cluster showed no motivations to assimilate into the majority culture (Shelvin et al., 2014). This cluster was identified by the lower mean for the private regard

subscale coupled with mean in the bottom third for the other MIBI-t subscales. Participants categorized as ethnically diffuse performed worse in the stereotype threat condition (M = 10.12) compared to the ethnically diffused individuals in the no-threat condition (M = 13.54). The high connected cluster contained participants who produced the greatest mean scores on each of the centrality, private regard, oppressed minority, and nationalist subscales. The researchers described this group as individuals who place importance on their Black American identity and understand the advantage of coordination with other minority groups in the face of oppression. Participants in the high connected profile performed worse under stereotype threat (M = 13.10) than their no-threat condition counterparts (M = 18.07). The researchers suggested that high connected individuals are more susceptible to stereotype threat because they are more concerned about negative stereotypes related to race.

The strength and nature of one's racial identity moderates the outcomes of stereotype threat (Schmader et al., 2015; Shelvin et al., 2014). Individuals who feel a strong connection to and place a high personal value on membership in their racial group tend to be more susceptible to the effects of stereotype threat. This vulnerability may stem from heightened reactivity to stereotype-related contextual cues and concerns about being stereotyped in specific domains. Interestingly, individuals who may be less sure about their racial identity may also be more vulnerable to the effects of stereotype threat. For instance, the impact of stereotype threat on Shelvin et al.'s (2014) Ethnically Diffuse cluster parallels Massey and Owens' (2014) conclusions that individuals who are not

connected to either minority or non-minority racial identities may be more susceptible to stereotype threat.

Domain Task Performance

As discussed in previous sections, has across several decades, stereotype threat researchers examined a wide variety of behavioral domains including but not limited to athletic performance, driving ability, and communication skills; however, academic performance is among the most commonly studied domains in stereotype threat research. The importance of studying academic performance through lens of the stereotype threat model can linked to the persistent academic gap between Black Americans and their White counterparts in the United States. Reading and mathematics achievement for children in the United States has steadily increase for nearly five decades; nevertheless, Black American children tended to score lower that their white counterparts even across other demographic areas such as socioeconomic class (*National Center for Education Statistics*, 2014). Similar gaps can be observed in higher education as well as across various occupational areas (Casad et al., 2017; Emerson & Murphy, 2014; Kalokerinos et al., 2014). Stereotype threat researchers aim to reduce such inequities by better understanding how stereotype threat impacts behavior and how to reduce negative outcomes.

In attempt to explore stereotype threat as an element of academic gaps, researchers have utilized various measures of academic performance that potentially reflect a student's real-life ability. Steele and Aronson (1995) set the standard for using academic performance measures in stereotype threat research in their initial research by

using a shortened variation of the GRE to measure performance in an academic setting (Steele & Aronson, 1995). Subsequent stereotype threat studies have included similar academic test instruments (e.g., Schuster et al., 2015). Standardized academic tests, however, are not the only form of task performance used in stereotype threat research. Some researchers developed unique tests for their research; for instance, Forbes and Leitner (2014) administered a timed 50-item multiple-choice multiplication test and (Galdi et al., 2014) presented a brief addition and subtraction test to participants to analyze academic performance under stereotype threat.

Despite the type or variation of academic test used, stereotype threat researchers have discovered that the difficulty of an academic test mediates outcomes for individuals under stereotype threat. Overall, researchers demonstrated that the stereotyped domain behavior must be moderately difficult to be affected by stereotype threat (Forbes & Leitner, 2014). Tasks that are too easy or have been well-learned by an individual do not seem to be impacted when an individual is place under stereotype threat. However, more difficult tasks that require more executive function resources such as attention and working memory are more susceptible to stereotype threat. Further, while single academic tasks such as a test or quiz have proven to be an adequate reflection of an individual's academic ability, such tests may not reflect long-term or sustained performance. That is, one academic test may not reflect how well or how poorly an individual performs overtime. A common approach to address the validity of an academic test measure in stereotype threat research is to use a more stable or consistence attribute of academic ability such as the SAT or GPA as a covariate (e.g., Steele & Aronson, 1995;

Tyler et al., 2016). Another study examined test performance over the course of a college semester finding that individuals under stereotype threat showed lower test performance over time when compared to individuals not under stereotype threat (Kapitanoff & Pandey, 2017).

Summary and Conclusion

In general, slurs are descriptive and emotionally charged words used to derogate, insult, and social oppress target groups (Archer, 2015; Croom, 2015). Slurs elicit negative emotions and thoughts in both the targets of the slurs and bystanders (Fasoli et al., 2016; Soral et al., 2018; Spotorno & Bianchi, 2015). Despite the ability of slurs to affect others, nearly all social groups targeted by slurs have appropriated those slurs and adapted them as ingroup terms with specific contextual rules to exert empowerment and control over the original term. However, members of social groups that employ appropriated slurs debate the true value of appropriated slurs. On one side of the debate ingroup members support the empowerment and control perspective while others posit any form of the original slur has negative effects on the group. Some researchers support the view that appropriated slurs are empowering and devalue the negative value of the original slur (e.g., Galinsky et al., 2013); however, no research has examined how appropriated slurs impact implicit cognition and subsequent behaviors. In this study, I addressed these gaps by examining the differences in cognitive (stereotype activation) and behavioral (academic task performance) outcomes when individuals are exposed to appropriated slurs. The results of the current study contribute to the knowledge on the debate (beneficial or harmful to the ingroup) on the value of appropriated slurs as well as expand

the understanding of the influence of contextual frames on stereotype threat and performance gaps.

To examine the potential effects of appropriated slurs on cognitive and behavior, I utilized the stereotype threat framework in this study. Stereotype threat is a process through which environment cues related to negative stereotypes about one's social group invoke cognitive interference leading to decreased performance in a stereotype-relevant task (Steele & Aronson, 1995). The basic components of the stereotype threat model are contextual frame, stereotype activation, cognitive interference, and task performance (Forbes & Leitner, 2014; Galdi et al., 2014; Schuster & Martiny, 2017). Considering the persistent academic performance gap faced by Black Americans, an abundance of stereotype threat research has focused on academic performance in Black Americans. Such research has found that subtle contextual cues, (frames) such as the diagnosticity of a test or the race of the instructor, activates the negative stereotypes placed on Black Americans, which reduced cognitive resources and diminished academic task performance (Shelvin et al., 2014; Steele & Aronson, 1995; Tyler et al., 2016). A potential source of an ingroup contextual frame may come in the form of appropriated slurs; therefore, in the current study, I utilized appropriated slurs as a contextual frame within the stereotype threat model. This joint examination of appropriated slurs and ingroup-based contextual frames contributes to the knowledge of the cognitive and behavior impact of exposure to appropriated slurs as well as deepen the current understanding of factors that contribute to invoking stereotype threat in Black Americans.

In the next chapter, Chapter 3, I present the rationale and methodology for this study. In part, the upcoming discussion will include operational definitions of the key variables, the data collection method and related instruments, analysis plan, and potential threats to the validity of the study. Further, the target population and sampling approach as well as the recruitment, participation, and ethnic procedures will be discussed.

Chapter 3: Research Method

Introduction

The purpose of this study was to examine how exposure to appropriated slurs influence stereotype threat. To this end, I employed a quantitative approach modeled after common research paradigms found in the stereotype threat literature. Specifically, in a randomized experiment with posttest-only with control group design, I administered two performance tasks and a survey to examine the relationship between exposure to appropriated slurs and the major outcome components of the stereotype threat paradigm (stereotype activation and stereotyped behavior outcomes) as well as the moderating effect of racial identity among Black American adults.

In this chapter, I present an overview of the research design. I first discuss the rationale for the study before presenting the target population and sampling method as well as recruitment procedures. Then, I discuss the data collection method and analysis plan in which I outline the operational definitions of the variables of interest, the instruments they were measured by, and the statistical tests that were applied to each research question as well as threats to validity. Finally, I discuss possible ethical considerations of the current study.

Research Design and Rationale

In the current study, I used a quantitative approach modeled after stereotype threat research designs found in the literature (e.g., Kellow & Jones, 2008; Pacilli et al., 2016; Steele & Aronson, 1995). Specifically, I used a randomized experiment with a posttest-only with control group design to examine how exposure to appropriated slurs affects

stereotype activation and stereotyped-domain task performance (academic test performance). A randomized design reduces the impact of confounding factors and decreases potential threats to internal validity (Shadish et al., 2002). Further, I selected the posttest-only design to avoid priming or sensitizing participants to the academic test; however, I attempted to establish a baseline of academic test performance by asking participants to respond to a Likert-type question ("I always did well on tests in school") after the academic test. This approach to establishing a baseline in the absence of a pretest measure has been effectively applied in various stereotype threat studies (e.g., Steele & Aronson, 1995). Moreover, the addition of a control group to the posttest design provides potential evidence for counterinference; the control group's behavior may be assumed to reflect accepted behavior in the absence of the dependent variable.

With the understanding that many adults have busy schedules, and priority is given to work, school, and family, I designed the current study in such a manner that participation time was less than 30 minutes. Further, I implemented an online participant interface that provided additional convenience and flexibility to participants; however, this online format may have introduced additional variation in the data due to possible environmental distractions, variations in the participants' computers and internet connections, and individual integrity during the testing sections of the study.

Methodology

Population

While nearly every contemporary minority group has employed appropriated slurs, I focused on Black American adults. According to the American Fact Finder (U.S.

Census Bureau, 2015), Black Americans make up approximately 13% of the population in the United States. Adults (age 18 and older) make up 73.8% of the Black American population, and 46.7% and 53.3% of those adults are female and male, respectively. While slurs and appropriated slurs are not limited to a single social group, I decided to focus on the Black American population for three reasons. First, Black Americans have been one of most targeted racial groups across the history of the United States, which may have produced the most deep-rooted set of racial stereotypes in society (Croom, 2015). Second, the appropriated slurs used in the Black American communities are among the most prominent examples of appropriated slurs and are more likely to be within potential participants' awareness—a requirement of stereotype activation. Third, the debate over the use of appropriated slurs in the Black American community has been well-documented from qualitative perspectives (Allan, 2015; Galinsky et al., 2013; Rahman, 2011).

Sampling and Sampling Procedures

I employed a convenience sampling approach recruiting from the available pools of sampling units such as the Walden University Participant Pool, Amazon Turk, and social media recruitment. The convenience sampling strategy was selected because of its ease of reaching potential participants in a relatively brief timeframe with minimal resources. A sample will ideally be representative of the population of interest; however, convenience sampling limits the ability to recruit participants who fully represent the target population (Frankfort-Nachmias & Nachmias, 2015). In the current study, for example, the ideal sample would not only match individuals who identify as Black

Americans but would also match the representation of those individuals across age, gender, socioeconomic status, education, etc., in the United States. The convenience sampling method was not likely to draw a sample that reflects those variations across the Black American population; nevertheless, this limitation could be addressed in the statistical treatment of the collected data if appropriate.

Sampling Frame

Establishing a sampling frame, a complete list of possible sampling units, is important to identify participants who meet the criteria of the population of interest. The sampling frame provides a method to ensure that sampling units drawn (those who participate) accurately represent the population of interest at the level they are meant to be studied (Frankfort-Nachmias & Nachmias, 2015). In consideration of the relatively broad population of interest in the current study, an exact sampling frame was not possible to establish. Nevertheless, the basic inclusion criteria for this study were (a) individuals who identify as Black Americans and (b) are 18 years of age or older.

In the context of this study, I defined the social label *Black American* or *African American* as individuals born and raised in the United States who identify with an African ancestry. This specificity of the inclusion criteria was to increase the likelihood that participants would be aware of both the original and appropriated versions of the slur (*nigger* and *nigga*, respectively) and the social rules and applications of both terms. These terms are socially ingrained into the culture of the United States through social use (within and between groups) and media (film, television, and music); individuals born and raised in the United States are most likely to be familiar with these terms (Croom,

2016; Gaucher et al., 2015). This awareness is vital in the stereotype threat model (Doyle & Voyer, 2016; Shelvin et al., 2014; Steele & Aronson, 1995). Moreover, exclusion of other racial groups was intended to focus the study on the selected appropriated slur and the population (Black Americans) that has the most personal connection to the underlying slurs to potentially invoke the cognitive interference aspect of stereotype threat. For instance, Asian American participants would not likely be negatively impacted by underlying stereotypes related to the appropriated slur *nigga*.

I limited the age criterion to adults for both logistical and ethical concerns. For instance, the videos in which the appropriated slurs appeared may not be suitable for children. Moreover, targeting a specific age range among adults would have not been practical for this study due to the possible variation in age of respondents.

Sample Size and Power Analysis

In the context of quantitative research, statistical power supports the validity of inferences drawn from the data. Specifically, statistical power is the probability of committing a Type II error based on the data being analyzed (Ott & Longnecker, 2015). I established the statistical power for each of the planned data analyses in the current study using G*Power 3.1 software. Moreover, I entered the standard parameters for statistical power ($\beta = .80$), alpha ($\alpha = .05$).

I addressed the research questions using two 2x2 ANOVA models to determine significant group differences in the mean outcomes of academic performance task and stereotype activation task scores. Prior to data collection, I derived a sample size for the ANOVA analyses through G*Power 3.1 by entering parameters for a fixed effects,

special main effects, and interactions ANOVA under the F-test family group. In addition to the standard levels of statistical power and alpha described above, I selected a medium effect size (0.25) with the numerator degrees of freedom (df) as 1 (the product of the number of levels in each condition minus 1) and the number of groups as 4, representing the number of conditions: (a) exposed to appropriated slur and low racial identity, (b) exposed to appropriated slur and high racial identity, (c) not exposed to appropriated slurs and low racial identity, and (d) not exposed to appropriated slurs and high racial identity. I based the selection of the effect size on stereotype threat research showing similar effect sizes (e.g., Lamont et al., 2015; Oliveira & Cabral-Cardoso, 2017). The results of the power analysis indicated a total sample size of 128 (32 participants per condition) at a critical F of 3.91.

Recruitment, Participation, and Data Collection

Recruitment

I recruited participants using the Walden University participant pool portal, Facebook, Twitter, LinkedIn, and Amazon Mechanical Turk (MTurk). In the recruitment materials, I provided a cover story that stated the study's focus was on the effects of comedy on problem solving. I presented the following text to potential participants: "Participants 18 years of age and older are wanted for a short survey on the effects of comedy on problem-solving ability."

My use of deception in the current study was to avoid invoking stereotype threat by priming the participants based on the actual nature of the study. The diagnosticity of the academic task and the focus on race were concealed until the end of the participation

activities to avoid setting contextual frames that may have unintentionally invoked stereotype threat. In stereotype threat research, several researchers have demonstrated that indicating a task as diagnostic of an individual's ability may invoke stereotype threat (e.g., Lambert et al., 2016; McGlone & Pfiester, 2015; Steele & Aronson, 1995). Similarly, adding race into the context of the current study may have produced similar effects; for instance, previous research has shown that mentioning racial demographics before a test may invoke stereotype threat in targeted groups (Steele & Aronson, 1995). Participants were presented with the justification of this deception during the debriefing process at the end of the study.

Individuals interested in participating in the current study were provided a link to the study website. Upon reaching the study website, potential participants were presented with the informed consent information. Individuals who consented to participate were shown the instructions. Individuals who did not wish to participate were able to exit the study and were redirected to a webpage thanking them for their consideration and instructing them to close their browser.

Participation

Individuals who gave informed consent moved through four consecutive areas of the study website. First, participants were randomly assigned to watch a prerecorded video of a Black American comedian either using the selected appropriated slur (experimental group) or not using the appropriated slur (control group). Random assignment was implemented through Survey Monkey's A/B Testing mechanism that directs participants to specified items based on a predetermined probability. In this case,

each new participant was directed to the experimental or the control condition video based on a .5 probability.

Second, after watching the video, all participants were directed to complete the academic performance measure labeled as part of the problem-solving task. Third, all participants were directed to complete the stereotype activation measure, which was labeled as a part of the problem-solving task. Next, participants completed the MIBI as a measure of racial identity. Last, participants were directed to a demographic questionnaire requesting information about their gender, age, educational background, and socioeconomic status. At the end of the research activity, participants were presented with a debriefing page that presented information about the actual purpose of the study as well as contact information if the participants had any questions or concerns or wished to report adverse events.

Data Collection

In this study, I used an anonymous online web-based survey design via SurveyMonkey to present information to participants and collect the data. I selected this mode of data collection to reduce overall resources and costs needed to administer the study to an acceptable sample size of participants in a brief timeframe. While online data collection provides great benefits to survey-type research, several limitations and potential issues should be identified including privacy issues (discussed in the Ethical Procedures section below) and the integrity of the data.

Instrumentation and Operationalization of Constructs

In the following section, I present the operational definitions and features of appropriated slurs, stereotype activation, academic task performance, and racial identity within the context of the current study. When appropriate, I provide detailed information about how specific constructs were measured. Moreover, I obtained gained expressed permission to use existing instruments to measure specific constructs in this study—all email correspondences of such permissions have been kept on file and appear in the appendices of this document.

Appropriated Slurs

Appropriated slurs are versions of terms originally used to demean and denigrate a group that have been refurbished by the targeted group as an ingroup term with specific contextual and social rules (Bianchi, 2014; O'Dea et al., 2015). For example, the term *nigger* as a long history in the United States as a derogatory term used to target Black Americans. The original slur, *nigger*, carries connotations of negative attitudes and emotions as well as negative stereotypes (e.g., lazy, stupid, criminals, etc.) toward Black Americans. The appropriated version of *nigger* is mostly commonly *nigga.* A term commonly used within some subsets of the Black American community as a contextually flexible term of endearment or solidarity but may also be used to show disapproval or disagreement. In the current study, I focused on the term *nigga* due to the widespread awareness of the term and its original form (*nigger*) as well as the underlying stereotypes connected to the original term. This appropriated slur served as the independent variable

in this study in which participants were exposed to short, prerecorded video clip of a Black American comedian either using the appropriated slur or not.

For this study, I developed the video material for the independent variable by employing a Black American confederate acting as the comedian. In two versions of the prerecorded video, the comedian delivered the same exact material apart from the appropriated slur. The comedian used the term *nigga* in the stereotype threat condition video and more generic (non-ingroup related) social terms in the no threat condition video. To ensure the ingroup context of the material was consistent across conditions, *nigga* was replaced with generic terms such as friend or brother so that the term conveyed ingroup connectedness or references where appropriate. Moreover, to reduce any confounding influences, each video was recorded in the same setting and under the same conditions with the same confederate who rehearsed and delivered the two versions of the scripted material with the same tone and manner. The jokes used in the videos were inspired by various comedy routines found through Google searches and rewritten for this study.

Stereotype Activation

In the stereotype threat framework, contextual frames activate implicit negative stereotypes that lead to cognitive interference and decreased task performance. Stereotype activation is the mental query of stored heuristics used to inform decisions or behavior; however, stereotype activation is not always within the conscious awareness of the individual experiencing the activation (Wang et al., 2017). While stereotype activation is an implicit cognitive process, it can be effectively measured using various

instruments such as implicit association tests and word fragment completion tasks. Word-fragment completion tasks are based on cognitive priming concepts that holds a target word is recognized with more ease if it is preceded by a related cue (Heyman et al., 2016). In Steele and Aronson's (1995) seminal research on stereotype threat they applied a simple word fragment task as a measure of stereotype activation they presented participants as a measure of cognitive processing. Steele and Aronson built their word-fragment task based on similar methods such as Gilbert and Hixon (1991) who noted that word-fragment tasks demonstrate the cognitive activation of recently primed or self-generated cognitive constructs. More recently, Salter et al. (2017) employed the word fragment task to examine how historic photographs within the context of racial injustice cognitively activate social justice concepts in individuals view the photos. They found that participants who were more visually attentive to photos containing injustice cues completed more social justice related word fragments compared to less attentive viewers. That is, contextual frames in the photographs increased the viewers' cognitive accessibility to social justice concepts.

In the context of this study, I defined stereotype activation as the cognitive availability of stereotype-related words. As such, I measured stereotype activation using a variation of Steele and Aronson's (1995) word-fragment model. Steele and Aronson administered an 80-item measure that consisted of words with blanks representing missing letters in each word (e.g., _ _ ACK). Participants were asked to complete each word fragment with the first word that comes to mind (Steele & Aronson, 1995). Twelve of the 80 word fragments were words associated with Black American race-related

concepts that were developed from a survey of White participants (N = 40) about social perceptions about Black Americans. Their results revealed significant main effects for race alone, $F(1, 61) = 13.77$, $p < .001$, stereotype threat condition alone, $F(2, 61) = 5.90$, $p < .05$, and the interaction between race and stereotype threat condition, $F(2, 61) = 3.30$, $p < .05$, indicating that Black American participants under stereotype threat completed more race-related word fragments (M = 3.7) than their counterparts who were not under stereotype threat (M = 2.1) or White participants in either stereotype threat condition. These results support the appropriateness of a word-fragment task in the current study such that similar results are expected among Black American participants that are exposed to appropriated slurs. If appropriated slurs hold similar contextual attributes to the manipulations used by Steele and Aronson, participants exposed to appropriated slurs (the stereotype threat condition) were expected to produce more stereotype-related fragments than participants not under stereotype threat.

For this study, I used 11 stereotype-related word fragments (the target words) used in Steele and Aronson's (1995) seminal research shuffled into an additional 17 non-stereotype-related filler word fragments. The purpose of the filler words was to reduce the possibility of participants discovering a theme or pattern to the stereotype-related words—I adapted this approach from Steele and Aronson (1995). I selected filler words from the English Lexicon Project (Balota et al., 2007) database by isolating words with similar attributes to the target words used by Steele and Aronson (1995). The English Lexicon Project is an open-source database that houses normative data for speeded naming and lexical decision-making for more than 40,000 words. The English Lexicon

Project's data was collected from approximately 1,200 participants across six universities in the United States and includes descriptive attributes (length, part of speech, frequency, etc.) and behavioral data related to each word including naming reaction time and accuracy. The naming reaction time and accuracy is the amount of time (in milliseconds) participants took to identify a word and if that identification was correct.

After downloading the English Lexicon Project's data set, I used the R-Programming Language (v. 4.0.1) and R-Studio (v. 1.3.1) interactive development environment to query the descriptive and behavioral data for the 11 target words and calculated summary statistics on their attributes—specifically, the minimum and maximum values for word lengths, frequencies, and reaction times. The R code I produced for this task is presented in the appendix. The target words ranged from four to eight characters in length, from 6,326 to 160,756 in frequency, and from 518.5 ms to 683.0 ms in mean naming reacting time. After removing the target words from the full database, I filtered the remaining entries (N = 79,661) to exclude words with special characters (i.e., apostrophes), capital letters, and with missing attribute data. Then, I selected words with attributes that fell within the attribute ranges of the target words. From the matched words (N = 3,126), I used a simple random sampling algorithm to select 33 (three times the number of targeted words) potential filler words. Finally, I reviewed the 33 filler words to ensure they were not synonymous to the targeted words.

To further ensure the filler words were comparable to the target words, I analyzed the missing letter patterns of the target words as Steele and Aronson (1995) presented them to participants. I calculated the portions of each combination of the number of blank

spaces and their position in the target word (beginning, end, or gap) by word length groups. For example, 66.7% of the four-letter target words have two blanks at the beginning of the words and 33.3% of the same length words have two blanks at the end of the word. I used these proportions to find the number of filler words to match to the characteristics of the target words. Continuing the current example, I modified five of the seven four-letter filler words to have two blank letters at the beginning of the word (0.667 x 7 = 5). I applied this method across all the letter length groups except for the six-letter and eight-letter groups. There were no six-letter target words from which to select a missing letter pattern and only one eight-letter target word; therefore, I arbitrarily selected from the five- and seven-letter patterns and applied them to the six- and eight-letter filler words. I reduced the final set of filler words to 11 through random selection to reduce the overall length of the survey. The final list of target and filler words (and their missing letter forms) are included as an appendix in this document.

Word-fragment task scores were calculated by the total number of fragments completed that match the potential stereotype-related answer. For instance, when participants were presented with the word-fragment, _ _ C E, possible answers include PACE, LACE, and RACE. I held the assumption, based on the previously discussed stereotype activation research and Steele and Aronson's (1995) seminal research, that individuals primed with race-related stereotypes through contextual cues in the environment would be more likely to select word-fragment responses related to the stereotype. In this example, I expected participants under stereotype threat (the appropriated slur condition) to select RACE more than other possible responses. I used

the same race-related word list Steele and Aronson cited in their 1995 studies (see Appendix B).

Academic Task Performance

Stereotype threat impacts behavioral outcomes related to relevant stereotypes that target an individual's social group. One highly studied behavior in the stereotype threat literature is academic test performance of racial minorities and females—groups that have been historically stereotyped as unintelligent or less able to perform in academic domains such as math or science (Tyler et al., 2016). A range of academic measures have been used in stereotype threat research including, but not limited to, exercise items selected from various mathematics textbooks, intelligence tests such as the Wonderlic Aptitude Test, and standardized academic tests such as the GRE and the SAT (e.g., Burkley et al., 2016; Steele & Aronson, 1995; Wright-Adams, 2014). In the tradition of stereotype threat research, I used selected items from the SAT to measure academic task performance.

The SAT, published and maintained by the College Board and Educational Testing Service group, is an instrument intended to measure an individual's verbal and mathematical abilities and is traditionally used to partially inform admissibility into institutions of higher education. The SAT was adapted from early intelligence assessments developed in the 1920s by a committee of psychologists for military recruitment (Gregory, 2007). The first official administration of the SAT was in 1926—over several decades this assessment became increasingly popular and remains a standard instrument.

In this study, I administered only items from the math section of the SAT. The SAT math section are composed of multiple choice and grid-in (entering a unique answer rather than selecting an existing answer) that either allow or prohibit the use of a calculator. Math items focus on three major areas of mathematic ability related to a wide range of college majors and career fields. These major areas are algebra, problem solving and data analysis, and advance mathematic concepts; moreover, the SAT math items incorporate some trigonometry and geometry related concepts. The full math section of the SAT consists of 58 questions to be answered in 80 minutes and is scored by the total number of correct answers.

In addition to its well-established use as a standardized academic test, the appropriateness of the SAT as a dependent variable in the current study was based on its successful use in various stereotype threat studies and similar research as either a covariate or dependent variable. Wright-Adams (2014) administered ten randomly selected math items from the SAT study guide as a dependent variable to study stereotype threat in Black American women. Robinson (2016) used the 15 items from the SAT as an outcome to explore the effects of gender and university affiliation stereotypes on female math performance. Similarly, research examining the correlation between academic performance and test anxiety used the SAT as a performance measure (Anis et al., 2016). Moreover, the stereotype threat model holds that the effects of stereotype threat only manifest in tasks that are at least moderately difficult for the individual (Steele & Aronson, 1995; Wright-Adams, 2014). The SAT items were assumed to provide the level of challenge necessary to observe stereotype threat.

In this study, I used a 10-item subset of randomly selected math items from the Official SAT Study Guide (College Board, 2017) as the academic performance variable (see Appendices D and E for selected items and permission of use). Like the SAT scoring method and other stereotype threat studies, the academic SAT items used in the current study were scored by the total number of correct answers participants submit. Specifically, one point was awarded for each correct answer and a sum of the points awarded will serve as the task score. As such, higher total scores (a maximum of 10) will indicate higher academic task performance.

Racial Identity

For the purposes of this study, I defined racial identity within the context of the MMRI (Sellers, Smith, et al., 1998). The MMRI's definition of racial identity is "the significance and qualitative meaning that individuals attribute to their membership" in their racial group (Sellers et al., 1998, p. 23). Racial identity has been implicated, either directly or indirectly, as an influencing factor in stereotype threat. Indirectly, Massey and Owens (2014) found that racial identity as a function of skin tone among Black American college students predicts their susceptible to stereotype threat; that is, *blacker* or darker skinned Black Americans who more likely to have a stronger racial identity were less impacted by stereotype threat conditions. More directly, Shelvin et al. (2014) demonstrated that both strong and weak racial identities can influence one's susceptibility to stereotype threat. One on hand, they found that a strong racial identity can increase stereotype threat effects but also act as protective factors in the absence of stereotype threat cues. On the other hand, they found that a weaker connection to one's racial group

may also increase one's vulnerability to stereotype threat—similar to Massey and Owens' (2014) proposition that lighter skinned Black Americans (who are less confident about their racial identity) are more vulnerable to stereotype threat.

I measured racial identity in this study with the MIBI (Sellers et al., 1997b). The MIBI was developed by (Sellers et al., 1997a) to examine the constructs of the MMRI, which states that racial groups such as Black Americans may have several social identities including, but not limited to, race that influences their cognition and behaviors in various social context. The MIBI attempts to capture three of the consistent dimensions of the MMRI—centrality, ideology, and regard—with a 27-item, 7-point Likert-type survey. Overall, higher total scores indicate stronger Black American racial identity.

In their initial examination of the MIBI, Sellers et al. (1997a) asked 474 Black American college students from introductory psychology courses at two Mid-Atlantic (United States) university to complete a 71-item MIBI and various race-related behavior surveys over the course of five academic semesters. While a factor analysis did not reach an acceptable level for all the items together (KMO < .60), analyses for each subscale reached acceptable levels indicating the MIBI possesses three distinct but interrelated constructs. Subsequent factor analysis for each construct was conducted resulting in a 51-item revision of the initial MIBI containing an eight-item centrality scale, a 36-item ideology scale, and a seven-item regard scale. An analysis of the inter-scale correlations indicated an acceptable internal validity. As predicted by Sellers et al. (1997a), the centrality scale was positively correlated with the private regard and nationalist attitudes subscales. Further, high centrality was negatively correlated with the assimilation and

humanistic subscales. The humanistic subscale showed a positive relationship with the assimilation subscale. The oppressed minority subscale was positively correlated with the assimilation subscale.

Sellers et al. (1997a) also examined the external validity of the MIBI by analyzing the relationships between the MIBI scores and the participants' race-related behaviors. Overall, a MANOVA revealed a significant relationship among the MIBI subscales and report of having a Black American best friend, $F(1, 472) = 9.74$, $p < .01$. When considering each subscale, participants who reported having a Black American best friend scored higher on the centrality subscale, $F(1, 472) = 12.35$, $p < .01$, and the nationalist scale, $F(1, 472) = 37.45$, $p < .01$. In contrast, participants with a Black American friend scored lower on the assimilation subscale, $F(1, 472) = 19.26$, $p < .01$, the humanist subscale, $F(1, 472) = 12.45$, $p < .01$, and the oppressed minority subscale, $F(1, 472) = 19.68$, $p < .01$. No significant relationship was found between best friend reports and the private regard subscale.

The relationship between enrollment in Black studies courses also showed an overall significance, $F(6, 467) = 3.44$, $p < .01$ (Sellers et al., 1997a). Participants who enrolled in at least one Black studies course scored higher scores on the centrality subscale, $F(1, 472) = 7.98$, $p < .01$ and nationalism subscale, $F(1, 472) = 18.32$, $p < .01$, but no other subscale presented a significant relationship with this behavior. A correlation analysis was conducted to examine the relationships between interracial contact and the MIBI subscales. Significant positive relationships were found between contact with other Black Americans and the centrality subscale ($r = .39$, $p < .01$), nationalism subscale ($r =$

.39, $p < .01$), and private regard subscale ($r = .27$, $p < .01$). In contrast, contact with Whites showed a negative relationship with the centrality subscale ($r = -.46$, $p < .01$) and the nationalist subscale ($r = -.41$, $p < .01$).

The MIBI has also been shown to be a reliable measure beyond Seller's initial development. Vandiver et al. (2009) found a relatively high reliability score for the centrality scale ($\alpha = .80$). Similarly, Helm (2001) examined the MIBI in a sample of 388 Black Americans and found reliability scores above .70 across the MIBI subscales—specifically, the centrality scale was found to have a .71 Cronbach's alpha. Simmons et al. (2008), however, found a slightly lower reliability score ($\alpha = .66$) for the centrality score.

The appropriateness of the MIBI as a moderating variable in the current study stemmed from its application in the Shelvin et al. (2014) study. The researchers found that racial identity profiles based on scoring patterns of the MIBI subscales highlighted individuals that were more susceptible to stereotype threat. Specifically, they found that individuals with a strong racial identity (Highly Connected profile) performed worse under stereotype threat than under no-threat conditions. This relationship may stem from highly identified individuals' greater concern about negative race related stereotypes; people who identify strongly with their race may feel more obligated to not live up to the negative stereotypes. Shelvin et al. (2014) also found that, even though they underperformed compared to their stereotype threat condition counterparts, individuals with strong racial identities in the no-threat condition still outperformed participants in

the other racial identity profile clusters. They argue this pattern suggests a protective factor of strong racial identities in the absence of stereotype threat.

Following this line of work, I used the centrality scale of the MIBI in the current study as a moderator of stereotype threat (see Appendices F and G for items and permissions). The centrality scale of the MIBI is a 10-item scale with each item measured on a 7-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). A minimum score of 10 suggests an individual does not hold their racial group as fundamental or dominant to their identity. In contrast, a maximum score of 70 indicates an individual's racial group is highly valued and dominant to their identity. The assumption that racial identity will moderate academic performance and stereotype activation aligns with the literature positing that stronger racial identity motivates individuals to disprove negative stereotypes; however, this motivation paradoxically leads to increases in stereotyped behaviors when those individuals are under threat.

Data Analysis Plan

I used SPSS v27.0 (IBM, 2020) software to conduct all data preparation and analyses. After downloading the final dataset from SurveyMonkey, I screened and cleaned the data by initially checking its accuracy, missingness, and outliers. Then, I analyzed each of the research questions using a series of ANOVA models. The follow sections detail the overall data screening, cleaning, and analysis plan.

Data Screening and Cleaning

Accuracy is ensuring that the data is received in the type and form that is expected. There are two fundamental types of data—numeric and categorical. Data may

inadvertently be converted from one for to another for several reasons; for example, if categorical data is downloaded from an online survey into a data file, the download process may convert the categories or factors into number rather than the expected words of the categories (Soley-Bori et al., 2013). Similarly, numerical or continuous data inaccuracies may be apparent when data points fall outside of an expected range. For instance, the expected range on the academic performance task in the current study was 0 to 10—one point for every current answer. However, if a score appears at a value of 55 points, this will indicate an accuracy issue. Further, accuracy issues may occur when participants input an incorrect or unexpected value when taking a survey or a test. For example, a participant who is asked to report the number of his or her children in an open text field on an online form may accidently input 77 rather than 7, indicating a typographic error.

In the current study, accuracy issues related to participants' input errors are expected to be unlikely because the online survey tool will be developed to accept only specific values connected to the expected values of the selected instruments. That is, the academic performance task (i.e., SAT), the MIBI centrality scale, and demographic items were presented as multiple-choice, single-selection items from which participants could only select one of the given options. The accuracy of the continuous variables with predetermined response items were screened by examining the ranges (the minimum and maximum) and a frequency table of the response values for each variable's collected data. The range of data identifies if there are any data points outside the expected values and the frequency table identifies how many of the unexpected values exist in the dataset.

The expected range for the academic performance task was 1 to 10 and for the MIBI centrality scale was 10 to 70. If the values of the data response are unexpected across the entire dataset this may indicate a coding issue as discussed earlier. In this case, I planned to remove such data points and mark them as *NA* or missing data.

Once the data was screened for accuracy and appropriately modified, I examined the missingness—or the lack of expected data—present in the dataset. Data can be missing completely at random (MCAR) or missing not at random (MNAR). MCAR data occurs when the probably of one variable missing data is not related to the probability of another variable missing data (Soley-Bori et al., 2013). MNAR data typically indicates an issue with the data collection or entry methods. For example, missing answers from the same portion of survey across several participants may be due to item wording issues or errors made during manual data entry by a researcher. If a MNAR issue is detected, further exploration of the data collection method is required to understand the source of the issue. In the current study, I planned to use Little's MCAR test to assess if the data collected from the online tools is missing at random. If the chi-squared outcome of Little's test is not significant ($p > 0.05$), I would fail to reject the null hypothesis of the missingness analysis that missing data is not randomly missing. In other words, a non-significant p-value from the Little's test would indicate that the missing data are missing completely at random. If the missingness analysis reveals a MCAR result, I planned to treat the missing data appropriately by variable including deletion or imputation. NMAR data may also lead to further analysis on data collection issues with the surveys and online collection tool (i.e., Survey Monkey).

Finally, I checked the dataset for outliers—data points that fall outside of the expected data range for any given variable. While a check of the expected ranges is performed during the accuracy screening, an outlier check should be conducted after the accuracy and missingness analyses are conducted because data modification at these two stages of the screening and cleaning process may change the data and create outlier that were not previously present. Moreover, at this stage of screening and cleaning, outlier analysis will become slightly more rigorous for continuous variables. For the continuous variables, I calculated z-scores and examine the scores that are 3.28 standard deviations above or below zero—scores beyond this threshold were considered outliers (Field, 2013). That is, a z-score either greater than 3.28 or less than -3.28 was considered an outlier and excluded from the analysis.

After the data was initially screened, I diagnosed the data for the basic assumptions of the ANOVA test. I discuss these assumptions and their results in Chapter 4.

Research Questions and Hypotheses

RQ1: What is the difference in academic test performance between Black American adults exposed to appropriated slur and Black American adults not exposed to appropriated slurs?

$H_0$1: Black American participants exposed to appropriated slurs will exhibit similar scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants not exposed to appropriated slurs.

$H_1$1: Black American participants exposed to appropriated slurs will exhibit significantly lower scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants not exposed to appropriated slurs.

RQ2: What is the difference in academic task performance between Black American adults who exhibit higher racial identity and Black American adults who exhibit lower racial identity?

$H_0$2: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, will exhibit similar scores on the academic task, as measured by the selected items of the SAT, compared to Black American participants who exhibit low racial identity scores.

$H_1$2: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, will exhibit significantly higher scores on the academic task, as measured by the selected items of the SAT, compared to Black American participants who exhibit low racial identity scores.

RQ3: To what extent does racial identity moderate the effects of appropriated slurs on academic test performance?

$H_0$3: Black American participants with high racial identity scores, as measured by the MIBI centrality scale, who are exposed to appropriated slurs will exhibit similar scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants with low racial identity scores who are exposed

to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

$H_1$3: Black American participants with high racial identity, as measured by the MIBI centrality scale, who are exposed to appropriated slurs will exhibit significantly lower scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants with low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

RQ4: What is the difference in the cognitive activation of negative stereotypes toward one's own racial group between Black American adults exposed to appropriated slur and Black American adults not exposed to appropriated slurs?

$H_0$4: Black American participants exposed to appropriated slurs will exhibit similar negative racial stereotype activation, as measured by the word fragment task, compared to Black Americans participants not exposed to appropriated slurs.

$H_1$4: Black American participants exposed to appropriated slurs will exhibit significantly more negative stereotype activation, as measured by the word fragment task, compared to Black American participants not exposed to appropriated slurs.

RQ5: What is the difference in the cognitive activation of negative stereotypes toward one's own racial group between Black American adults who exhibit higher racial identity and Black American adults who exhibit lower racial identity.

$H_0$5: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, will exhibit similar negative racial stereotype

activation, as measured by the word fragment task, compared to Black American participants who exhibit low racial identity scores.

$H_1$5: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, will exhibit significantly lower negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants who exhibit low racial identity scores.

RQ6: To what extent does racial identity moderate the effects of exposure to appropriated slurs on the cognitive activation of negative stereotypes toward one's own racial group in Black American adults?

$H_0$6: Black American participants with high racial identity scores, as measured by the MIBI centrality scale, who are exposed to appropriated slurs will exhibit similar negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

$H_1$6: Black American participants with high racial identity, as measured by the MIBI centrality scale, who are exposed to appropriated slurs will exhibit significantly higher negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants with low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

Considering the focus on group difference in the stated hypotheses, I employed a set of two-way independent analysis of variance (ANOVA) models to address the research questions. ANOVA models with more than one independent variable, known as factorial ANOVAs, are a special case of linear regression models in which two or more categorical independent variables are examined for differences on a dependent variable (Toothaker, 1993).

Similar models have been applied throughout both seminal and recent stereotype threat research discussed in previous sections. For instance, in their seminal work on stereotype threat, Steele and Aronson (1995) employed several factorial ANOVAs to answer their research questions across several studies. In their initial 2x3 factorial study design, they used an ANCOVA (analysis of covariance) model to examine the group differences between Black American and White participants assigned to either a stereotype threat or no-threat condition while controlling for previous SAT scores (a continuous variable)—they used similar models in subsequent studies. The two categorical independent variables (race and experimental condition) allowed the researchers understand group differences leading them to uncover several aspects of the stereotype threat phenomenon and set the model for future research. Similarly, Galdi et al. (2014) used an ANOVA model—in a 2x2 factorial design—to examine the differences in math performance among male and female children place in either a stereotype threat or no-threat condition (coloring picture with subtle gender-based stereotypes or a neutral landscape picture). This statistical model allowed the researchers

to support their hypotheses that subtle contextual cues or implicit associations can invoke stereotype threat.

For the models in this study, one ANOVA model used the academic performance task scores as the dependent variable and the other ANOVA model used the word-fragment task scores as the dependent variable. Moreover, the independent variables in both statistical models were the appropriated slur condition and a two-level (high and low) dichotomized MIBI centrality variable. The appropriated slur condition was represented as a binary variable in which 0 indicates assignment to the control group (no slur exposure) and 1 indicates assignment to the experimental group (slur exposure). Similarly, the MIBI centrality measure was represented as a binary variable calculated using a median split dichotomization method resulting in binary categories in which 0 indicates participants whose MIBI centrality scores were under the median of the collected scores (labeled as the low centrality group) and 1 indicates participants whose MIBI centrality scores were at or over the median (labeled as the high centrality group). I used a dichotomized MIBI central variable in the current statistical models (factorial ANOVA) based on research design and theoretical frameworks found in the stereotype threat literature as well as to obtain an interpretation of the data that would be more directly aligned with the current hypotheses. In the following sections, I elaborate on my rationale of using a dichotomized MIBI centrality variable drawing on both stereotype threat research and statistical literature.

I first highlight examples of dichotomizing the MIBI within the context of stereotype threat research before turning to the advantages and drawbacks of using the

median split technique. As part of a larger seminal study on the relationship between racial identity (through the framework of the multidimensional model of racial identity) and academic achievement, Sellerset al. (1998) examined the moderating effects of racial centrality on the relationship between racial ideology and academic outcomes (i.e., GPA). The researchers conducted a median split on the sample's MIBI centrality scores to create low and high levels of this measure and utilized this variable in their statistical models. Overall, Sellers et al. found a positive relationship between racial centrality and a negative relationship between racial ideology and GPA. Specifically, participants with strong racial identities (centrality) had higher GPAs while participants with higher ideology scores, indicating distant from their racial group, had lower GPAs. Additionally, the researchers found that centrality moderated the relationship between ideology and academic performance such that participants in the high centrality group (i.e., strong connection to their racial identity) who did not endorse an assimilation or nationalist ideology had significantly higher GPAs than those in the low centrality group. Similarly, Shelvin et al. (2014) used a hierarchal clustering to classified participants into groups (clusters) based on their MIBI subscale means. The clustering technique allowed the researchers to identify six distinct groups including high connected and low connected that described participants who produced the highest and lowest mean centrality scores, respectively. Moreover, the high connected group was described by the researchers as individuals who placed more importance on their racial identity while the low connected group placed less importance on this aspect of self. While hierarchal clustering is a far

more complex method compared to median splitting, it demonstrates the utility of converting continuous MIBI scores into descriptive groupings.

Turning to the statistical justification of using a median split to dichotomize the MIBI centrality score, I will now discuss the benefits and hazards of this method. For instance, while dichotomization supports more direct and easily understandable interpretations related to group differences along racial identity, researchers often express concerns about loss of information when converting a continuous variable to a dichotomous variable (Iacobucci et al., 2015). This argument holds that the fidelity for scores nearest to the median will be hidden by a simplistic binary grouping. That is, scores just below and above the median will be forced in to the lower or upper category implying the are inherently different; however, scores near the median may be more alike than different. Similarly, dichotomization also removes the variability of scores within a group. For instance, in a dichotomized group, scores just over the median are not conceptually different than highest scores in the variable.

In other words, researchers should be concerned about losing insights that stem from the variability of a continuous variable when overgeneralizing scores into dichotomous groupings (Iacobucci et al., 2015). Importantly, the loss of fidelity resulting from dichotomization may contribute to Type II errors (lowering effect sizes and statistical power) when deciding whether to reject the null hypothesis. Recent research has addressed the traditional concerns about median split dichotomizing demonstrating that under certain conditions this method can be appropriate and robust. Across two studies, for instance, Iacobucci et al. (2015) conducted Monte Carlo simulations to better

understand the impacts median split dichotomizing has on statistical modeling. The results replicated previous claims about the impact of dichotomization one the relationship between the independent and dependent variables—that is, a loss of power and effect. However, their simulation also demonstrated that such as loss in power is negligible and can be mitigated with a strong research design and sampling approach.

Another concern about dichotomizing variables increases the possibility of making Type I errors. Dichotomizing several variables, for instance, in the context of a multiple regression model may contribute to the likelihood of committing a Type I error. That is, a continuous variable may not reach statistical significance in a regression model but may become significant in the same model as a dichotomous variable. While there are many published claims to the spurious effects of dichotomizing variables, Iacobucci et al. (2015) demonstrated that such spurious effects are often achieved in the context of statistical models in which more than one continuous variable is converted to a dichotomized variable. Therefore, the researchers suggested that limiting dichotomization to one independent variable and using ANOVA models will mitigate the Type I error concerns. Moreover, they found that limiting dichotomization to a single independent variable within the context of an orthogonal experimental design and appropriate statistical method (e.g., ANOVA) mitigates the spurious effects of dichotomizing.

Multicollinearity plays a significant role in how the median split influences a statistical model. Specifically, when there is a relatively high correlation between independent variables, it is more likely that median split produced factor may produce spurious effects. However, such effects can easily be checked by examining the data set

before running an analysis and by thoughtful research design planning. Moreover, Iacobucci et al. (2015) noted that the 2x2 factorial design, such as the design I utilized in this study, remains especially robust when applying the median split dichotomization to only one of the independent variables. Previous research that aligns with the theoretical framework and focus of the current study as well as recent elucidation on the use median split dichotomized variables supports the implementation of the 2x2 factorial design with the median split conversion of one independent variable in the current study. I continue this section with a brief discussion on the details of the two-way ANOVA models including the expected outputs and potential interpretations.

In the statistical analysis plan for the current study, the academic task performance ANOVA model was intended to address the hypotheses for research questions one through three while the word-fragment task ANOVA model was intended to address the hypotheses for research questions four through six. For each model, I produced the main effects and interaction effects of the independent variables on the dependent variable. Prior to conducting the data analysis, I set the following criteria for evaluation of each model and interpretation of the research questions. A significant main effect ($p < 0.05$) of the slur condition in the academic performance task model would indicate that a significant difference exists between the mean academic task scores of participants exposed to appropriated slurs and participants not exposed to the appropriated slurs. If true, an examination of the mean scores for each group would reveal the nature of the difference between the experimental and control groups. Given a significant main effect and a lower mean academic task score for the experimental

(exposure to the appropriated slur) group compared to the control group (no appropriated slurs), there would be evidence to reject the null hypothesis for research question one and support the research hypothesis that Black American participants exposed to appropriated slurs perform worse on the academic task than Black Americans who were not exposed to the appropriated slurs.

Similarly, a significant main effect ($p < 0.05$) of the racial identity in the academic performance task model would indicate that a significant difference exists between the mean academic task scores of participants in the high racial identity group and participant in the low racial identity group. Further, examination of the mean scores for each racial identity group to understand the nature of the difference between the high and low racial identity groups. Given a significant main effect and a higher mean academic task score from the high racial identity group, there would be evidence to reject the null hypothesis for research question two and support the research hypothesis that participants with high racial identities perform better on the academic task compared to participant with low racial identities.

A significant interaction effect ($p < 0.05$) in the academic performance task model would indicate that racial identity level influences the effect of exposure to appropriated slurs on academic task performance. An examination of group means, and a simple slopes analysis will allow me to expand on the nature of the differences between the four interaction groups. If the mean academic performance task score for the high racial identity/appropriated slur group is the lowest compared to the other groups, I will have evidence to reject the null hypothesis for research question three and support the research

hypothesis. That is, Black Americans with high racial identities who are exposed to appropriated slurs will performance worse on an academic test compared to participants with low racial identities exposed to the same slurs and compared to those with high and low racial identities not exposed to appropriated slurs.

A significant main effect ($p < 0.05$) of the slur condition in the word-fragment model would indicate that a significant difference exists between the mean stereotype activation scores of participants exposed to appropriated slurs and participants not exposed to the appropriated slurs. Further, if significant, an examination of the mean scores for each group to understand the nature of the difference between the experimental and control groups. Given a significant main effect and a higher mean stereotype activation score for the experimental (exposure to the appropriated slur) group compared to the control group (no appropriated slurs), I will have evidence to reject the null hypothesis for research question four and support the research hypothesis that Black American participants exposed to appropriated slurs produce more than Black Americans who were not exposed to the appropriated slurs.

Similarly, a significant main effect ($p < 0.05$) of the racial identity levels in the word-fragment model would indicate that a significant difference exists between the mean stereotype activation task scores of participants in the high racial identity group and participant in the low racial identity group. As such, examination of the mean scores for each racial identity group to understand the nature of the difference in stereotype activation between the high and low racial identity groups. Given a significant main effect and a lower mean stereotype activation score from the high racial identity group, I

will have evidence to reject the null hypothesis for research question five and support the research hypothesis that participants with high racial identities produce less negative racial stereotype word fragments related to their own racial group compared to participants with low racial identities.

A significant interaction effect ($p < 0.05$) in the word-fragment model would indicate that racial identity level influences the effect of exposure to appropriated slurs on negative stereotype activation. An examination of group means, and a simple slopes analysis will allow me to expand on the nature of the differences between the four interaction groups. If the mean stereotype activation score for the high racial identity/appropriated slur group is the highest compared to the other groups, I will have evidence to reject the null hypothesis for research question six and support the research hypothesis. That is, Black Americans with high racial identities who are exposed to appropriated slurs will produce more negative stereotype word fragments compared to participants with low racial identities exposed to the same slurs and compared to those with high and low racial identities not exposed to appropriated slurs.

Support for the research hypotheses for research questions one through three would suggest that exposure to appropriated slurs decrease academic task performance; however, a more central racial identity—that is, holding one's racial identity as important—may exacerbate the effects of exposure to appropriated slurs on academic task performance. Similarly, support for the research hypotheses for research questions four through six would suggest that exposure to appropriated slurs increase negative racial stereotype activation toward one's own racial group; however, a more central racial

identity may exacerbate the effects of appropriated slurs on negative stereotype activation. Taken together, support for research hypothesis for the current research questions may suggest that exposure to appropriated slurs invoke stereotype threat as indicated by increased negative racial stereotype activation and decreased academic task performance in Black Americans. Moreover, such results would suggest that racial identity moderates the effects of appropriated slurs on stereotype threat. I discuss the actual results in Chapter 4.

Threats to Validity

In the context of research, validity refers to the "approximate truth of an inference" put forth by conclusions of the researcher (Shadish et al., 2002, p. 34). Validity is the estimate of how the researcher's observations reflect reality within and outside of the context of the study—known as internal and external validity, respectively. Validity, however, is not a property of any given empirical method or process but a characteristic of the assertion of knowledge based on the selected method. Moreover, validity is not a certainty and may be inaccurate or incorrect when making inferences. Such errors are called threats to validity and these threats should be identified and addressed to minimize their impacts on the posited inferences. In the following sections, I discuss threat to both internal and external validity for the current study and potential actions to mitigate those threats if needed.

Threats to Internal Validity

One of the goals of empirical research is to support the stated relationship between variables as they are measured or manipulated—the internal validity (Shadish et

al., 2002). To that end, researchers attempt to demonstrate that an independent variable both temporally precedes and covaries with the dependent variable or variables. Moreover, researchers must present data-driven rationale that no other plausible explanation exists for the relationship between the independent and dependent variables. Threats to internal validity are the possible causes or influences on the dependent variables other than the application of the independent variable. To understand the nature of the relationship between the dependent and independent variables in the context of the current study, it would be judicial to discuss possible threats to internal validity including selection and attrition.

Sample selection, as a threat to internal validity, occurs when some subset of participants (e.g., those in the experimental condition) differ from another subset before they are exposed to any of the experimental manipulations (Frankfort-Nachmias & Nachmias, 2015). Such a difference could influence the interpretation of the results leading to inferences about the experimental conditions that are false. In the current study, I addressed selection bias by implementing a random assignment procedure that places participants into the experimental or control conditions based on computer-initiated algorithms rather than the participant qualifications, characteristics, or availability. That is, after participants consent to participation, they were randomly directed to either the experimental or control versions of the manipulation stimuli. Similarly, self-selection bias occurs when individuals chose to participate in a study because of their interest in the research topic (Heinen et al., 2018). Such a weighted

sample may lead to skewed responses. In this study, however, I initially used deception to conceal the true goals of the study, which may reduce self-selection bias.

Attrition refers to participants not completing the research activities and may occur for several reasons including, in the context of current study, loss of internet connectivity, distraction from research activities, or loss of interest (Frankfort-Nachmias & Nachmias, 2015). Attrition is a type of selection threat which cannot fully be addressed by random assignment. However, I monitored the data collection for completed research records and extended the recruitment data collection period until the desired sample size has been reached as the best approach to address attrition.

Threats to External Validity

External validity is the ability to generalize the conclusion of a study to the wider population outside of the selected sample (Frankfort-Nachmias & Nachmias, 2015). In this study, one threat to external validity lies in the potentially limited variation of characteristics within the sample size. That is, Black Americans do not all share the same perspectives, values, backgrounds, and social environments which may lead quite different outcome when exposed to appropriated slurs (Allan, 2015). For example, during the recruitment process in this study, potential participants may have belonged to a generational cohort that perceive appropriated slurs differently than individuals from another cohort. Further, nearly every minority group in the United States has adopted an appropriated slur for ingroup communication; however, other groups (e.g., lesbians, Hispanics, women) may have different emotional and cognitive reactions to appropriated slurs when compared to Black Americans based on historic social experiences (Gaucher

et al., 2015). Also, considering the potential participants are a subset of the population that has access to the internet and the motivation to participate in online research, the sample will likely not match the overall characteristics of the target population. To address such issues of external validity, I avoided making overly generalized statements about population outside of the selected sample during the conclusions and discussion of the results. Instead, I made every attempt to suggest future research to elucidate the potential finding of the current study to the wider population and across various groups.

Ethical Procedures

In this study, I adhered to the American Psychological Association's (APA) Code of Ethnic as well as the guidelines established by Walden University's Institutional Review Board (IRB) to ensure ethical research practices. To that end, I submitted the proposed research to Walden University's IRB office for review and approval (Code of Ethics, Section 8.01)—approval number: 04-09-20-0194883. The major ethnics concerns during the current included informed consent, privacy and anonymity, and use of deception.

Informed Consent

The APA provides overall guidance around informed consent in Section 8.02 of the Code of Ethics. Overall, the informed consent process provides participants with detailed information about the nature of the study and the risks and benefits of participation. Informed consent information may include the purpose of the research, the activities and time commitment expected during participation, the participant's right to decline or withdraw from participation and the consequences of such actions, limitation

on confidentiality and privacy during data collection and analysis, and any physical or psychological risks associated with participation.

In this study, potential participants accessed the study website in SurveyMonkey via the provided link where they will be presented with the informed consent information. Participants were informed that the current research seeks to examine the effects of comedy on problem solving. While this was not the true intention of the study, the rationale and justification for this initial deception will be discussed in the next section. Further, participants were informed that participation will involve watching a short video of a standup comedian and completing a brief problem-solving challenge that should take approximately 20 to 25 minutes.

Participants were also informed that their participation would be completely voluntary and anonymous, and they may decline or withdraw from participation at any time without penalty or repercussions. Individuals who consented were automatically directed to the research activity sections of the survey. Individuals who did not wish to participate were directed to a thank you message ending their interaction with the research site. Participants who completed the research activities were directed to a debriefing page that will explain the actual purpose of the study and the rationale behind the initial deception.

My contact information and contact information for Walden University's Institutional Review Board office was provided to the participants should they have any concerns or questions about the current study. Moreover, in the event that a participant experienced emotional or psychological distress as a direct result of participation in the

current study, a tollfree number to a national mental health line was provided during the consent process.

Deception in Research

Ethical research includes transparent and honest interactions with participants. In some cases, however, researchers may employ deception to gain access to more natural and honest reactions from participants. In this study I sought to use such deception to conceal the true purpose of the study—examining the influence of appropriated slurs on stereotype threat. The APA recognizes this need and provides guidance for best practices when deception is needed during research. The APA Code of Ethics, Section 8.02(a) states that researchers do not use deception in research unless they can provide justification that such actions support the study's scientific merit. The rationale for the use of deception during the recruitment and informed consent process in this study was based on previous scientific finding as previously discussed. The use of deception was justified to avoid priming participants and invoking stereotype threat from a cue or information in the informed consent about the actual nature of the study. That is, the diagnosticity of the academic task and the focus on race was hidden until the end of the participation activities to avoid setting contextual frames that may have invoked stereotype threat.

Relatedly, the APA Code of Ethics, Section 8.02(b) notes that deceiving potential participants about research activities could predictably cause physical pain or emotional distress. While participants may not approve of the appropriated slurs used in the current study, I assumed because of the ingrained and wide-spread use of the selected terms in

film, television, music, and society that even unexpected exposure to the appropriated slur will not cause physical harm or emotional distress in participants.

APA Section 8.02(c) of the Code of Ethics requires researchers to disclose the use of deception and its justification to participants after their participation and allow participants to withdraw their data from further use in the study. In accordance with this guideline, as well as Section 8.08, participants were debriefed about the use of deception and its justification in the context of this study. In this study, participants were debriefed about the actual focus of the study upon completion or exiting of the research activities. Participants were presented with a webpage explaining the use of deception and the justification for its use. Moreover, participants were presented with an opportunity to withdraw their data from the study after learning about the use of deception.

Privacy

According to the APA Code of Ethics, psychologists and researchers should plan and implement procedures to protect the privacy and confidentiality of participants. Overall, researchers are obliged to protect confidential information and data obtained from participants in both physical and digital form; moreover, researchers should clearly inform participants about those precautions and procedures as well as any limitations on privacy and confidentially. In this study, I collected data via an anonymous online survey tool eliminating the intentional transmission and storage of personal identifying information such as name or contact information. However, online communications may transmit virtual identification information such as IP addresses or geolocation information that could be linked to an individual and encroach on their privacy. If such

information would have been collected, it would have been immediately removed from the datasets during the data cleaning process and deleted from the online collection. Further, this risk will be presented to participants during the informed consent process so that they understand the potential risks of participating in an online study.

Despite the anonymous nature of the data collected in the current study, security and safeguarding procedures will be implemented to protect it. Data collected via the survey website were downloaded over a trusted and secure network through the designated research computer to an external password-protected hard drive. The hard drive was only connected to the research laptop when the data were being download, processed, or analyzed. When not in use, this hard drive was stored in a locked file box only accessible to the primary investigator.

Summary

In this chapter, I discussed the structure and rationale behind the quantitative posttest-only with control group design I have selected to address the research questions in this study. As a part of the methodology discussion, I defined the overall population of interest, the sample framing and sampling procedure, as well as a statistically derived sample size for this study. Further, I presented the procedures for recruitment, participation, and data collection methods via online resources. I have operationally defined the key constructs of appropriated slurs, stereotype activation, academic task performance, and racial identity within the scope of the current study. Moreover, I have presented instrumentation that will measure these constructs and laid out the statistical analysis procedures in which a series of ANOVAs were used to statistically analyze the

research questions and hypotheses. I have also addressed threats to validly and possible ethical concerns (focusing on the use of deception in recruitment) as they relate to this study.

In the next chapter, Chapter 4, I present the details of the implementation of the data collection, data processing, and analysis outlined in the current chapter. Further, I present the results of the data analysis. Later, in Chapter 5, I interpret the results of the data analysis as well as discuss the limitations and implications of the current study's finding.

Chapter 4: Results

The debate about the impact of appropriated slurs within social groups has been largely unexamined in the social sciences literature. Research on slurs and appropriated slurs has demonstrated some perceived empowerment and control when people use appropriated slurs in specific social situations (Galinsky et al., 2013). However, such research lacks measurement related to possible impacts on underlying cognitive processes or resulting behavior outcomes. Moreover, in the stereotype threat literature, researchers have focused on a wide range of contextual frames that invoke stereotype threat. Thus far, stereotype threat research has not been conducted examining if appropriated slurs act as a contextual frame within the stereotype threat model. The purpose of this study was to address these gaps in the knowledge about appropriated slurs by examining how exposure to appropriated slurs influences stereotype threat in Black American participants. I employed the stereotype threat model to quantitatively examine differences in stereotyped domain behavior and stereotype activation between participants who were exposed and participants who were not exposed to the selected appropriated slur. As such, the statistical procedures and results of this study will be presented to answer the following:

RQ1: What is the difference in academic test performance between Black American adults exposed to appropriated slur and Black American adults not exposed to appropriated slurs?

$H_0$1: Black American participants exposed to appropriated slurs will exhibit similar scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants not exposed to appropriated slurs.

$H_1$1: Black American participants exposed to appropriated slurs will exhibit significantly lower scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants not exposed to appropriated slurs.

RQ2: What is the difference in academic task performance between Black American adults who exhibit higher racial identity and Black American adults who exhibit lower racial identity?

$H_0$2: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, will exhibit similar scores on the academic task, as measured by the selected items of the SAT, compared to Black American participants who exhibit low racial identity scores.

$H_1$2: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, will exhibit significantly higher scores on the academic task, as measured by the selected items of the SAT, compared to Black American participants who exhibit low racial identity scores.

RQ3: To what extent does racial identity moderate the effects of appropriated slurs on academic test performance?

$H_0$3: Black American participants with high racial identity scores, as measured by the MIBI centrality scale, who are exposed to appropriated slurs will exhibit similar

scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants with low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

$H_1$3: Black American participants with high racial identity, as measured by the MIBI centrality scale, who are exposed to appropriated slurs will exhibit significantly lower scores on the academic performance task, as measured by the selected items of the SAT, compared to Black American participants with low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

RQ4: What is the difference in the cognitive activation of negative stereotypes toward one's own racial group between Black American adults exposed to appropriated slur and Black American adults not exposed to appropriated slurs?

$H_0$4: Black American participants exposed to appropriated slurs will exhibit similar negative racial stereotype activation, as measured by the word fragment task, compared to Black Americans participants not exposed to appropriated slurs.

$H_1$4: Black American participants exposed to appropriated slurs will exhibit significantly more negative stereotype activation, as measured by the word fragment task, compared to Black American participants not exposed to appropriated slurs.

RQ5: What is the difference in the cognitive activation of negative stereotypes toward one's own racial group between Black American adults who exhibit higher racial identity and Black American adults who exhibit lower racial identity.

$H_0$5: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, will exhibit similar negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants who exhibit low racial identity scores.

$H_1$5: Black American participants who exhibit high racial identity scores, as measured by the MIBI centrality scale, will exhibit significantly lower negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants who exhibit low racial identity scores.

RQ6: To what extent does racial identity moderate the effects of exposure to appropriated slurs on the cognitive activation of negative stereotypes toward one's own racial group in Black American adults?

$H_0$6: Black American participants with high racial identity scores, as measured by the MIBI centrality scale, who are exposed to appropriated slurs will exhibit similar negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants low racial identity scores who are exposed to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

$H_1$6: Black American participants with high racial identity, as measured by the MIBI centrality scale, who are exposed to appropriated slurs will exhibit significantly higher negative racial stereotype activation, as measured by the word fragment task, compared to Black American participants with low racial identity scores who are exposed

to appropriated slurs and Black American participants with high or low racial identity scores who are not exposed to appropriated slurs.

In this chapter, I will first discuss data collection procedure including the timeframe of data collection and any discrepancies in the data collection plan. Then, I will report the sample descriptive statistics, statistical assumption diagnostics, and statistical analysis findings of the statistical models used to evaluate the proposed hypotheses. Additionally, I will discuss the consideration and exclusion of potential covariates collected during the current study.

Data Collection

Time Frame, Recruitment, and Response Rates

Data collection for the current study began in September 2020 and continued through October 2020; 239 participants accessed the online survey via the Survey Monkey WebLink Collector URL posted on social media platforms, the Walden University participant pool, and the Cloud Research platform. The initial 7 weeks of recruitment were conducted through the Walden University Participant Pool website and recruitment posts through social media platforms (i.e., Facebook, LinkedIn, and Twitter). Over the course of this initial period, 44 participants consented to participate and two provided data that were suitable for analysis based on collection monitoring analyses I conducted at that time.

Considering the low recruitment and completion rates during this period, I submitted and received approval for a change in recruitment procedures to the Walden University IRB. The new recruitment procedure leveraged a partnership with Cloud

Research (powered by Amazon TurkPrime) to reach a wider pool of participants. Further, the new procedure included nominal monetary ($5.00) compensation for participating. The Cloud Research recruitment effort was deployed early October 2020 for approximately 1 week. During this time, 151 participants consented to participate and completed the study.

I terminated all recruitment efforts in October 2020. Over the course of the entire recruitment period across both recruitment platforms, 239 participants consented to participate. Of those who consented, 165 participants agreed to submit their responses after being debriefed and 132 participants indicated that they identified as Black American as defined in this study. After filtering the data based on consent, debrief acceptance, and self-identification, the final sample size reached 130 (a response rate of 54.4%); however, this sample size was reduced further during the data screening process, discussed below.

Emergence of Adverse Events

Over the course of the data collection phase of the study, there were no reports of psychological harm or other adverse events reported by participants. Further, I received only one email from a participant who commented on a typo in the consent form, which was corrected immediately.

Data Screening and Cleaning

Prior to applying the statistical models, I screened and cleaned the data for accuracy and missingness. To assess the accuracy of the data, I conducted a series of minimum and maximum value calculations for each variable. As expected, all scores fell

within the possible range for each variable. Next, I examined potential response biases in the data; specifically, I assessed the participants' response patterns for evidence of nondifferentiation bias and extreme completion speeds. Nondifferentiation bias is failure of participants to differentiate between items with their given answer (Lavrakas, 2013); for instance, participants may provide the same response option to all the questions on a survey (i.e., straight lining).

To assess potential nondifferentiation bias, I computed the standard deviations for each participants' responses to the academic performance task and the racial identity scale. Standard deviations of multiple-choice surveys that equal zero suggest a participant simply selected all the same response rather than mindfully processing the items (Leiner, 2019). I observed no participants with a standard deviation of zero for the academic performance task (minimum = .52). For the racial identity scale, I observed one participant whose MIBI standard deviation equaled zero; further, this participant also produced an elapsed completion time of 8 minutes. This participant was excluded under the extreme completion speed criteria.

Rapid completion time does not necessarily indicate that collected data are poor quality (Leiner, 2019). For instance, participants may be experts on the topics being assessed, leading to quick answers. However, if there is no reasonable explanation for extremely rapid completion times, it can be assumed that participants with such response times did not carefully or accurately answer the questions. Considering the nature of the mathematical (10 items) and word-fragment (28 items) tasks in the current study, it can be assumed that participants with extraordinarily short completion times did not provide

valid data. To address this issue, I excluded participants based on the following elapsed completion time criteria. I calculated each participant's elapsed completion time for the entire sample (M = 31.68, SD = 14.77), the difference between the time at which a participant completes the research activities by exiting the website and the time at which that participant began the research activities by entering the research website. The entry and exit times were collected by SurveyMonkey's metadata variables.

While it is certainly plausible that some participants took longer to complete the research activities, it is unlikely that participants with extremely short completion times would have been able to complete the tasks mindfully and accurately. To eliminate participants who may have rushed through the research activities, I excluded participants whose elapsed time was less than one standard deviation from the mean (< 16.91 minutes). This exclusion also eliminated participants who met the nondifferentiation screening criteria previously discussed. While this exclusion likely decreased the potential effects of response biases, it reduced the overall sample size from 130 to 118 participants, which is less than my predetermined target sample size (N = 128) based on the power analysis discussed in Chapter 3. The decreased sample size will impact the statistical power of the statistical models used to evaluate the stated hypotheses in the current study. I further discuss the reduction of statistical power and its potential effects on the current study in subsequent sections of this chapter.

As discussed in Chapter 3, I examined missingness in the collected data by employing Little's MCAR test for the items related to the key variables (i.e., academic performance task, word-fragment task, and the MIBI). The results produced all

nonsignificant outcomes, $p > .05$, indicating that any missing data are missing completely at random. There were no discernible patterns or specific issues concerning missing data points. To further assess missingness in the data related to the key variables, I examined the rate of missing data points for items related to each key variable. A cumulative percentage across all three scales for the key variables was calculated and any participant exhibiting more than 20% of their data missing would be eliminated; however, the maximum missing percentage reached 10.34% and data from all 118 participants were retained.

Screening Covariates

In any study, characteristics of participants or situational factors that are not the primary focus of the researchers' interests may still contribute to the statistical outcomes of the study; such extraneous variables are known as *covariates*. The potential impact of extraneous variables in research can be addressed by experimental control or statistical control (Ott & Longnecker, 2015). Experimental control is applied through experimental design procedures such as random assignment. Statistical control involves measuring potential variables and incorporating them into the statistical analyses. Both approaches for extraneous control are different means to the same end. In this study, I employed random assignment of participants to the experimental treatments (exposure to slurs or no exposure), thus exercising direct experimental control over possible extraneous variables.

However, I also collected additional information from participants, including demographic information (gender, education attainment, and age) and their perceptions of prior test performance in school to ensure participants' characteristics beyond the key

variables of interest did not impact the statistical results. I did not formally hypothesize covariates in the current study; however, I did screen each of the additional variables collected to assess their potential effect on the dependent variables. For a suspected extraneous variable to qualify as a covariate, it must produce a statistical effect on the key dependent variable (Ott & Longnecker, 2015). As such, I produced appropriate statistical models using each potential covariate as the sole independent variable with each of the key dependent variables. I set the criteria that, if these models produced statistically significant results, indicating an effect on the dependent variable, I would include that specific variable in the primary ANOVA models as a covariate.

I conducted simple linear regressions to assess participants' age as a potential covariate. Age did not exhibit a statistically significant relationship with the academic performance task scores or the word-fragment task scores ($p > .05$). I conducted t-tests to assess participants' gender as a potential covariate. I found no significant differences between male and female participants along the academic performance task scores or the word-fragment task scores ($p > .05$). To assess both education level and the participants' perceptions of their past test performance, I conducted a series of one-way ANOVAs. Because some education response options produced small groups (e.g., doctoral degree), I collapsed education into three primary groups: high school plus ($N = 40$), undergraduate degree ($N = 56$), and graduate degree ($N = 22$). Participants in the high school plus category reported they either attained a high school diploma or some college but no degree. Participants in the undergraduate degree category reported attaining either an associate's degree or a bachelor's degree. Participants in the graduate degree group

reported attaining a master's degree or a doctorate. Similarly, I collapsed the response groups into three groups to reach reasonable group sizes for the ANOVA. The agree group (N = 77) contained participants who either responded *strongly agree* or *agree* to the question about prior test performance. The disagree group (N = 24) contained participants who either responded *strongly disagree* or *disagree* to the question. The neither group (N = 17) contained participants who responded *neither agree nor disagree*. Each of the one-way ANOVAs for education and perceived test performance failed to reach statistical significance for either dependent variable. Considering the lack of statistically significant results produced be each potential covariate screening, I disqualified age, gender, educational attainment, and perceptions of prior test performance in school as covariates in the formal statistical models used to address the stated hypotheses in this study.

Statistical Results

Descriptive Statistics

The final sample size (N = 118) consisted of participants who were randomly assigned to either the control group (N = 58) or the experimental group (N = 60) in which they viewed a video without the appropriated slur or with the appropriated slur, respectively. The current sample consisted of female (N = 62) and male (N = 56) adults over the age of 18 (M = 37.94 years, SD = 11.50).

Table 1 shows the education distribution of the current sample. The largest educational groups included participants with some college (23.7%) and participants with a bachelor's degree (32.2%). I utilized a chi-square test to determine how the

representative the current sample was of the target population. That is, I compared the rates of education in the current sample to the rates of education among Black Americans according to the American Community Survey (American Community Survey, 2019). Overall, the chi-square test was significant, $p < .01$, indicating a difference among the rates of the education groups.

To understand this difference further, I examined the adjusted residuals for each group. If an adjusted residual was greater than the absolute value of 1.96, I evaluated it as statistically significant (Kateri, 2014). In addition to comparing the actual rates of the education groups, the sign of the adjusted residual (positive or negative) indicated if the sample education group was under- or overrepresented compared to the population education group. There were significant differences between the sample and population rates for high school graduates, bachelor's degree, and graduate degrees groups. Specifically, high school graduates were underrepresented in the sample compared to the population while bachelor's and graduate level education groups were overrepresented in the sample. The *some college, no degree* and associate's degree level groups in the sample were not significantly different from the population.

As a control question related to academic task performance, participants were asked if they always did well on tests in school. More than half (51.7%) of the participants responded with *agree* to this question. Table 2 shows the distribution of the responses to this question.

Table 1

Sample Education Demographics

	Sample		ACS 2019[a]		Adjusted residuals for sample group
	n	*%*	*n*	*%*	
High school graduate or equivalent	12	10.2	8732838	57.96	–6.0
Some college, no degree	28	23.7	6518381	43.26	–.9
Associate's degree	18	15.3	2384596	15.83	1.9
Bachelor's degree	38	32.2	3811181	25.29	4.8
Graduate/professional degree	22	18.6	2352885	15.62	3.2

[a] American Community Survey (2019) conducted by the U.S. Census Bureau

Table 2

I Always Did Well on Tests In School

	n	*%*
Strongly agree	16	13.6
Agree	61	51.7
Neither agree nor disagree	17	14.4
Disagree	21	17.8
Strongly disagree	3	2.5

Statistical Assumptions

Statistical assumptions refer to the requirement of statistical analyses that the sample data approximates the features of the population from which it is sampled (Cohen, 2013). The violation of any given assumption can reduce the robustness and accuracy of statistical analyses (Toothaker, 1993). In this section, I present the evaluation of the statistical assumptions for both the academic performance task model and the word-fragment ANOVA models before discussing the statistical results of those models in the next section.

I considered three main statistical assumptions—normality, homogeneity of variances, and influential outliers—related to the ANOVA models in the current study. In his classical text, Toothaker (1993) discussed that normality is among the least likely assumptions to be met in a multiple comparison analyses (e.g., ANOVAs). He argued that expectations of normality may not be realistic because very few dependent variables that are of interest to social science researchers will display normality. Further, Toothaker posited that researchers are typically unaware of the actual attributes of the population distribution and even extremely non-normal data can be obtained from normal population. In other words, meeting the assumptions of normality may not be as vital as the other assumptions. Nevertheless, I assessed the assumption of normality by both examining histograms across subgroups and by calculating the KS statistic for each dependent variable.

Histograms for the academic performance task scores by condition group (experimental and control) appeared approximately normal. Further, the Shapiro-Wilk test for the condition groups produced nonsignificant results ($p > .001$). Similarly, the histograms for the academic performance task by racial identity group (lower and upper MIBI) appeared approximately normal and the Shapiro-Wilk test was non-significant ($p > .001$) for both groups. Together, these results suggest that the assumption of normality was met for the academic performance task. In contrast, the histograms for the word-fragment task scores by condition group and racial identity group appeared to deviate from an approximately normal distribution. A slight positive skew appeared in the control condition group and both racial identity groups. The Shapiro-Wilk results for this

dependent variable were significant ($p < .001$) across all groups. The results indicate that the assumption of normality for the word-fragment task score was violated. Despite the deviation of the word-fragment task scores from normality, I hold that the robustness of ANOVA models with moderate size samples will compensate for this issue (Toothaker, 1993).

The overall purpose of an ANOVA test is to determine whether each of the groups of interest have a common mean for the dependent variable (Gastwirth et al., 2009). Moreover, the *F-test* used to test for common means across groups assumes that each group will have common variances. If common variances are not true, the accuracy of the *F-test* will falter. Therefore, a method to assess the equality of variance is needed to ensure accurate results in an ANOVA model. One such method is the Levene's test, which evaluates the null hypothesis that dependent variable variances are the same across the levels of the independent variables. The results of the Levene's tests in this study was non-significant ($p > .05$) for the academic performance task and word-fragment task models, indicating this assumption was met.

I assessed the assumption that no influential outliers appear in the data using the Cook's distance measure. In Cook's (1979) seminal work, he demonstrated that large D (for distance) values indicates the presence of influential outliers in data that would skew the results of a linear model. The calculated Cook's distances in the current dataset did not exceed .07, indicating no influential outliers were present.

Statistical Power

Beyond the observed outcomes of any given dataset, the reality of a statistical effect existing can be true or false (Cohen, 2013). As such, when hypothesis testing, a researcher's decision to reject or retain the null hypothesis can be a correct or incorrect decision. When discussing the probability that a researcher will correctly diagnostic the null for an effect that does not truly exist in the population, it is referred to as statistical power. On one hand, if a researcher rejects the null when there is no true effect, he or she has made a correct inference about the data. On the other hand, if a researcher retains the null in the same situation (no true effect), a Type II error (i.e., a false negative) has been committed.

The statistical power of a test, represented by the value of beta, determines the probability of making a Type II error (Cohen, 2013). As discussed in Chapter 3, I selected the initial beta of .80 based on social science standards to determine an appropriate sample size for this study. This beta value declared that there was an 80% chance of detecting a true effect if I rejected the null hypotheses. Inversely, this standard indicated that there is a 20% chance of committed a Type II error if I retained the null hypothesis. While the standard convention of statistical power (b = .80) was employed to determine the ideal sample size for the current study, the final sample size fell short of that goal due to exclusion criteria previously discussed (i.e., non-differentiation and elapsed time issues).

As previously discussed, the target sample size (N = 128) was reduced to 118 participants through response bias screening. The reduced sample size will reduce the

actual statistical power of the current statistical analyses. I conducted a post hoc power analysis using G*Power by entering the previously discussed parameter—alpha (.05), effect size (.25)—and the actual sample size used in the analyses (N = 118) to evaluate the actual statistical power of the current study. The power analysis results indicate the actual statistical power was .77 indicating that there is a slightly increased chance of committed a Type II error when retaining the null hypothesis. That is, there is a 23% chance that I did not detect a true effect—exposure to appropriated slurs decreasing academic task performance and increasing stereotype activation—by retaining the null hypothesis in this study.

Even though a decrease in statistical power is not ideal in empirical research, the nominal increase in the probability of a Type II error in the current study can still be acceptable. That is, Cohen (2013) noted that a beta that was substantially smaller than .80 would pose too much of a risk of Type II error. In this study, I do not perceive the .03 decrease in power materially smaller to pose a major concern.

Statistical Analysis Findings

In the following sections I will discuss statistical results including descriptive statistics for the model variables as well as statistical significance, post-hoc comparisons, and effect sizes for each ANOVA model for the current study. Further, I will evaluate each of the current research questions and their associated hypotheses based on the presented statistical results.

Academic performance task results

I employed a two-way ANOVA model using the academic performance task as the dependent variable to examine research questions one through three. I calculated the academic performance score by summing the number of math items participants correctly solved (M = 5.25, SD = 2.61). The independent variables in this model were experimental condition and MIBI median split groups. Participants were randomly assigned to the control condition (N = 58) or the experimental condition (N = 60) and were grouped into the lower-MIBI (N = 58) or the upper-MIBI (N = 60) group based on their total MIBI score in relation to the overall sample median for the current sample (*median* = 29).

The main effect of slur condition was not significant, $p > .05$. Therefore, I retained the null hypothesis for research question one indicating that there was no statistically significant difference in the academic performance task scores between participants who were exposed to the appropriated slur (M = 5.23) and participants who were not exposed to the appropriated slur (M = 5.26). The main effect of MIBI level was not significant, $p > .05$. Therefore, I retained the null hypothesis for research question two indicating no statically significant difference in the academic performance task scores between lower-MIBI (M = 5.71) and upper-MIBI (M = 4.81) groups. Further, the interaction effect of condition and MIBI was not significant, $p > .05$. Therefore, I retained the null hypothesis for research question three indicating no statistically significant difference in the academic performance task scores exist among the slur conditions and MIBI subgroups.

Table 3

Two-Way ANOVA: Academic Performance Task

Source	Type III sum of squares	df	Mean square	F	Sig.	Partial eta squared
Corrected Model	26.707[a]	3	8.902	1.319	.272	.034
Intercept	3246.096	1	3246.096	481.112	.000	.808
Condition	.091	1	.091	.013	.908	.000
MIBI_median_split	24.093	1	24.093	3.571	.061	.030
Condition * MIBI_median_split	2.373	1	2.373	.352	.554	.003
Error	769.166	114	6.747			
Total	4043.000	118				
Corrected Total	795.873	117				

a. R Squared = .034 (Adjusted R Squared = .008)

Table 4

Descriptive Statistics for Academic Performance Task

				95% confidence interval	
	n	M	SD	LB	UB
Control					
Lower MIBI	31	5.548	2.604	4.624	6.473
Upper MIBI	27	4.926	2.598	3.936	5.916
Total	58	5.237	2.604	4.560	5.914
Experimental					
Lower MIBI	27	5.889	2.598	4.899	6.879
Upper MIBI	33	4.697	2.597	3.801	5.593
Total	60	5.293	2.610	4.625	5.961
Total					
Lower MIBI	58	5.719	2.604	5.041	6.396
Upper MIBI	60	4.811	2.601	4.144	5.479
Total	118	5.246	2.608	4.775	5.716

Word-Fragment Task Model

I employed a two-way ANOVA model using the word-fragment task scores as the dependent variable to examine research questions four through six. I calculated the word-fragment task scores by summing the number of stereotype-related word matches entered by the participants (M = 1.48, SD = .93). The independent variables in this model were the experimental condition and the MIBI median split groups. Participants were randomly assigned to the control condition (N = 58) or the experimental condition (N = 60) and were grouped into the lower-MIBI (N = 58) or the upper-MIBI (N = 60) group based on their total MIBI score in relation to the overall sample median for the current sample (*median* = 29).

The main effect of slur condition was not significant, $p > .05$. Therefore, the null hypothesis for research question one was retained indicating that there is no statistically significant difference in the word-fragment task scores between participants who were exposed to the appropriated slur (M = 1.60, SD = .96) and participants who were not exposed to the appropriated slur (M = 1.36, SD = .89). The main effect of MIBI level was not significant, $p > .05$. Therefore, the null hypothesis for research question two was retained indicating no statistically significant difference in the word-fragment task scores between lower-MIBI (M = 1.44, SD = .94) and upper-MIBI (M = 1.51, SD = .93) groups. Further, the interaction effect of condition and MIBI was not significant, $p > .05$. Therefore, the null hypothesis for research question three was retained indicating no statistically significant difference in the word-fragment task scores exist among the slur conditions and MIBI subgroups.

Table 5

Two-Way ANOVA: Word-Fragment Task

Source	Type III sum of squares	df	Mean square	F	Sig.	Partial eta squared
Corrected Model	2.055[a]	3	.685	.786	.504	.020
Intercept	258.246	1	258.246	296.145	.000	.722
Condition	1.621	1	1.621	1.859	.175	.016
MIBI_median_split	.074	1	.074	.084	.772	.001
Condition * MIBI_median_split	.317	1	.317	.363	.548	.003
Error	99.411	114	.872			
Total	361.000	118				
Corrected Total	101.466	117				

a. R Squared = .020 (Adjusted R Squared = -.006)

Table 6

Descriptive Statistics for Word-Fragment Task

				95% confidence interval	
	n	M	SD	LB	UB
Control					
Lower MIBI	31	1.290	.935	.958	1.623
Upper MIBI	27	1.444	.935	1.088	1.800
Total	58	1.367	.937	1.124	1.611
Experimental					
Lower MIBI	27	1.630	.935	1.274	1.986
Upper MIBI	33	1.576	.963	1.254	1.898
Total	60	1.603	.937	1.363	1.843
Total					
Lower MIBI	58	1.460	.937	1.216	1.703
Upper MIBI	60	1.510	.937	1.270	1.750
Total	118	1.483	.931	1.315	1.651

Summary

From the data analyses produced in the current study, I failed to find statistical evidence for the research hypothesis. These results demonstrated no observable effect of

exposure to the appropriated slur on academic performance task and word-fragment task scores. Racial identity appeared to have no significant effect on academic performance task and word-fragment task scores. Moreover, the interaction between the appropriated slur condition and racial identity did not produce an observable effect on academic performance task and word-fragment task scores. Taken together, these results point to several possible conclusions about the impact of exposure to appropriated slurs on stereotype threat. In Chapter 5, I will present the interpretation of the results discussed in the current chapter to provide a final summary of this study. As a part of the final discussion, I will present the limitations, recommendations for future research, and the implications of the current study on social change.

Chapter 5: Discussion

In this study, I investigated a long-standing debate on the impact of appropriated slurs by addressing two gaps in the appropriated slur and stereotype threat literature (Croom, 2016; Gaucher et al., 2015). As such, I focused on how exposure to appropriated slurs might impact cognitive processes and subsequent behaviors; I attempted to address the question of whether appropriated slurs are harmful or beneficial to ingroup members. I examined the effects of exposure to appropriated slurs on stereotype threat to elucidate the impact of appropriated slurs on the cognitive and behavioral experiences of ingroup members. Specifically, I used a randomized experiment with a posttest-only with control group design to test how exposure to an appropriated slur affected two components of stereotype threat—negative stereotype activation about one's own group and academic task outcomes—in Black American participants. Results revealed no statistically significant differences for stereotype activation or academic task performance between participants exposed to appropriated slurs and participants not exposed to appropriated slurs. Further, racial identity had no impact on these results. Overall, the results suggest that mere exposure to an appropriated slur does not invoke stereotype threat in Black American adults. I will discuss these finding further in the following section.

Interpretation of the Findings

I examined two fundamental components of stereotype threat—stereotyped domain performance and stereotype activation (Steele & Aronson, 1995)—in the current study to explore the effects of exposure to appropriated slurs on stereotype threat. In the following sections, I discuss my decisions on the stated null hypotheses based on

ANOVA model results and I discuss an interpretation of the combined results in the context of the stereotype threat model.

Stereotyped Domain Performance

The first three research questions addressed the stereotyped domain performance component of stereotype threat by analyzing participants' scores on a brief academic performance task after being exposed or not exposed to the appropriated slur. Moreover, racial identity was considered as a possible moderating factor between appropriated slur exposure and academic performance task outcomes. I did not obtain statistically significant outcomes for academic task performance for RQ1 and RQ3. Participants exposed to the appropriated slur performed similarly on the academic performance task to participants not exposed to the appropriated slur (RQ1). Participants in the lower racial identity and upper racial identity groups also performed similarly on the academic performance task (RQ2). Further, there was no significant interaction between racial identity and slur exposure on the academic task scores (RQ3). These findings indicate that the effects of appropriated slur exposure on academic performance task scores is not dependent on racial identity. All four subgroups (i.e., slur condition-low MIBI, slur condition-high MIBI, control-low MIBI, and control-high MIBI) exhibited similar scores on the academic performance task.

Stereotype Activation

The next three research questions addressed the stereotype activation component of stereotype threat by analyzing the amount of stereotype-related words participants produced in a word-fragment task after being exposed or not exposed to the appropriated

slur. As with the academic performance task model, racial identity was considered as a possible moderating factor between appropriated slur exposure and stereotype activation. I did not observe statistically significant outcomes for the word-fragment research questions nor did racial identity produce statistically significant results. Participants who were exposed to appropriated slurs produced, on average, the same number of stereotype-related words as participants not exposed to the appropriated slur (RQ4). Further, the number of stereotype-related word fragments was not affected by racial identity (RQ5) or the interaction between racial identity and slur exposure (RQ6). This finding indicates that the effects of appropriated slur exposure on negative stereotype activation is not dependent on racial identity. All four subgroups (i.e., slur condition-low MIBI, slur condition-high MIBI, control-low MIBI, and control-high MIBI) produced similar numbers of stereotype-related words on the word-fragment task. While I retained all the stated null hypotheses in the current study, the findings provide some insights into gaps in two areas of research: stereotype threat and appropriated slurs.

General Discussion

Contextual cues, stereotype activation, and stereotyped domain behavior represent the three fundamental components of the stereotype threat model (Steele & Aronson, 1995). In the context of this model, these results taken together suggest that exposure to appropriated slurs does not invoke stereotype threat. In this study, I employed exposure to an appropriated slur as the contextual cue to examine its effects on the other components of stereotype threat. I operationalized stereotype activation and stereotyped domain behavior as a word-fragment task and academic task, respectively. As discussed,

I observed no main effect of exposure to the appropriated slur on either stereotype activation or academic performance task. Exposure to the appropriated slur did not produce stereotype activation nor stereotyped domain behavior (i.e., academic performance task) outcomes different than when not exposed to the appropriated slur in the current sample of participants. Taken together, these findings suggest two possible insights into appropriated slurs. First, the null effects suggest that exposure to appropriated slurs may not be sufficient contextual cues in the stereotype threat paradigm. Mere exposure to an appropriated slur may not invoke stereotype threat and decrease performance in the stereotyped domain. Second, these findings suggest that appropriated slurs, at least mere exposure to the terms, may not have harmful impacts on ingroup members. To my awareness, this is the first stereotype threat study to employ appropriated slurs as a contextual cue, thus addressing a gap in the area.

In the appropriated slur literature, appropriated slurs are implicated as harmful to ingroup psyche or empowering to the ingroup through the transformation of the negative power of the original slur into a positive ingroup term (Archer, 2015; Galinsky et al., 2013; Gaucher et al., 2015). However, such researchers failed to address measured cognitive and behavioral outcomes related to exposure to appropriated slurs. This study presents support for the perspectives that refurbishing slurs decreases their negative connotations and reduces negative target group outcomes (e.g., Galinsky et al., 2013; Gaucher et al., 2015). This outcome suggests that the appropriated slur did not possess the required connection to the underlying negative stereotypes needed to cognitively trigger stereotypes. Thus, the appropriation of the slur may have broken or diminished the

link to the negative connotations of the original slur. These results align with prior theoretical and empirical work in which researchers posit the negativity of the original slur is severed by the appropriation process. For instance, Croom (2013), Galinsky et al. (2013), and Archer (2015) have contended that the contextual flexibility of slurs lends itself to developing new rules and applications in which the slur can become nonoffensive and supportive.

Galinsky et al. (2013) offered empirical support that self-labeling using an appropriated slur increases the speaker's and observer's perceptions of social power in the speaker of the appropriated slur; such perceptions contrast the intent of original slurs, such as social oppression (Croom, 2015). Archer (2015) also proposed that the contextual flexibility of slurs allows people to change the social meaning and rules of the word, developing a new term separate from its root. Further, Gaucher et al. (2015) found that women reported higher levels of self-assurance and lower levels of fear and were less likely to endorse gender-based stereotypes when exposed to appropriated slurs in supportive situational contexts. Similarly, the results of this study indicate that exposure to an appropriated slur does not increase negative stereotype activation, suggesting the link between the negativity of the original slur and the appropriated slur has been weakened.

While the current results suggest appropriated slurs may not have negative impacts, I also did not observe evidence that suggests positive effects on these outcomes. Galinsky et al. (2013) and Gaucher et al. (2015) claimed that appropriated slurs may act as term for empowerment and increased self-assurance. These positions align with the

perspective that appropriated slurs are beneficial to the social group that was targeted by the original slur. This study's results, however, did not demonstrate decreased negative stereotype activation or increased academic task performance. That is, exposure to the appropriated slur did not appear to benefit the participants' outcomes in this study.

Limitations of the Study

The current study, as with many empirical studies, was not without limitations that could have impacted the validity and generalizability. These limitations included control over the testing environment, the context of the exposure to the appropriated slur, and the statistical power. In the following section, I briefly discuss these limitations.

Two of the possible limitations, control over testing environment and situational environment, for the current study rise from its online data collection format. As participants engaged in the research activities remotely, in settings of their choosing, I was unable to monitor the key variable tasks such as the academic performance task. Such tasks in similar studies were monitored in-person and under time constraints by the research staff. In the current study, I was unable to monitor and limit participants from completing the tasks using help from other resources such as the internet or other people. Further, normal academic testing environments typically have a time constraint; however, I did not apply such controls to the research activities due to lack of technical capabilities of the online tools. The unconstrained time limit may have relieved some of the natural stress or urgency that participants experienced in previous stereotype threat studies.

Similarly, traditional laboratory research allows the researcher to control the situational environment in which participants engage the research activities. Because

participants were able to participate remotely, I had no control over their environments. Specifically, in the current study, I attempted to manipulate contextual cues (exposure to the appropriated slur) to invoke stereotype threat. However, a contextual cue does not exist in a vacuum and may be influenced by other existing situational cues. For example, Pennington et al. (2019) discussed that gender-based stereotype threat primes could have been unintentionally compounded by the gender of an experimenter who interacted with participants. That is, a female researcher interacting with participants may have buffered the negative stereotype prime of females performing poorly in math.

Similarly, in the current study, Black American participants may have been in environments (e.g., their homes) in which other ingroup members were present and offered a moderating influence on the stereotype threat prime. The environmental context in which an appropriated slur is used appears to play a role in how it is perceived by others. For instance, Galinsky et al. (2013) and Gaucher et al. (2015) found that appropriated slurs used in social change or group solidarity increased the positive perceptions of appropriated slurs. However, I was unable to integrate the control of situational contexts into the current study.

Another potential limitation to the current study was statistical power. As previously discussed, statistical power refers to the likelihood of mistakenly retaining the null hypothesis (Cohen, 2013). The conventional and target statistical power that I established in Chapter 2 (.80), was used to establish the required sample size in the current study (N = 128). However, due to exclusion of some participant data, the final sample size fell short of that criteria (N = 118); thus, the statistical power was reduced. In

other words, the smaller sample size increased the possibility that I would erroneously retain the null hypotheses in the current study. That is, there is an increased possibility that I did not find an effect of exposure to appropriated slurs when one exists. Nevertheless, the decreased statistical power was not substantially lower than the conventional standard (.80) and the increase risk of a Type II error is nominal.

Additionally, this study may have included limitations such as sampling bias, response bias, and confounding variables. By using specific participant pool platforms, a sampling bias related to other characteristics of the participants may have been present. For instance, Walden University's participant pool and MTurks' registered participants may overrepresent certain demographics within the Black American community. In fact, I did observe that graduate degree education in the sample was at a higher rate when compared to the American Community Survey rates. This overrepresentation may have influence how well participant perform on the academic performance task or how they perceive appropriated slurs. However, as discussed in previous sections, I attempted to address the response bias and confounding variable limitations within the scope of the current research. For example, I analyzed the data for potential response bias and excluded such observations as well as eliminated possibly confounding variables through statistical analysis (discussed in the data screening section).

Recommendations

The results of the current study provide a glimpse into the complexity of appropriated slurs and stereotype threat as well as set the stage for further research to better understand the intricacies of how appropriated slurs impact cognitive and

behavioral outcomes. In the following sections, I briefly recommend directions for future research based on the findings in the current study.

Slurs and appropriated slurs are highly contextually flexible terms that can be used as terms of respect and endearment or as terms of disgust and hate. The meaning and value of such terms change within the contexts of different settings, speakers, and conversations. In this study, a Black American confederate posed as a comedian who integrated the selected appropriated slur into his materials. The appropriated slur was not directed at the viewer and was used in the context of talking about friends. In this contextual environment, the appropriated slur may have different meaning and cognitive impacts as opposed to other situations. I recommend that future research focus on changing the situational contexts in which an appropriated slur is used to better understand the possible differences in its impact across those situations. For example, future researchers may consider using appropriated slurs as a contextual cue in academic or professional settings rather than settings in which participants may feel more connected to their ingroup (e.g., at home online). Moreover, situational contexts could also include exposure to other participants who do not belong to the ingroup of the appropriated slur. That is, proximity to outgroup members may affect how participants process appropriated slur used by their ingroup (Stone et al., 1997).

While theoretical models of slur appropriation are seemingly applicable to all social groups, it would be naïve to assume that all social groups, cognitively and socially, process appropriated slurs equally. I also recommend that further research on the effects of appropriated slurs expand to other social groups (racial, ethnic, gender, sexual

orientation, age cohorts, etc.) that used such terms. Moreover, as appropriated slurs may change in application and context over time, I recommend further research to explore the influence of age on the impacts of appropriated slur.

Lastly, I recommend that future research on the effects of appropriated slurs employ measures of cognitive activity and behavioral outcomes that produce richer data. One such possibility for cognitive measures is the Implicit Associations Test (IAT). Instruments such as the IAT allow researchers to infer implicit bias and stereotype activation based on participant behaviors such as reaction times to specific stereotype-related behaviors (Greenwald et al., 1998). Stereotype threat researchers have posited that contextual cues that invoke negative stereotypes would activate implicit associations even when the targets do not endorse the stereotypes (Galdi et al., 2014; Schmader et al., 2015). Such research has detected significant effects of contextual cues on IAT outcomes; that is, participants exposed to stereotype-related contextual cue exhibited increased levels of implicit bias toward their own social group. Such findings suggest the activation of negative stereotypes in participants when exposed to a related contextual cue. Research on the effects of appropriated slurs using sophisticated instruments such as the IAT may produce more nuanced understanding of the cognitive processes underlying exposure and used of appropriated slurs.

Similarly, cognitive-based instruments that measure functions such as attention, inhibition, and working memory can provide a nuanced understanding cognitive-behavioral impacts of appropriated slurs. For instance, Chalabaev et al. (2016) utilized both academic mathematics task and Stroop tests, a cognitive performance measure, to

examine how embodied triggers influence stereotype threat. Liu et al. (2017) also leveraged cognitive measures to investigate the relationship of self-concepts, working memory, and stereotype threat in older adults. I recommend that future research use similar approaches in which such technical instrumentation of cognitive performance is paired with convention performance tasks (i.e., academic performance) to provide deeper insights into the influence of appropriated slurs. Moreover, future researchers should consider qualitative or mixed methods approaches to exploring the effects of appropriated slurs to gain a deep and rich understanding of this type of communication.

Implications

In this study, I explored the impact of exposure to appropriated slurs on stereotype threat among Black American adults. The current results found no effect of appropriated slurs on negative stereotype activation nor academic task performance as measured by word-fragment and selected SAT math problems, respectively. These results suggest that exposure to appropriated slurs do not have a negative impact on the components of stereotype threat and may not have negative impacts on ingroups. In the following sections I will discuss how these results may have positive social change implications in stereotype threat interventions and interpersonal communications.

Beyond breaching an under-researched area of cognitive impacts of ingroup use of appropriated slurs, the current research also contributes to the discussion of the value of appropriated slurs (Bianchi, 2014; Croom, 2015). Specifically, the current study suggests that exposure to appropriated slurs does not necessarily invoke negative stereotype activation or related behavioral outcomes. This appears to support slur

appropriation models that posit the appropriation process diminishes the impact of the negative context of the original slur (Galinsky et al., 2003). However, while no negative effects were observed, the current study did not produce positive effects of exposure to appropriated slurs. This is contrary to the stereotype lift or empowerment perspective that states appropriated slurs have positive effects beyond dismantling the negativity of the original (Gaucher et al., 2015). Moreover, the current results provide the beginnings of a deeper understanding of interpersonal communication and media consumption of social groups that employ appropriated slurs. That is, critiques of appropriated slurs in music and film can be better informed by the current results—specifically, that mere exposure to appropriated slurs may not have negative impacts on consumers. Similarly, parents and other caretakes (e.g., counselors, mentors, etc.) may better understand the impacts of music, film, and television products that include appropriated slurs on their children's cognition and behaviors.

As we learn more about the impacts of appropriated slurs, they may have functions in affirmation interventions to address stereotype threat such as brief writing interventions. The efficacy of self-affirmation interventions is inconsistent despite such interventions being often employed (Bratter et al., 2016). Despite some positive results, the outcomes of affirmation-based intervention for stereotype threat have been inconsistent (Jordt et al., 2017). Such complexity may lay in integrating social identity, especially identities targeted by negative stereotypes, into affirmation exercises. Typical self-affirmation approaches include asking participants to write about why specific values (e.g., spending time with friends, music, etc.) are important to them.

Other affirmation approaches ask participants to focus on their connection or value in the stereotyped domain (Bernacki et al., 2016). The underlying rationale to increase the participants' self-esteem. However, self-esteem also stems from social connections such as belonging to groups—even if the group membership is voluntary or assigned. As such, exploring the integration of group belongingness and values into affirmation intervention would be a valuable venture. Like individual affirmation interventions, group-based affirmations may also focus on social groups and domain-related groups such as racial groups or math groups, respectively. The current study may offer some initial insights into the integration of social group connectedness. If ingroup terms such as appropriated slurs do truly mitigate the connection to the negativity of the original slur, asking participants to write about their social groups using any ingroup terms they would naturally use may help compartmentalized negative stereotype threat activation.

Slurs are sociopsychological terms that utilize referential and expressive functions to bind together negative stereotypes and affective reactions against a target group (Croom, 2016). The true power of slurs lies in their ability to invoke their underpinning negative evaluation of the target and produce emotional reactions in both targets of the slur and others (Burkley et al., 2016; Fasoli et al., 2016; Jeshion, 2013b; Soral et al., 2018). Slurs, as such, are a highly viable and impactful component of social control or oppression. Humans as social beings, however, appear to resist such control by engaging social and group processes such as slur appropriation to reduce the negative impacts of the original slurs (Archer, 2015; Galinsky et al., 2013). While slur appropriation is

viewed as potentially empowering by some, others believe any form or connection to the original slur remains harmful to the targeted group (Croom, 2013).

Conclusion

In this study, I explored the value of appropriated slurs by testing the effects of exposure to an appropriated slur on stereotype threat in Black American adults. As one would expect from such a complex concept, the results presented more questions than clear cut answers. Overall, I found no statistically significant effects of exposure to appropriated slurs on stereotype activation or academic task performance—two main components of stereotype threat. These results may suggest that appropriated slurs do not have a negative impact on the hearer. However, I also did not observe positive effects (e.g., increased academic task performance) in the results. Nevertheless, these initial findings in a novel line of research offer a pathway to future research to elucidate the complexities of appropriated slurs. The contextual and semantic flexibility of appropriated slurs, for instance, offer areas of exploration to better understand how these terms function across different situations, conversations, group, and subgroup configurations (Bianchi, 2014). As appropriated slurs become a more common and deeply grained component of social life, researchers and social scientists should strive to theoretically and empirical understand their intricate nature and impacts on cognition and behavior.

References

Allan, K. (2015). When is a slur not a slur? The use of nigger in *Pulp Fiction*. *Language Sciences, 52*, 187–199. https://doi.org/10.1016/j.langsci.2015.03.001

Angle, J. W., Dagogo-Jack, S. W., Forehand, M. R., & Perkins, A. W. (2017). Activating stereotypes with brand imagery: The role of viewer political identity. *Journal of Consumer Psychology, 27*(1), 84–90. https://doi.org/10.1016/j.jcps.2016.03.004

Anis, Y., Krause, J. A., & Blum, E. N. (2016). The relations among mathematics anxiety, gender, and standardized test performance. *Research in the Schools, 23*(2), 28–37.

Archer, D. (2015). Slurs, insults, (backhanded) compliments and other strategic facework moves. *Language Sciences, 52*, 82–97. https://doi.org/10.1016/j.langsci.2015.03.008

Baldwin, T., de Marneffe, M.-C., Han, B., Kim, Y.-B., Ritter, A., & Xu, W. (2015). Shared tasks of the 2015 Workshop on noisy user-generated text: Twitter lexical normalization and named entity recognition. *Proceedings of the Workshop on Noisy User-Generated Text*, pp. 126–135. https://doi.org/10.18653/v1/w15-4319

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014

Bernacki, M., Nokes-Malach, T., Richey, J. E., & Belenky, D. M. (2016). Science diaries: A brief writing intervention to improve motivation to learn science. *Educational Psychology, 36*(1), 26–46.

https://doi.org/10.1080/01443410.2014.895293

Bianchi, C. (2014). Slurs and appropriation: An echoic account. *Journal of Pragmatics, 66*, 35–44. https://doi.org/10.1016/j.pragma.2014.02.009

Bratter, J. L., Rowley, K. J., & Chukhray, I. (2016). Does a self-affirmation intervention reduce stereotype threat in Black and Hispanic high schools? *Race and Social Problems, 8*(4), 340–356. https://doi.org/10.1007/s12552-016-9187-4

Brochu, P. M., & Dovidio, J. F. (2014). Would you like fries (380 calories) with that? Menu labeling mitigates the impact of weight-based stereotype threat on food choice. *Social Psychological and Personality Science, 5*(4), 414–421. https://doi.org/10.1177/1948550613499941

Burkley, M., Andrade, A., & Burkley, E. (2016). When using a negative gender stereotype as an excuse increases gender stereotyping in others. *Journal of Social Psychology, 156*(2), 202–210. https://doi.org/10.1080/00224545.2015.1090945

Casad, B. J., Hale, P., & Wachs, F. L. (2017). Stereotype threat among girls. *Psychology of Women Quarterly, 41*(4), 513–529. https://doi.org/10.1177/0361684317711412

Chalabaev, A., Radel, R., Masicampo, E. J., & Dru, V. (2016). Reducing stereotype threat with embodied triggers: A case of sensorimotor–mental congruence. *Personality and Social Psychology Bulletin, 42*(8), 1063–1076. https://doi.org/10.1177/0146167216651407

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Taylor & Francis Group.

Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American*

Statistical Association, *74*(365), 169–174. https://doi.org/10.1080/01621459.1979.10481634

Covarrubias, R., Herrmann, S. D., & Fryberg, S. A. (2016). Affirming the interdependent self: Implications for Latino student performance. *Basic and Applied Social Psychology*, *38*(1), 47–57. https://doi.org/10.1080/01973533.2015.1129609

Croom, A. M. (2013). How to do things with slurs: Studies in the way of derogatory words. *Language & Communication*, *33*(3), 177–204. https://doi.org/10.1016/j.langcom.2013.03.008

Croom, A. M. (2015). Slurs, stereotypes, and in-equality: A critical review of "How epithets and stereotypes are racially unequal." *Language Sciences, 52*, 139–154. https://doi.org/10.1016/j.langsci.2014.03.001

Croom, A. M. (2016). An assessment of the negative and positive aspects of stereotypes and the derogatory and nonderogatory uses of slurs. In A. Capone & J. L. May (Eds.) *Interdisciplinary studies in pragmatics, culture and society* (pp. 791–822). Springer

Davis, C., Aronson, J., & Salinas, M. (2006). Shades of threat: Racial identity as a moderator of stereotype threat. *Journal of Black Psychology*, *32*(4), 399–417. https://doi.org/10.1177/0095798406292464

Deviyanti, D. (2015). The role of women's identification with math and academic major in women's susceptibility to stereotype threat and stereotype lift. https://dspace.library.colostate.edu/handle/10217/167037

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial

test performance: A meta-analysis. *Learning and Individual Differences*, 47, 103–116. https://doi.org/10.1016/J.LINDIF.2015.12.018

Emerson, K. T. U., & Murphy, M. C. (2014). Identity threat at work: How social identity threat and situational cues contribute to racial and ethnic disparities in the workplace. *Cultural Diversity and Ethnic Minority Psychology*, 20(4), 508–520. https://doi.org/10.1037/a0035403

Emerson, K. T. U., & Murphy, M. C. (2015). A company I can trust? organizational lay theories moderate stereotype threat for women. *Personality and Social Psychology Bulletin*, 41(2), 295–307. https://doi.org/10.1177/0146167214564969

Fasoli, F., Paladino, M. P., Carnaghi, A., Jetten, J., Bastian, B., & Bain, P. G. (2016). Not "just words": Exposure to homophobic epithets leads to dehumanizing and physical distancing from gay men. *European Journal of Social Psychology*, 46(2), 237–248. https://doi.org/10.1002/ejsp.2148

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. Sage.

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.

Forbes, C. E., & Leitner, J. B. (2014). Stereotype threat engenders neural attentional bias toward negative feedback to undermine performance. *Biological Psychology*, 102, 98–107. https://doi.org/10.1016/J.BIOPSYCHO.2014.07.007

Frankfort-Nachmias, C., & Nachmias, D. (2015). *Research methods in the social sciences* (8th ed.). Worth.

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: when automatic associations disrupt girls' math performance. *Child Development*, 85(1), 250–263. https://doi.org/10.1111/cdev.12128

Galinsky, A. D., Hugenberg, K., Groom, C., & Bodenhausen, G. (2003). The reappropriation of stigmatizing labels: Implications for social identity. *Identity Issues in Group-Research on Managing Group and Teams*, 5, 221–256. http://www.emeraldinsight.com/doi/abs/10.1016/S1534-0856%2802%2905009-0

Galinsky, A. D., Wang, C. S., Whitson, J. A., Anicich, E. M., Hugenberg, K., & Bodenhausen, G. V. (2013). The reappropriation of stigmatizing labels: the reciprocal relationship between power and self-labeling. *Psychological Science*, 24(10), 2020–2029. https://doi.org/10.1177/0956797613482943

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science*, 24(3), 343–360. https://doi.org/10.1214/09-STS301

Gaucher, D., Hunt, B., & Sinclair, L. (2015). Can pejorative terms ever lead to positive social consequences? The case of slut walk. *Language Sciences*, 52, 121–130. https://doi.org/10.1016/J.LANGSCI.2015.03.005

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*1, 60(4), 509.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of latino women. *Personality and Social Psychology Bulletin*, 28(5), 659–670.

https://doi.org/10.1177/0146167202288010

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Gregory, R. J. (2007). Psychological testing: history, principles, and applications. (5th ed.). Allyn & Bacon.

Gupta, V. K., Turban, D. B., & Pareek, A. (2013). Differences between men and women in opportunity evaluation as a function of gender stereotypes and stereotype activation. *Entrepreneurship: Theory and Practice*, 37(4), 771–788. https://doi.org/10.1111/j.1540-6520.2012.00512.x

Hakim, A., & Quartiroli, A. (2016). Minority, student, and athlete: multiracial division I college athletes' stereotype threat experiences. https://search.proquest.com/openview/67b87bdff4891704f19c5f88c48a5b6a/1?pq-origsite=gscholar&cbl=18750&diss=y

Heidrich, C., & Chiviacowsky, S. (2015). Stereotype threat affects the learning of sport motor skills. *Psychology of Sport and Exercise*, 18, 42–46. https://doi.org/10.1016/J.PSYCHSPORT.2014.12.002

Heinen, E., Mackett, R., Van Wee, B., Ogilvie, D., & Panter, J. (2018). Residential self-selection in quasi-experimental and natural experimental studies: an extended conceptualization of the relationship between the built environment and travel. *Journal of Transport and Land Use*, 11(1), 939–959.

https://doi.org/10.5198/jtlu.2018.1165

Henry, P. J., Butler, S. E., & Brandt, M. J. (2014). The influence of target group status on the perception of the offensiveness of group-based slurs. *Journal of Experimental Social Psychology*, 53, 185–192. https://doi.org/10.1016/j.jesp.2014.03.012

Heyman, T., Hutchison, K. A., & Storms, G. (2016). Is semantic priming (ir)rational? Insights from the speeded word fragment completion task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10), 1657–1663. https://doi.org/10.1037/xlm0000260

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*, 25(4), 652–665. https://doi.org/10.1016/j.jcps.2014.12.002

Jeshion, R. (2013a). Slurs and stereotypes. *Analytic Philosophy*, 54(3), 314–329. https://doi.org/10.1111/phib.12021

Jeshion, R. (2013b). Expressivism and the offensiveness of slurs. *Philosophical Perspectives*, 27(1), 231–259. https://doi.org/10.1111/phpe.12027

John-Henderson, N. A., Rheinschmidt, M. L., Mendoza-Denton, R., & Francis, D. D. (2014). Performance and inflammation outcomes are predicted by different facets of SES under stereotype threat. *Social Psychological and Personality Science*, 5(3), 301–309. https://doi.org/10.1177/1948550613494226

Jordt, H., Eddy, S. L., Brazil, R., Lau, I., Mann, C., Brownell, S. E., King, K., & Freeman, S. (2017). Values affirmation intervention reduces achievement gap

between underrepresented minority and white students in introductory biology classes. *CBE—Life Sciences Education*, 16(3), ar41. https://doi.org/10.1187/cbe.16-12-0351

Kalokerinos, E. K., von Hippel, C., & Zacher, H. (2014). Is stereotype threat a useful construct for organizational psychology research and practice? *Industrial and Organizational Psychology*, 7(03), 381–402. https://doi.org/10.1111/iops.12167

Kapitanoff, S., & Pandey, C. (2017). Stereotype threat, anxiety, instructor gender, and underperformance in women. *Active Learning in Higher Education*, 18(3), 213–229. https://doi.org/10.1177/1469787417715202

Kateri, M. (2014). Contingency table analysis: Methods and implementation using R. *In Contingency Table Analysis: Methods and Implementation Using R*. Springer New York. https://doi.org/10.1007/978-0-8176-4811-4

Kellow, J. T., & Jones, B. D. (2008). The effects of stereotypes on the achievement gap: reexamining the academic performance of African American high school students. *Journal of Black Psychology*, 34(1), 94–120. https://doi.org/10.1177/0095798407310537

Lambert, A. E., Watson, J. M., Stefanucci, J. K., Ward, N., Bakdash, J. Z., & Strayer, D. L. (2016). Stereotype threat impairs older adult driving. *Applied Cognitive Psychology*, 30(1), 22–28. https://doi.org/10.1002/acp.3162

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(1), 180–193. https://doi.org/10.1037/a0038586

Lavrakas, P. (2013). Nondifferentiation. *In Encyclopedia of Survey Research Methods.* Sage Publications, Inc. https://doi.org/10.4135/9781412963947.n334

Leiner, D. J. (2019). Too fast, too straight, too weird: non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, 13(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403

Liu, P., Zhao, F., Zhang, B., & Dang, Q. (2017). Small change makes a big splash: the role of working self-concept in the effects of stereotype threat on memory. *Journal of Psychology: Interdisciplinary and Applied*, 151(7), 613–631. https://doi.org/10.1080/00223980.2017.1372340

Martin, D. B. (2009). Researching race in mathematics education. *Teachers College Record*, 111(2), 295–338. https://s3.amazonaws.com/academia.edu.documents/31971516/Article1.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1509482672&Signature=1RO%2FRw4WzGPG8Msbfqy6lpu3lZA%3D&response-content-disposition=inline%3Bfilename%3DResearching_Race_In_Mathematics_Educatio.pdf

Marx, D. M., & Roman, J. S. (2002). Female role models: protecting women's math test performance. *Personality and Social Psychology Bulletin*, 28(9), 1183–1193. https://doi.org/10.1177/01461672022812004

Massey, D. S., & Owens, J. (2014). Mediators of stereotype threat among black college students. *Ethnic and Racial Studies*, 37(3), 557–575.

McCambridge, J., de Bruin, M., & Witton, J. (2012). The effects of demand

characteristics on research participant behaviours in non-laboratory settings: a systematic review. *PloS One*, 7(6), e39116. https://doi.org/10.1371/journal.pone.0039116

McGlone, M. S., & Pfiester, R. A. (2015). Stereotype threat and the evaluative context of communication. *Journal of Language and Social Psychology*, 34(2), 111–137. https://doi.org/10.1177/0261927X14562609

McKinley, C. J., Mastro, D., & Warber, K. M. (2014). Social identity theory as a framework for understanding the effects of exposure to positive media images of self and other on intergroup outcomes. *International Journal of Communication*, 8, 20. http://ijoc.org/index.php/ijoc/article/view/2276

Musu-Gillette, L., de Brey, C., McFarland, J., Hussar, W., Sonnenberg, W., & Wilkinson-Flicker, S. (2017). Status and trends in the education of racial and ethnic groups 2017. https://nces.ed.gov/pubs2017/2017051.pdf

Najdowski, C. J., Bottoms, B. L., & Goff, P. A. (2015). Stereotype threat and racial differences in citizens' experiences of police encounters. *Law and Human Behavior*, 35(5), 463.

"National Center for Education Statistics." (2014). Reading and mathematics score trends. http://nces.ed.gov/programs/coe/indicator_cnj.asp

Neguţ, A., & Sârbescu, P. (2014). Problem music or problem stereotypes? The dynamics of stereotype activation in rock and hip-hop music. *Musicae Scientiae*, 18(1), 3–16. https://doi.org/10.1177/1029864913499180

O'Dea, C. J., Miller, S. S., Andres, E. B., Ray, M. H., Till, D. F., & Saucier, D. A.

(2015). Out of bounds: factors affecting the perceived offensiveness of racial slurs. *Language Sciences*, 52, 155–164. https://doi.org/10.1016/j.langsci.2014.09.005

Oliveira, E., & Cabral-Cardoso, C. (2017). Older workers' representation and age-based stereotype threats in the workplace. *Journal of Managerial Psychology*, 32(3), 254–268. https://doi.org/10.1108/JMP-03-2016-0085

Ott, R. L., & Longnecker, M. L. (2015). An introduction to statistical methods and data analysis (7th ed.). Cengage Learning.

Pacilli, M. G., Tomasetto, C., & Cadinu, M. (2016). Exposure to sexualized advertisements disrupts children's math performance by reducing working memory. *Sex Roles*, 74(9–10), 389–398. https://doi.org/10.1007/s11199-016-0581-6

Pennington, C. R., Litchfield, D., McLatchie, N., & Heim, D. (2019). Stereotype threat may not impact women's inhibitory control or mathematical performance: providing support for the null hypothesis. *European Journal of Social Psychology*, 49(4), 717–734. https://doi.org/10.1002/ejsp.2540

Picho, K., & Schmader, T. (2018). When do gender stereotypes impair math performance? A study of stereotype threat among ugandan adolescents. *Sex Roles*, 78(3–4), 295–306. https://doi.org/10.1007/s11199-017-0780-9

Popham, L. E., & Hess, T. M. (2015). Age differences in the underlying mechanisms of stereotype threat effects. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 70(2), 223–232.

https://doi.org/10.1093/geronb/gbt093

Rahman, J. (2011). The N word: its history and use in the African American community. *Journal of English Linguistics*, 40(2), 137–171. https://doi.org/10.1177/0075424211414807

Riciputi, S., & Erdal, K. (2017). The effect of stereotype threat on student-athlete math performance. *Psychology of Sport and Exercise*, 32(August), 54–57. https://doi.org/10.1016/j.psychsport.2017.06.003

Robinson, M. M. (2016). Effects of stereotype threat on college women's math performance and test anxiety. *McNair Scholars Journal*, 17, 106–114.

Saad, C. S., Meyer, O. L., Dhindsa, M., & Zane, N. (2015). Domain identification moderates the effect of positive stereotypes on Chinese American women's math performance. *Cultural Diversity and Ethnic Minority Psychology*, 21(1), 162–167. https://doi.org/10.1037/a0038428

Salter, P. S., Kelley, N. J., Molina, L. E., & Thai, L. T. (2017). Out of sight, out of mind: racial retrieval cues increase the accessibility of social justice concepts. *Memory*, 25(8), 1139–1147. https://doi.org/10.1080/09658211.2016.1274037

Schmader, T., Block, K., & Lickel, B. (2015). Social identity threat in response to stereotypic film portrayals: effects on self-conscious emotion and implicit ingroup attitudes. *Journal of Social Issues*, 71(1), 54–72. https://doi.org/10.1111/josi.12096

Schuster, C., & Martiny, S. E. (2017). Not feeling good in STEM: effects of stereotype activation and anticipated affect on women's career aspirations. *Sex Roles*, 76(1–

2), 40–55. https://doi.org/10.1007/s11199-016-0665-3

Schuster, C., Martiny, S. E., & Schmader, T. (2015). Distracted by the unthought – suppression and reappraisal of mind wandering under stereotype threat. *PLOS ONE*, 10(3), e0122207. https://doi.org/10.1371/journal.pone.0122207

Sellers, R. M., Chavous, T. M., & Cooke, D. Y. (1998). Racial ideology and racial centrality as predictors of african american college students' academic performance. *Journal of Black Psychology*, 24(1), 8–27. https://doi.org/10.1177/00957984980241002

Sellers, R. M., Rowley, S. A. J., Chavous, T. M., Shelton, J. N., & Smith, M. A. (1997a). Multidimensional inventory of black identity: A preliminary investigation of reliability and construct validity. *Journal of Personality and Social Psychology*, 73(4), 805–815. https://doi.org/10.1037/0022-3514.73.4.805

Sellers, R. M., Rowley, S. A. J., Chavous, T. M., Shelton, J. N., & Smith, M. A. (1997b). Multidimensional inventory of black identity. PsycTESTS. https://doi.org/doi:Full; Full text; 999903182pass:[_]full_001.pdf

Sellers, R. M., Smith, M. A., Shelton, J. N., Rowley, S. A. J., & Chavous, T. M. (1998). Multidimensional model of racial identity: a reconceptualization of African American racial identity. *Personality and Social Psychology Review*, 2(1), 18–39. https://doi.org/10.1207/s15327957pspr0201_2

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Cengage Learning.

Shelvin, K. H., Rivadeneyra, R., & Zimmerman, C. (2014). Stereotype threat in African American children: the role of Black identity and stereotype awareness. *International Review of Social Psychology*, 27(3–4), 175–204.

Silverman, A. M., & Cohen, G. L. (2014). Stereotypes as stumbling-blocks. *Personality and Social Psychology Bulletin*, 40(10), 1330–1340. https://doi.org/10.1177/0146167214542800

Simmons, C., Worrell, F. C., & Berry, J. M. (2008). Psychometric properties of scores on three black racial identity scales. *Assessment*, 15(3), 259–276. https://doi.org/10.1177/1073191108314788

Soley-Bori, M., Horn, M., Morgan, J., & Min Lee, K. (2013). Dealing with missing data: key assumptions and methods for applied analysis. https://www.bu.edu/sph/files/2014/05/Marina-tech-report.pdf

Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, 44(2), 136–146. https://doi.org/10.1002/ab.21737

Spotorno, N., & Bianchi, C. (2015). A plea for an experimental approach on slurs. *Language Sciences*, 52, 241–250. https://doi.org/10.1016/j.langsci.2015.04.004

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Stepanova, E. V., Bartholow, B. D., Scott Saults, J., Friedman, R. S., Chi, J., & Hollis, A. M. (2018). Effects of exposure to alcohol-related cues on racial prejudice. *Social*

Psychology, 49(2), 113–124. https://doi.org/10.1027/1864-9335/a000334

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

Stone, J., Perry, W., & Darley, J. (1997). *White men can't jump*: evidence for the perceptual confirmation of racial stereotypes following a basketball game. *Basic and Applied Social Psychology*, 19(3), 291–306. https://doi.org/10.1207/s15324834basp1903_2

Stout, J. G., Dasgupta, N., Hunsinger, M., & Mcmanus, M. A. (2010). STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). https://doi.org/10.1037/a0021385

Thames, A. D., Hinkin, C. H., Byrd, D. A., Bilder, R. M., Duff, K. J., Mindt, M. R., Arentoft, A., & Streiff, V. (2013). Effects of stereotype threat, perceived discrimination, and examiner race on neuropsychological performance: simple as black and white? *Journal of the International Neuropsychological Society:* JINS, 19(5), 583–593. https://doi.org/10.1017/S1355617713000076

Toothaker, L. E. (1993). Multiple comparison procedures. Sage Publications, Inc.

Tyler, K. M., Thompson, F. A., Gay, D. E., Burris, J., Lloyd, H., & Fisher, S. (2016). Internalized stereotypes and academic self-handicapping among Black American male high school students. *Negro Educational Review*, 67(1–4), 5.

U.S. Census Bureau. (2015). Table DP05: American Community Survey Demographic and Housing Estimates.

https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_SPT_DP05&prodType=table

U.S. Census Bureau. (2019). Sex by educational attainment for the population 25 years and over (Black or African American alone). https://data.census.gov/cedsci/table?q=Black or African American Education&tid=ACSDT1Y2019.B15002B&hidePreview=false

Vandiver, B. J., Worrell, F. C., & Delgado-Romero, E. A. (2009). A Psychometric examination of multidimensional inventory of black identity (MIBI) scores. *Assessment*, 16(4), 337–351. https://doi.org/10.1177/1073191109341958

Wang, P., Yang, Y.-P., Tan, C.-H., Zhao, X.-X., Liu, Y.-H., & Lin, C.-D. (2016). Stereotype activation is unintentional: behavioural and event-related potential evidence. *International Journal of Psychology*, 51(2), 156–162. https://doi.org/10.1002/ijop.12135

Wang, P., Zhou, P., Tan, C.-H., & Zhang, P.-C. (2017). Effect of self-efficacy in stereotype activation. *Social Behavior and Personality: An International Journal*, 45(3), 469–476. https://doi.org/10.2224/sbp.5201

Wasserberg, M. J. (2014). Stereotype threat effects on African American children in an urban elementary school. *The Journal of Experimental Education*, 82(4), 502–517. https://doi.org/10.1080/00220973.2013.876224

Wright-Adams, V. (2014). Stereotype threat and the academic task performance of African American women. Available from Dissertations & Theses @ Walden University; ProQuest Dissertations & Theses Global.

Appendix A: R-Programming Code for Word Fragment List Development

```
####read in word database from http://elexicon.wustl.edu/WordStart.asp####
df <- read.csv(file = "projects/word_fragments/EngLexiconDB2019April.csv",
stringsAsFactors = F)
df[,5:ncol(df)] <- as.data.frame(
 lapply(X = df[,5:ncol(df)],
 FUN = as.numeric),
 stringsAsFactors = F
)
summary(df)

####targeted words####
###attributes and summary statistics
targeted_words <-
c("race","lazy","black","poor","class","brother","white","minority","welfare","color","t
oken")
targeted_words_attr <- df[df$Word %in% targeted_words,] #query target words from
ELP data
summary(targeted_words_attr)

###filter ELP and match target word attributes
##remove targeted words from df
df_filler_selection <- droplevels(df[!(df$Word %in% targeted_words),])
##remove words with special charaters
df_filler_selection <- droplevels(df_filler_selection[!grepl(pattern = "[']", x =
df_filler_selection$Word),])
##remove words with capital letters
df_filler_selection <- droplevels(df_filler_selection[!grepl(pattern = "^[A-Z]", x =
df_filler_selection$Word),])
##length match
df_filler_selection <- droplevels(df_filler_selection[df_filler_selection$Length >=
min(targeted_words_attr$Length) &
 df_filler_selection$Length <= max(targeted_words_attr$Length),])
##freqency match
df_filler_selection <- droplevels(df_filler_selection[df_filler_selection$Freq_HAL >=
min(targeted_words_attr$Freq_HAL) &
 df_filler_selection$Freq_HAL <= max(targeted_words_attr$Freq_HAL),])
##mean naming latency (raw) match
df_filler_selection <-
droplevels(df_filler_selection[df_filler_selection$I_NMG_Mean_RT >=
min(targeted_words_attr$I_NMG_Mean_RT) &
 df_filler_selection$I_NMG_Mean_RT <=
max(targeted_words_attr$I_NMG_Mean_RT),])
```

```
##remove NAs
df_filler_selection <- df_filler_selection[,-which(colnames(df_filler_selection) ==
"I_NMG_Obs")] #remove entire column due to NAs
df_filler_selection <- df_filler_selection[complete.cases(df_filler_selection),] #select
rows with no NAs

##look
summary(df_filler_selection)

###randomly select filter words from matched word bank
n <- nrow(targeted_words_attr) * 3
set.seed(47)
df_filler_final <- df_filler_selection[sample(x = nrow(df_filler_selection), size = n),]
summary(df_filler_final)

####select missing letters from filler words####
###target word database of blanks and position
targeted_words <-
c("race","lazy","black","poor","class","brother","white","minority","welfare","color","t
oken")
targeted_words_length <- nchar(targeted_words)
targeted_words_blank_n <- c(2,2,2,2,2,5,3,6,4,3,3)
targeted_words_blank_position <- factor(x = c("begining",
 "end",
 "begining",
 "begining",
 "gap",
 "end",
 "begining",
 "end",
 "end",
 "end",
 "end"))
targeted_words_df <- data.frame(targeted_words,
 targeted_words_length,
 targeted_words_blank_n,
 targeted_words_blank_position)
###
tab1 <- round(prop.table(ftable(targeted_words_df$targeted_words_blank_n,
targeted_words_df$targeted_words_blank_position,
targeted_words_df$targeted_words_length), margin = 2)*100, digits = 1)
tab2 <- as.data.frame(tab1); colnames(tab2) <-
c("n_blanks","pos_blank","word_len","freq")
```

```
tab3 <- as.data.frame(table(nchar(df_filler_final$Word))); colnames(tab3) <-
c(“word_len”,”word_count”)
tab4 <- merge(tab2,tab3)
#calculate number of word to match to missing/blank letter attributes
tab4$apply_to_n_filler_words <- round((tab4$freq/100)*tab4$word_count)
```

Appendix B: Word Fragment Task Word List

Target Word Fragments

1. _ _ C E (RACE)
2. L A _ _ (LAZY)
3. _ _ A C K (BLACK)
4. _ _ O R (POOR)
5. C L _ S _ (CLASS)
6. B R _ _ _ _ _ (BROTHER)
7. _ _ _ T E (WHITE)
8. M I _ _ _ _ _ _ (MINORITY)
9. W E L _ _ _ _ (WELFARE)
10. C O _ _ _ (COLOR)
11. T O _ _ _ (TOKEN)

Filler Word Fragments

1. W E S _ _ _ _(WESTERN)
2. _ _ Y A L (ROYAL)
3. _ _ _ E D (NAMED)
4. H E L _ _ _ _ (HEALTHY)
5. _ _ N U (MENU)
6. _ _ R N (BURN)
7. C L E _ _ _ _ (CLEARLY)
8. _ _ N E (LONE)
9. _ _ A L (GOAL)
10. B E _ M _ (BECOME)
11. C O _ _ _ _ _ _ (COMBINED)
12. L E _ _ (LEGS)
13. P U _ _ _ _ _ (PURPOSE)
14. A P _ _ _ _ _ (APPLIED)
15. M A _ _ _ (MAKING)
16. R E _ L _ (REPLY)
17. _ _ N G (KING)

Appendix C: Permission to Use the SAT

Charles,
Thank you for providing more detail regarding your study.

Your request is **APPROVED**.

Permission is granted on a one-time, non-exclusive and non-transferable basis, provided you agree to the following terms and conditions:

1. When administering the questions, you shall state that the test administration is for research purposes only and not as an agent of the College Board or SAT® exam program.
2. You shall distribute the test questions directly to only test takers participating in your study.
3. When using the questions from the practice test:
    a. You shall limit the distribution to one handout per test taker and the handout must be distributed as a stand-alone document and not incorporated into your own publication.
    b. The College Board logo, SAT® trademark and copyright information remains intact within the printouts.
    c. You shall print or copy the pages exactly as they appear and not alter the content of the items.
    d. Display a proper source citation in the handout.
4. If at some point you publish your work and need a copyright source citation for the test questions, you must include the proper citation, as example of which is below:
   Source: Official SAT® Practice Test, © 2016. The College Board. www.collegeboard.org. Used with permission.

Please let me know if you have any questions or need further information.

Thanks,
The College Board

Appendix D: Permission to Use MIBI

Dear Charles,

Dr. Sellers asked me to respond to your email and apologize for the delay. You have his permission to use the MIBI.

Sincerely,

Office of the Vice Provost for Equity and Inclusion and Chief Diversity Officer

Appendix E: Demographic Questionnaire

1. Please indicate your age:
2. Please indicate your gender:
    a. Male
    b. Female
3. Please indicate your race/ethnicity:
    a. American Indian or Alaskan Native
    b. Asian
    c. Black or African American
    d. Hispanic
    e. Native Hawaiian or Pacific Islander
    f. Two or More Races
    g. White
4. Please indicate your education level:
    a. High School Graduate or Equivalent
    b. Technical/Vocational Trade Program or Certified
    c. Associates Degree
    d. Bachelor’s degree
    e. Master’s Degree
    f. Doctoral Degree