A Meta-Analysis of Experimental Examinations of Racism in U.S. Pre-K-12th Grade Educational Contexts

by

Jayley Janssen

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2023 by the
Graduate Supervisory Committee:

Elanor Seaton, Chair
Cindy Faith Miller
Jose Causadias
Justin Jager

ARIZONA STATE UNIVERSITY

May 2023

ABSTRACT

Racism primarily has been examined in United States (U.S.) educational contexts via survey and qualitative investigations, which reveal Black, Indigenous, and People of Colors' (BIPOCs') self-reported experiences of individual racism. Missing from this research is a focus on the perpetrators of racism and the institutional nature of racism in U.S. education. Experimental examinations of racism in U.S. education are efficacious in identifying perpetrators of racism and offer objective evidence of racism, which can underscore the importance of qualitative and survey findings. However, experimental examinations of racism in educational contexts are infrequent, examine a wide array of perpetrators (e.g., peers, teachers, principals) and educational domains (grading, diagnoses, liking), and reveal mixed findings. Thus, I utilized meta-analytic techniques to explore the magnitude of the effect of racism in pre-K-12th grade contexts as demonstrated by experimental examinations. In addition, I explore moderation of the effect of racism by 1) the type of experimental technique, 2) the specific perpetrator, 3) the BIPOC target, 4) the region of the US, and 5) the design of treatment assignment. Using 71 effect sizes from 57 studies, I found evidence that racism is present in U.S. pre-K-12th grade education for BIPOC [Cohen's d=0.15, 95% CI(0.05, 0.25)]. Heterogeneity existed in the overall effect, and moderators included the U.S. region, and the level of racism (e.g., individual versus institutional) explained the heterogeneity of the effects. The findings are discussed in terms of theoretical and methodological implications and future directions for research.

DEDICATION

This dissertation is dedicated to strong women; To women who blaze the trail that moves us forward; To Christina Janssen, Pamela Leffringhouse, Eleanor Seaton, and Eva Marie Shivers. In unique ways, each of you exemplifies the strength and power of trailblazing women. You made this dissertation possible and are the reason I get to engage in this meaningful work. This dissertation is for you.

ACKNOWLEDGMENTS

Thank you, first and foremost, to my committee members, Dr. Eleanor Seaton, Dr. Cindy F. Miller, Dr. Jose Causadias, and Dr. Justin Jager. I feel so lucky to have been guided by you throughout my doctoral milestones. Together, you were a dream committee that emboldened my pursuit of this research niche. Eleanor, my advisor and mentor, thank you for sharing your expertise on racism, for offering me so many opportunities to share my passion for experimental research, and for your generosity and patience guiding me as a White woman to critically examine racism's impact across all levels of society. Cindy, thank you for your sharing your exceptional educational perspective that so wonderfully combines research and practice. You consistently expanded my understanding of the implications of this research. Jose and Justin, thank you for sharing your methodological and analytic brilliance. You helped me reach levels of rigor that I could not have achieved on my own.

To all my colleagues, thank you for holding my hand and walking with me along the journey toward this degree. Specifically, thank you to Dr. Brittany Alexander, Clarissa Abidog, Dr. Diana Jenkins, Aubery Hoeffer, and Kamryn Morris. Sharing in the struggles and triumphs of this journey made this degree not only possible but bearable, and sometimes even fun!

To Dr. Katie Zeiders and Dr. Antionette Landor. Thank you for opening this world to me. You were the first to teach me about research and the first to believe that I was capable of obtaining a Ph.D.

To my husband and partner in everything, Alex. You are my biggest supporter, and you have sacrificed so much for me to complete this degree. Thank you for your unconditional love, for keeping me fed and caffeinated, for sitting with me while I worked, and for always believing I could do it. I love you and our little parenthesis!

To my family. Thank you for your patience, encouragement, and celebration as I navigated each milestone of this degree. My parents: Christina Janssen, Greg Janssen, and Dwight Heubner; My grandparents: Pamela Leffringhouse, Thomas Clark, and Ruth Ann and Gene Janssen; My siblings: Jacqueline and Chase Hildebrand, Janelle Janssen; and my second family: Melinda, Jim, and Nick Rinella.

To my niece and my nephew, Adeline and Oliver Hildebrand. Thank you for your Facetimes filled with smiles and laughter. Your calls were bright spots as I completed this dissertation. You mean the entire world to me.

To Jerry. You sat with me while I wrote, were by my side while I coded, and brought me so much comfort as I completed this degree. I am so grateful to you. Finally, to all my friends and anyone who ever spoke a word of encouragement as I completed this degree. Thank you from the bottom of my heart.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

A Meta-Analysis of Experimental Examinations of Racism in U.S. Educational Institutions

Research on the United States (U.S.) educational institutions has largely avoided examining the perpetrators and institutional nature of racism, a system of dominance in which dominant members with power and privilege create and sustain systems that exclude Black, Indigenous, and People of Color (BIPOC) from power, esteem, and status (Harrell, 2000). Instead, BIPOC accounts of individual racism (Estrada & Hondagneu-Sotelo, 2011; Hardie & Tyson, 2013; Seaton & Iida, 2019; Wang & Yip, 2020; Zeiders, 2017) and its correlates (e.g., physical and mental health outcomes, decreased academic trajectories, resiliency strategies; Benner & Graham, 2013; Chavous et al., 2008; Crengle et al., 2012; Priest et al., 2013; Seaton & Yip, 2009) have been the foci. As a result, efforts to reduce racism in educational institutions have targeted the classroom context or peer and teacher interactions with BIPOC. However, historically little attention has been given to the need for educational reform: the identification and modification of persons (e.g., peers, teachers, administration), practices, and policies complicit with White Supremacy. White supremacy is defined as the attitudes, ideologies, and politics associated with the structuring of racial order where "White" and "Whiteness" constitute a superior and dominating race at the detriment of other racial and ethnic groups (Bonilla-Silva, 2001; Fredrickson, 1981).

Evidence documenting racism in educational institutions predominantly emerges from qualitative and survey data. Though efficacious in providing rich and in-depth data, qualitative and survey methods are not without their limitations. By nature, both qualitative and survey findings are subjective and correlational. This affords legislators

and leaders in the U.S. education system to discount evidence of racism and allows them to continue to prescribe to color evasive and post-racial ideologies (Annamma et al., 2017; Bonilla-Silva, 2015).

Experimental probes of racism in educational institutions have the potential to bridge gaps that currently exist in survey and qualitative research with successful triangulation of scientific understanding of racism in educational contexts. Experimental probes involve manipulating variables, often referred to as treatment and controls, to establish cause and effect (Reichardt, 2019). Results from experimental investigations of racism offer objective and indisputable evidence of the existence of racism and denote the significance of the aforementioned survey and qualitative findings demonstrating BIPOC's discriminatory experiences (Giulietti et al., 2019; Milkman et al., 2012, 2015; Stewart & Uggen, 2020). Further, given their ability to determine cause and effect, experimental probes are particularly useful in identifying the perpetrators of educational racism and racist policies and practices, which, when viewed in aggregate, can demonstrate the institutional nature of educational racism.

Despite the considerable strengths of experimental methods, the adoption of experimental techniques to examine racism in educational settings is infrequent. As a result, information about the implementation of these methods, specifically in educational contexts, is sparse. In addition, the findings from research utilizing these techniques are mixed. Experimental examinations of educational contexts have identified the perpetration of racism by peers, teachers, administration, and policies. Such objective evidence of racism from multiple persons and levels of the education system contributes to the understanding of the institution of racism in the education system – the

combination of racist behaviors and policies that work together to inhibit BIPOC's academic trajectories (Teeger, 2015). However, experimental examinations of educational contexts have also reported null findings and, in some cases, a pro-BIPOC bias in which BIPOC targets were given preferential treatment.

To understand the overall effect of racism in educational contexts, the results from all experimental examinations need to be quantitatively synthesized. In addition, a systematic aggregation of the evidence of racism from all persons and levels of the U.S. education system is required to further scientific understanding of the institutional nature of educational racism for researchers, practitioners, and policymakers. Such information is essential for the current climate of the U.S. education system. As racism in the U.S. gains national and global scrutiny from White individuals, increased criticism and attention have been given to the U.S. education system for contributing to the oppression of BIPOC (Chavez, 2021; Laughland, 2022). Further, there have been public outcries for reform in U.S. education policies and procedures to be promotive for BIPOC and contribute to anti-racist development (Chatterji, 2020; Meckler & Natanson, 2021). Though racism traces to the very origins of the U.S. education system (Takaki, 2008; Noltemeyer et al., 2012), the increased intention for reform sparks the need for scientific understanding backed by strong theoretical and methodological evidence. Synthesized experimental evidence of both the perpetrators of interpersonal racism and institutional racism in the U.S. education system can be particularly useful to guide educational reformations to focus on dismantling racist persons, policies, and procedures complicit with White supremacy.

The present study will utilize meta-analytic techniques to systematically and quantitatively summarize the results of experimental examinations of racism across pre-K through 12th-grade educational contexts. Two research questions guide the present meta-analysis: 1) what is the magnitude of the effect of racism in pre-K through 12th-grade educational contexts as demonstrated by experimental examinations? and 2) is the overall effect size moderated by the individual (e.g., student peers, teachers, principals) or study characteristics (e.g., type of experimental technique or region)?

Interpersonal, Institutional, and Structural Racism

Racism is multilevel and multidimensional (Jones, 1997). Interpersonal or individual racism is the behavioral component of racism (referred to as interpersonal racism throughout; Jones, 1997; Williams et al., 2003). Interpersonal racism is a personal manifestation of racism exhibited by White dominant group members' interactions with BIPOC (Jones, 1997; Lawrence & Keleher, 2004; Teeger, 2015). White people's enactment of interpersonal racism is motivated by their personal beliefs and biases about race and societal ideologies that are socialized from birth (Jones, 1997; Teeger, 2015). Importantly, the theoretical grounding that racism is rooted in power and dominance has implications for those *who* can exhibit individual racism. Through this theoretical lens, only White individuals in the U.S. can exhibit individual racism. Though BIPOC can prescribe to racial stereotypes (e.g., overgeneralized labels), prejudice (e.g., judgmental attitudes), and exhibit discrimination (e.g., unfair treatment), these phenomena are distinct from racism (Harrell, 2000; Jones, 1997; Lott & Maluso, 1995). The ability to enact interpersonal racism today is directly tied to the long history of White racial dominance and power in the U.S. that formed the current societal circumstances (Harrell,

2000). Thus, the ability to enact interpersonal racism hinges on the ability to benefit from the history of racial dominance and power, a benefit only afforded to White individuals. Institutional racism comprises inequitable policies, practices, and opportunities that are produced and perpetuated by institutions (e.g., the education system) to oppress BIPOC (Jones, 1997; Lawrence & Keleher, 2004; Teeger, 2015). All individuals within an institution can enact the institution's power when they enforce racist policies and procedures outlined by the institution. Finally, structural racism, or the system of racial hierarchy and inequity characterized by White Supremacy, is strengthened and perpetuated by the interaction and interconnections of racist societal institutions (Gee & Hicken, 2021; Lawrence & Keleher, 2004; Teeger, 2015). Structural racism perpetuates, normalizes, and legitimizes historical, institutional, and interpersonal racism and produces compounding oppressions for BIPOC (Gee & Hicken, 2021; Lawrence & Keleher, 2004; Teeger, 2015). Lawrence and Keleher (2004) explain:

> *"Structural Racism lies underneath, all around, and across society. It encompasses (1) history, which lies underneath the surface, providing the foundation for White supremacy in this country, (2) culture, which exists all around our everyday lives, providing the normalization and replication of racism and (3) interconnected institutions and policies, they key relationships and rules across society providing the legitimacy and reinforcements to maintain and perpetuate racism" (p. 2).*

This review probes both interpersonal and institutional racism. Interpersonal racism is probed via the aggregation of experimental evidence exhibited in perpetrator-target relationships that are not situated in a position of power (e.g., peer-peer

relationships, parent-school relationships, teacher-teacher relationships). Institutional racism is probed via the aggregation of experimental evidence of racist policies, practices, and procedures in the education system (e.g., hiring, curriculum, educational policy) and via the aggregation of experimental evidence of perpetrators in positions of power exhibiting racism (e.g., teacher-student, counselor-student, principal-student). Further, this review investigates connections between these two types of racism via the overall estimated effect of racism in pre-K to 12th grade. An estimated effect provides a nuanced understanding of the ways in which these two types of racism reinforce each other to limit BIPOC's academic trajectories.

Experimental Examinations of Racial Discrimination

Experimental examinations have long been established as the gold standard, given their ability to determine cause and effect (Imai et al., 2008; Reichardt, 2019). By deliberately imposing a treatment, the effect can be measured (Imai et al., 2008; Reichardt, 2019). Examining racism presents a unique problem for traditional experimental methods, given that race cannot be directly manipulated or randomly assigned. As a result, researchers have experimentally varied the "apparent" race or the level of "apparent" racist treatment of a fictitious target (Blank et al., 2004). For this meta-analysis, the definition of an experimental examination encompasses the following: 1) true experiments, in which all factors affecting the phenomena of interest are controlled, and manipulations are introduced in laboratory settings; 2) quasi-experiments, in which factors affecting the phenomena are controlled, but treatments are introduced in a natural or real-world setting, and 3) natural experiments in which factors affecting the phenomena are not controlled, and the treatment is naturally occurring, or unplanned (for

examples see Dee, 2004; Giulietti et al., 2019; Glock et al., 2015; Lalani & Bhutta, 2019; Milkman et al., 2012, 2015; Richeson & Nussbaum, 2004; Thornhill, 2019).

Two types of imposed treatments will be examined for studies included in this meta-analysis review. First, studies were included in which the apparent race of a target was manipulated by providing race-related cues (e.g., racially salient names, images) to determine the effect of the apparent race on the responder (e.g., if participants' direct response is different for one race treatment when compared to another; Blank et al., 2004). Such examinations measure the behavior of potential perpetrators toward targets of different races (Blank et al., 2004). Second, studies were included in which the target's experience of racism was manipulated to determine how persons with power and privilege in the education system respond to racism (i.e., their vicarious response; Blank et al., 2004). Together, these domains shed light on how racism pervades and persists in U.S. educational contexts.

Experimental Techniques

Within the three categories of experimental investigations (e.g., true experiments, quasi-experiments, natural experiments), five experimental techniques have been used to investigate racism in U.S. educational contexts: 1) audit experiments, 2) vignette experiments, 3) experimental tasks, 4) confederate experiments, and 5) natural experiments (Janssen et al., 2022; Morris et al., 2022). Each experimental technique is described, and findings of racism across perpetrators are discussed. In addition, we note when samples included both White and BIPOC participants. The inclusion of BIPOC participants when examining individual racism defies our theoretical understanding that individual racism is rooted in establishing and maintaining racial power and privilege

(Harrell, 2000). However, BIPOC may perpetuate racism by enacting the racist power of the educational institution when holding positions of power (Harrell, 2000; Jones, 1997). Thus, findings are interpreted with caution.

Audit Experiments

Audit experiments originated from literature examining the labor market (Pager & Shepherd, 2008). Researchers selected, trained, and matched individuals to ensure equally qualified testers that differed on specific domains (e.g., race, gender, and qualifications; Pager & Shepherd, 2008). These equally qualified testers were sent out into the field to observe the effect of such characteristics on a specified outcome (Gaddis, 2018). However, as technology has evolved, researchers have translated the experimental method to be conducted online. Instead of using in-person testers, fictitious testers or profiles are created and distributed online via emails or online platforms (e.g., online application portals).

These technological advances in electronic correspondence facilitated the use of audit experiments to examine educational contexts. Researchers were able to construct fictitious, matched email or application profiles that varied by race-ethnicity to determine if the recipient would exhibit racism in how or if they responded to inquiries (Gaddis, 2018). Despite the transition to an electronic format, the hallmark of the audit experiment remains - participants are unaware of their participation in the experiment, reducing social desirability bias and allowing participants' true racialized attitudes and behaviors to be observed (Gaddis, 2018; Janssen et al., 2022).

Though the electronic nature of audit correspondence experiments facilitates a relatively cost-free experimental design, eliciting participants' racialized attitudes and

behaviors electronically results in the reliance on racially salient names to convey the race of the target. As a result, the construction and selection of names represent a pivotal step. To construct names, researchers have utilized resources on common and popular names by race-ethnicity: The 2000 U.S. decennial census, New York City Health Department's Bureau of Vital Statistics birth records, and popular websites suggesting baby names for different racial groups (e.g., Babycenter.com), Social Security Administration Data, and respective state departments of public health data. In addition, researchers have conducted preliminary experiments to test and validate the racial salience of names (Milkman et al., 2012, 2015; Thornhill, 2019). Audit correspondence examinations have been used to probe racism in educator relationships with students and to probe racism in hiring decisions within the education system. Notably, all audit experimental examinations probing "racism" in U.S. education utilize samples that include both White and BIPOC educators. However, it can be argued that the audit methodology seeks to uncover institutional racism, which can be enacted by any person within the institution who holds a position of power.

Racism in educator-student relationships. By constructing email correspondence from fictitious students, researchers can probe racism in educator responses. Audit correspondence examinations have been used to probe how White and BIPOC collegiate faculty responsiveness regarding grading, doctoral mentorship, and even admission decisions vary by the race of the student. Findings indicate harsher and more negative outcomes for BIPOC versus White targets (Milkman et al., 2012, 2015; Robinson et al., 2019; Thornhill, 2019). However, only two audit correspondence examinations have probed pre-K-12th grade educators (Guilietti et al., 2017; Janssen et

al., 2022). Janssen and colleagues (2022) sent emails asking White and BIPOC school counselors for enrollment information from mothers of prospective students. The mothers of prospective students were presented as Black, Asian, Latina, and White who had recently moved into the district. The results revealed that Asian email inquiries were answered less frequently by counselors when compared to all other racialized email inquiries in schools with higher prestige, located in rural areas, and served students with lower socioeconomic statuses (Janssen et al., 2022).

Giulietti and colleagues (2017) emailed public service providers, including school district offices. The sample of school district administrators included White and BIPOC administrators. Emails were sent from fictitious Black and White senders and asked a simple or complex inquiry. The findings revealed that White inquiries received significantly more responses than their Black counterparts from school district offices (Giulietti et al., 2017). Further, Black inquiries nearly doubled in rural areas compared to urban areas, leading to the conclusion that racial disparities in responsiveness were most prevalent in the Midwest (Giulietti et al., 2017).

Racism in hiring. In contrast to the audit examinations utilizing email correspondence, Boyd-Swan and Herbst (2019) sent electronic job applications from fictitious teachers to a sample of White and BIPOC hiring administrators at child care centers. Authors utilized software created by Lahey and Beasley (2009) to randomly create resumes that varied not only by the race of the applicant, signaled by racially salient names, but also by the qualifications (e.g., work history as none, six months, or two years of experience; high school diploma, associate's, or bachelor's degree, etc.). The results from the resume experiment demonstrated racial disparities in hiring as Black

and Latine applicants received significantly fewer responses and interview requests than White applicants (Lahey & Beasley, 2009).

Vignette Experiments

Compared to other experimental techniques, vignette experiments are the most frequently utilized in research on racism in U.S. education. Vignettes include fictitious scenarios that vary by the race of the target (e.g., Black versus White student) and/or vary by the treatment or circumstance of the target (e.g., a Latina/o/e versus White student who unfairly versus fairly received a failing grade). The constructed vignettes are presented to participants, and their reactions or self-reported hypothetical responses are assessed to estimate the effect of treatment (reference?).

Experimental vignettes can be text-based or video-based scenarios. Text-based experimental vignettes require participants to read a constructed story and then respond to a series of prompts or questions (Morris et al., 2022). Similar to audit experiments, text-based vignettes rely on racially salient names to convey the race of the target. However, in some cases, researchers explicitly conveyed the student's race in the text to ensure the salience of the manipulation (e.g., Jamal is a Black fourth-grade student). Further, in some cases, researchers utilized photos to bolster the realism of the constructed scenarios (e.g., images that varied by ethnicity-race of the subject, images of the school, classroom, or educational setting; Ash & Crammer, 2020; Biefeld, 2020; Jarvis & Okonofua, 2020). Alternatively, video-based experimental vignettes record scenarios between actors to simulate racially charged situations (Halberstadt et al., 2018; Mead, 2010; Neal et al., 2003; Spinrad, 2022). Participants' facial reactions can be assessed while watching the video, and their responses to a series of prompts or questions can be collected.

In experimental vignettes, participants are aware of their participation and the fictitious nature of the scenario. Thus, the constructed scenarios must be both realistic and designed to place the participant as an actor in the constructed scenario to successfully reveal the participant's perceptions, values, and behaviors. Researchers have validated their hypothetical scenarios via surveys and focus groups prior to conducting the experiment (Ash & Crammer, 2020; Brown, 2006; Dicker et al., 2012; Ford, 1997; Harris, 2014; Nichols & Reason, 2009; Tittler & Wade, 2019; Wolsko et al., 2000; Zhu & Bresnahan, 2018).

Similar to audit experiments, to construct racially salient names, researchers have utilized resources on common and popular names by race-ethnicity: previous vignette studies, birth certificate data, and U.S. Census data (Bertrand & Mullainathan, 2004; Fryer & Levitt, 2004; Greenwald, McGhee, & Schwartz, 1998; Killen et al., 2002; Newton, Dickter, & Gyurovski, 2011; Alonso- Marsden 2017; Levitt & Dubner, 2005; McGhee, & Schwartz, 1998; Milkman, Akinola, & Chugh, 2012). In addition, researchers have conducted preliminary experiments to test and validate the racial salience of names (Dicker et al., 2012; Fish et al., 2015; Ford, 1997; Inman et al., 1998; Kunesh & Noltemeyer, 2019; McGinnis, 2017; Noltemeyer et al., 2012; Steinberg, 1980; Tittler & Wade, 2019; Westmoreland et al., 2018; Zhu & Bresnahan, 2018). However, given that participants are aware of their participation, researchers conducting vignette experiments can also implement manipulation checks to ensure the experimental stimuli elicited the expected ethnic-racial group. Vignette experiments have been used to probe racism in education by probing educator-student relationships, peer-student relationships, and racism in hiring and enrollment decisions.

Racism in educator-student relationships. Scenarios constructed to examine teacher, principal, and counselor racism typically present student targets' misbehaviors at school and vary the ethnicity-race of the student. Notably, all vignette experimental examinations probing racism in teachers, principals, and counselors' disciplinary decisions utilize samples that include both White and BIPOC educators. The evidence suggests that teachers (Fish et al., 2015; Halberstadt et al., 2018; Kunesh & Noltemeyer, 2019; Neal et al., 2003; Okonofua et al., 2015; Xie & Cook, 2015) and principals (Jarvis & Okonofua, 2020) exhibit harsher discipline reactions toward Black student targets compared to White student targets. However, examinations have also revealed that teachers assign equal discipline reactions for Black student targets compared to White students and for Asian student targets compared to White students (Barry et al., 2020; Chang & Sue, 2003; Elhoweris et al., 2005; Noltemeyer et al., 2012; Briscoe-Jin, 2020). Further, when examining guidance counselor disciplinary decisions, vignette experiments found equal discipline reactions for Black student targets compared to White students (Dameron et al., 2019; Starcke & Porter, 2019; Hoffman & Cancelli, 2019). In one examination, researchers found a reverse effect, such that guidance counselors indicated less severe discipline for scenarios depicting Black student targets compared to White student targets (Buckley, 2020).

Vignette experiments have also probed racism in teachers' and counselors' judgment of students' cognitive and behavioral abilities. Evidence suggests that among samples of White and BIPOC teachers, racial disparities were exhibited in disability referrals for Black student targets compared to White student targets (Harris, 2013; Mead, 2011). However, examinations have also revealed that samples of White and BIPOC

teachers equally refer White and Black student targets for disability referrals (Martin, 2014; Calhoun, 1975). Further, examinations of White and BIPOC counselors' diagnoses of students' cognitive disabilities also revealed null results for the comparison of Black versus White targets (Sadeh & Sullivan, 2016; Sullivan et al., 2019). Only one vignette experimental probe utilized an all-White sample (Levy, 2004). The results revealed no evidence of individual racism in White counselors' judgment of students' cognitive disabilities (Levy, 2004).

Finally, vignette experiments have been used to examine racial disparities in educators' judgment of students' academic abilities. Notably, all of these vignette experiments included both White and BIPOC educators in the sample. Teachers and counselors often assess students' abilities and performance when recommending students to be placed in gifted programs, to be held back/advance a grade, and for diagnostic assessments. For scenarios presenting fictitious student academic profiles (e.g., grades, test scores, enrollment in gifted programming) that vary by student race-ethnicity, the evidence suggests that teachers indeed exhibit harsher treatment toward Black student targets compared to White student targets (Comeaux, 2013; Fisher, 2019; Parks & Kennedy, 2007; Quinn, 2020a). White students were treated as more academically competent and skilled even when their academic profile matched their Black counterparts. However, the evidence is not as clear for counselors. Crosby and Monin (2006) found evidence of counselors' harsher treatment toward Black academic profiles compared to White student profiles (Crosby & Monin, 2006). Yet, Saunders and Merlin-Knoblich (2021) found no effect when comparing counselor treatment to Black, White, and Latine students' profiles.

Racism in peer relationships. Vignette scenarios have been used to examine racism in peer relationships. Participants read scenarios that depicted the personality and activity characteristics and the race-ethnicity of a peer target. After reading the profiles, the participants were asked to respond with their own thoughts and beliefs about the peer target. Rhee-Worobec's (2000) examination revealed that White youth reported more empathy and helping toward the White target versus Black peer target. Results examining samples of White and BIPOC participants have also revealed that youth assigned fewer positive ratings toward Black and Chinese peer targets and harsher behavior attributions toward Black peer targets when compared to White peer targets (Biefeld, 2000; Moskowitz & Druckman, 2016; Steinberg, 1980). Further, in a sample of only White youth, Killen and Stangor (2001) probed peer relationships by encouraging the participants to imagine witnessing a scenario in which individual racism occurred. The examination revealed that White youth expressed disapproval of vignettes describing exclusions of a Black and a White peer based on racial stereotypes. Interestingly, examinations that included both White and BIPOC youth in the sample revealed similar results (Wegamnn, 2017).

Racism in hiring and enrollment decisions. Vignette experiments have also been used to probe racism in hiring and enrollment. However, in contrast to audit experiments, vignette experiments probed stakeholders outside the immediate educational context: school board members and parents. Inman and colleagues (1998) presented a scenario that described a Black, Latino, and White candidate experiencing racial exclusion in their candidacy for the school board to a sample of Black, Latino, and White male participants. The vignettes also manipulated the power of the fictitious perpetrator

by describing an elected board member versus a civilian. After reading the vignette, White and BIPOC participants were more likely to identify exclusion as race-based against the Latino candidate. Participants' rejection of the race-based exclusion was particularly high when it was exhibited by a person in a position of power (i.e., a current school board member versus a community member; Inman et al., 1998).

In addition, a vignette experiment probing White and BIPOC parents presented a scenario that encouraged them to imagine sending their 5-year-old to school the following year (Billingham & Hunt, 2016). The ethnic-racial composition of the school was manipulated. The results revealed that parents were less likely to enroll their children in schools with higher proportions of BIPOC students (Billingham & Hunt, 2016).

Experimental Tasks

Experimental tasks model situations in which the participant must make choices or decisions that affect the outcomes for themselves or the group around them (Gozlie, 2019; Maclin, 2020; Sailer et al., 2017). Experimental tasks have been used to understand the processes and mechanisms underlying dyadic and group interactions (Gozlie, 2019; Maclin, 2020; Sailer et al., 2017). Tasks vary from audit and vignette experiments because they require participants to complete brief assignments with a brief or no cover story (Janssen et al., 2022; Morris et al., 2022). Experimental tasks often take place in an unfamiliar setting, such as a laboratory. Thus, similar to vignette experiments, participants are aware of their participation in the examination. However, experimental tasks may facilitate the feeling of being observed and have increased social desirability bias (Janssen et al., 2022; Morris et al., 2022). Further, participation in forced-choice tasks may not mimic real-world situations.

Experimental tasks required respondents to finish a given assignment (e.g., grading an essay that varied by the race-ethnicity of the writer) or to relay opinions across several prompts (e.g., deciding which ethnic-racial teacher's course to enroll; Janssen et al., 2022; Morris et al., 2022). Experimental tasks utilized a number of tools to carry out the experimental design: text-based materials, verbal discussion and observations, and computer software games. To manipulate the race of the target, tasks typically use racially salient names, photos, or explicitly name the race of the target. Similar to audit and vignette experiments, validating the racial salience of the treatment is important for the experiment's integrity (Janssen et al., 2022; Morris et al., 2022). Interestingly, experimental task examinations often vary in their inclusion and complexity of an experimental cover story. That is, some examinations craft a scenario to accompany the experiment, creating a more realistic reason to complete the task. However, other implementations provide no cover story and only require the participant to complete the task at hand. Experimental tasks have been used to probe racism in educator-student relationships, peer-student relationships, and educational curricula.

Racism in educator-student relationships. Experimental tasks have been used to probe racism in teachers, counselors, and principals. To assess racial attributions, pictures and brief texts depicting target students were presented to teachers and counselors to assess racial bias in attributions about their academic potential and behaviors. The results revealed that the sample of White and Black preschool teachers exhibited negative attributions about Black student targets academic potential when compared to White student targets, an effect that was particularly present for Black boys (Adams, 1978). Further, results revealed that the sample of White and BIPOC counselors

exhibited more positive expectations for Asian student targets' academic potential when compared to Black and White student targets (Chen & Weseley, 2011).

Experimental tasks have also been used to probe racial disparities in grading. Teachers, counselors, and admissions personnel were tasked with grading student targets' assignments, reviewing transcripts for advanced placement, and evaluating school program acceptance. The results revealed that the sample of White and BIPOC teachers exhibited a pro-Black bias in grading assignments, such that Black student targets received higher scores than their White counterparts for the same assignment (Haughton, 2014). Findings from a sample of White and BIPOC admissions personnel find a similar pattern, with the results from a task of rating honor society applicants revealing a pro-Black bias in acceptance rates for Black student targets when compared to White student targets (Axt, 2017). However, findings from a sample of White and BIPOC counselors do not follow the same trend. Guidance counselors provided harsher evaluations and fewer recommendations for Black female student targets regarding advanced placements compared to White student targets (Francis et al., 2019).

Racism in peer relationships. Experimental tasks have also been used to probe peer relationships. Such experiments are most common in collegiate settings (see Bell et al., 2019; Cunningham et al., 2009; Ruscher et al., 2010; Vanman et al., 1997; Harber, 1998; Sadler, 2002). However, one examination exists in pre-K-12th grade contexts (McGlothlin et al., 2005). White students' attributions toward peer targets were probed by presenting images to convey the race of the peer (McGlothlin et al., 2005). White students' liking, skepticism, and stereotypical attributions of their peers were recorded.

The findings revealed no evidence of White students' racism toward Black versus White peer targets (McGlothlin et al., 2005).

Racism in educational curriculum. Lastly, experimental tasks were used to probe how school curriculum contributes to racism. White and BIPOC participants were tasked with reading from textbook passages manipulated to be racialized with the following: 1) essentialist conceptualizations of race and 2) racialized genetics (Donovan, 2014, 2016). Findings revealed that racialized passages elevated middle and high school students' negative beliefs about BIPOC compared to students who did not read racialized passages (Donovan, 2014, 2016).

Confederate

Confederate experiments include research actors selected to differ on a specific domain (e.g., race). The confederate actors are trained and matched to participate in the study with actual participants. The use of confederate actors creates social contexts where participants' reactions to the designated manipulation can be captured (Highhouse, 2009; Imai et al., 2008; Kuhlen & Brennan, 2013; Reichardt, 2019). Though implementing the most control of all the methods, confederate experiments may still be susceptible to social desirability bias (Janssen et al., 2022; Morris et al., 2022). Though the confederate interaction simulates a real-world experience, engaging in a laboratory setting may create unnatural participant reactions. In addition, the cause of the effect can be confounded with the confederate mood, unconscious behaviors, and inaccurate timing of the confederate performance.

Racism in peer relationships. Confederate experimental designs have been used to assess racism in peer interactions by having a trained White or BIPOC confederate

implement a racially charged confrontation or initiate a conversation on racialized topics (Janssen et al., 2022; Morris et al., 2022). In addition, confederate experimental designs have also been used to examine how White students exhibit individual racism and White privilege in public versus private settings. To my knowledge, confederate experiments have only been conducted in collegiate contexts and reveal mixed findings of the presence of racism (Blanchard et al., 1994; Durm & Barrios, 1999; Hackney & Glaser, 2013; Harber, 2004; Littleford et al., 2005; Moss & O'Brien, 2018; Zabel et al., 2015).

For example, confederate experiments simulating racial confrontations among student peers have revealed that White students were more receptive to feedback from a White versus Black confederate peer (Durm & Barrios, 1999). Yet, the results from Harber (2004) suggest White students had more positive attributes toward Black versus White confederate peers, even when the Black confederates exhibited disagreeable and argumentative behavior.

Further, confederate experiments examining how White students exhibit individual racism and White privilege in public and private settings have also revealed mixed results. White students participated alongside a Black or White confederate peer to answer questions about both campus racial harassment and their own scores on implicit bias tests (Blanchard et al., 1994; Moss & O'Brien, 2018). The confederate either explicitly acknowledged racism or did not. The results revealed that both hearing a peer condemn campus racism or hearing a peer admit to possessing implicit racial bias increased the White student's own likelihood of also acknowledging such racism (Blanchard et al., 1994; Moss & O'Brien, 2018). Hackney and Glaser (2013) examined how White students would react in a classroom where the professor racially profiled a

Black confederate for cheating on an assignment. The results revealed that when professors explicitly profiled the Black confederate for cheating, White students were more likely to exercise their racial privilege and cheat on the assignment than when profiling was missing or when White confederates were targeted (Hackney & Glaser, 2013).

Natural Experiments

The final experimental technique used to probe racism in educational contexts is natural experiments. In contrast to the other experimental techniques, the treatment in natural experiments is not manipulated by the researcher. Instead, the treatment and control conditions occur naturally or outside the researcher's control (Thapar, 2019). To capture cause and effect, researchers examine data from before and after the natural event. Researchers examine treatment effects by identifying existing data that was measured both before and after the naturally occurring treatment. Evidence from natural experiments informs understanding of institutional racism in U.S. education.

Racism in educational policy. Natural experiments have identified racism in school policy. Hoffman (2014, 2016) investigated how the implementation of a zero-tolerance disciplinary policy impacted White and BIPOC student expulsion and suspension rates. The policy change removed authority from teachers and guidance counselors and placed the responsibility on principals, who were required to adhere to rigid disciplinary guidelines. Data collected two years prior and one year after policy implementation were utilized. Results revealed that Black students were disproportionately impacted by this policy when compared to White, Latino, Asian, and

Native students. That is, suspension and expulsion rates increased for Black students because of the policy implementation (Hoffman, 2014, 2016).

Further, Tuttle (2020) examined the racist effects of a school initiative to integrate a majority-White county district and a majority-Black city district. The implemented initiative randomly assigned students to be bussed to new schools or stay at home. Assignments were conducted using the student's grade level and the first letter of their last name. The author utilized school, yearbook, and confidential census data to link student assignments in elementary school to the same person once they reached adulthood (Tuttle, 2020). Findings revealed that the experience of busing to a city school had little effect on White students' academic and professional trajectories in adulthood; however, evidence of White flight or White disenrollment from the integrated schools was evident after the initiative's implementation (Tuttle, 2020).

Meta-Analysis

The results from these five experimental techniques–audit experiments, vignette experiments, task experiments, confederate experiments, and natural experiments–suggest that racism is exhibited by peers, teachers, administrators, and school policies, hiring, and curriculum. Such findings offer objective evidence of the institution of racism in the education system or the combination of racist behaviors and policies that work together to inhibit BIPOC's academic trajectories. However, research utilizing the five experimental techniques has also reported mixed findings. The summarized findings offer contradicting evidence of who may exhibit racism. Moreover, across experimental techniques, findings of the perpetrators of racism are often conflated with the inclusion of both White and BIPOC participants as perpetrators. This is particularly problematic for

examinations of individual racism because individual racism exhibited by Whites is reinforced by societal power and dominance (Harrell, 2000), whereas discrimination exhibited by BIPOC is distinct. Experimental understanding of how racism is perpetuated in educational institutions will remain limited until examinations adopt this lens.

In addition, experimental evidence of racism is spread across a variety of educational domains. Racism has been examined in grading decisions, academic tracking, behavioral and cognitive disability diagnoses, and perceptions of liking (Sullivan et al., 2019; Golson, 2022; Fisher, 2019; Janssen et al., 2022; Sedlacek, 2021; Copur-Gecturk, 2020; Francis, 2019; Hailey, 2022). Racism operates uniquely across these educational domains, which can create inconsistent interpretations of the presence of racism. For example, racism motivates the disproportionately high discipline and academic tracking of Black and Latine students compared to White students but motivates the low grading and low ADHD diagnosis of Black and Latine students compared to White students (Townsend, 2000; Travers et al., 2013). In addition, racism uniquely oppresses each ethnic-racial group in the education system based on distinct histories and racial stereotypes (Takaki, 1993). For example, Asian students are plagued with the model-minority myth, which is in contrast to Black students who are stereotyped as uninterested in academic achievement (Ladson-Billings, 2009; Takaki, 2008). Despite these important distinctions in how racism operates across educational domains and ethnic-racial groups, previous research has done very little to summarize experimental findings. Of the aggregations that do exist, quantitative summaries are restricted to a single domain (e.g., grading; Malouff & Thorsteinsson, 2016; Tenenbaum & Ruck, 2007)

Further, experiments on racism in educational settings have been conducted in a variety of U.S. states and regions. There is experimental evidence that suggests racism is more prevalent in educational contexts in the southern than those in other U.S. regions (Janssen et al., 2022). However, experimental work also documents higher rates of racism in other regions, such as the Midwest (Giulietti et al., 2019). Importantly, evidence documenting regional differences in racism in educational contexts is currently only available from large, multi-state experimental examinations – primarily audit examinations, given their use of electronic correspondence. Other experimental methods that utilize more traditional sampling strategies (i.e., vignette and task experiments) often present findings from a single U.S. state or region. Thus, understanding about the way in which the presence and prevalence of racism varies across educational contexts in the U.S. is unclear.

Finally, the evidence suggests that the type of experimental manipulation may contribute to identifying evidence of racial discrimination. Experiments that elicit participants' own expression of racial discrimination versus experiments that manipulate the witnessing of racial discrimination seem to have important implications for the findings. Experiments manipulating the witnessing of racial discrimination to assess participants' reactions often found that students, counselors, and teachers rejected the expression of racial discrimination in educational contexts. However, evidence of racial discrimination was found from participants in experiments where the manipulation elicited participants' own racially discriminatory behaviors and attitudes.

To address these gaps in scientific understanding, a meta-analysis is needed to estimate the effect size of racism in pre-K-12th grade educational contexts and examine

potential moderators of the effect. Meta-analytic techniques provide a method to quantitatively summarize results across studies identified in systematic reviews (Card, 2017), emphasizing estimation rather than relying on significance testing (Cumming, 2013). The goal of meta-analyses is to estimate the magnitude of an effect and to explain the heterogeneity between studies (see Gurevitch et al., 2018).

The present meta-analysis will allow for the quantification of the effect of the occurrence of racism in the U.S. pre-K-12 education system provided by experimental methodologies (e.g., natural experiment, quasi-experiment, true experiment). Additionally, meta-analytic techniques afford examination of sample characteristics (e.g., primary, secondary) and study characteristics (e.g., natural experiment, quasi-experiment, true experiment), which alter or moderate the magnitude of the effect. In addition, given the increase in statistical power associated with meta-analytic techniques (Cohn & Becker, 2003), the estimates resulting from the present study are more precise and generalizable than a single study.

As of this writing, there are no published meta-analyses on experimental examinations of racism in educational institutions that aggregate findings across educational domains and across perpetrators. Thus, a quantitative aggregation is needed in order to summarize objective evidence of racial discrimination across pre-K through 12th-grade educational contexts. Estimating the magnitude of racism and its moderators in U.S. pre-K-12th grade education will identify specific barriers and persons that inhibit BIPOC students' academic trajectories. Further, these results would advance the notion of institutional racism, or that racism from multiple sources in the educational system intersects to create an institution that disenfranchises BIPOC.

The Current Review

The present study utilized meta-analytic techniques to quantitatively summarize the results of experimental examinations of racism across pre-Kindergarten through 12th-grade educational contexts. The present meta-analysis proposed two aims. First, to explore the magnitude of the effect of racism in pre-K through 12th-grade educational contexts as demonstrated by experimental examinations. Second, the following moderators of the effect were explored: 1) the five experimental techniques, 2) the specific perpetrator (e.g., peer, teacher, counselor, principal), 3) BIPOC target (e.g., Black, Asian, Latina/o/e, Native), 4) region of the U.S., and 5) type of elicited response (i.e., participant's direct response, participant's vicarious response).

Method

This meta-analysis was conducted following PRISMA (2015) guidelines and was pre-registered in PROSPERO on December 28, 2022[1].

Inclusion and Exclusion Criteria

Variables

Studies were included if the experimental manipulation sought to uncover White racist attitudes or behaviors, White perpetuation of racism, or racism in educational policies, procedures, or curriculum. Studies with participant self-report of racial attitudes, racial bias, or racial discrimination were excluded. Studies were included if the experimental manipulation probed one or more of the following BIPOC as the targets of racism: Asian, Black, Latina/o/x, Native, Pacific Islander, bi- and multi-racial, and/or a

[1] The preregistered protocol can be viewed at https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022369194

consolidated group of two or more BIPOC groups. Studies in which the experimental manipulation probed White and non-Hispanic individuals as the target of racism were excluded, given such investigations defy the definition that racism may not be experienced by Whites given that they are the dominant racial group (see Harrell, 2000). Studies were included if the experimental manipulation or cover story was explicitly related to the educational context at pre-K-12th grade educational institutions in the U.S. Studies that examine collegiate contexts were excluded, given that collegiate education is not compulsory.

Population and geographic location

The population of interest included Whites within the U.S. education system, and studies were included if White participants were in the sample. Studies that examined only BIPOC as participants were excluded, given that the findings do not contribute to the understanding of White's perpetuation or complicity with racism in U.S. education. Unfortunately, the majority of the literature examining racism in the education system utilize samples of White and BIPOC individuals. Mixed ethnic-racial samples from studies that examined *institutional racism* (e.g., principal expulsions, teacher enforcement of harsh discipline or grading, educational policies) were included in the present meta-analysis. As stated in our definition of racism (Harrell, 200), components highlighting the centrism of power in racism, such experimental evidence can be attributed to an individual's enactment of the institution's power via the enforcement of racist policies, practices, and procedures.

Studies that examined *individual racism* in mixed ethnic-racial samples were included only if disaggregated statistics for the White participants were provided. Thus,

studies exploring student-to-peer perpetration, teacher-to-fellow-teacher perpetration, and parent-to-school perpetration of racial discrimination were included only if statistics for a White subsample could be extracted from the report or obtained from the author.

Studies examining populations outside of the United States were excluded, given that racism in the U.S. may be unique from manifestations of racism in other countries or contexts.

Research Method

Only studies that employed an experimental method were included. Interventions were not included. Systematic reviews and meta-analyses, qualitative studies, commentaries, reviews, editorials, and corrections were excluded based on the methodological inappropriateness of the research methodology.

Experimental Treatment Design

Studies were not excluded by experimental treatment design. Specifically, studies using a between-subjects design or a within-subjects design were included. In a between-subjects design, participants are assigned to a single treatment condition, often at random (Bal et al., 2011). For example, half of the participants may be exposed to the White target, and half of the participants may be exposed to the Black target. In a within-subjects design, participants are exposed to two or all treatment conditions (Bal et al., 2011). For example, participants are exposed to both White and Black targets.

Language

Studies were included if they were published in English only.

Time Period

All studies, regardless of the date of publication, were included.

Animal Studies

Only studies examining human participants were included.

Publication Status

Published studies, dissertations, theses, unpublished briefs, working papers, and conference presentations were included.

Moderators

The current study examined whether racism varied as a function of 1) the five experimental techniques, 2) the specific perpetrator (e.g., peer, teacher, counselor, principal), 3) the BIPOC target (e.g., Black, Asian, Latina/o/e, Native), 4) the region of the US, and 5) type of elicited response (i.e., participant's direct response, participant's vicarious response).

Literature Search

Peer-reviewed reports were obtained from Psych Info, Web of Science, ERIC, and PubMed. The search was completed up to January 1, 2023. Unpublished manuscripts were identified by searching ProQuest Dissertations & Theses Global to reduce publication bias. Search strings and relevant restrictions used for each database can be viewed in Table 1.

Unpublished reports were also identified by reviewing the American Educational Research Association (AERA) conference paper and presentation repository from conferences spanning the past five years (i.e., 2018 and 2022). The final method of report collection identified relevant published and unpublished reports from the reference lists of all included records (i.e., backward and forward searches).

All reports identified from the database searches were extracted and imported to EndNote. EndNote was used to remove duplicate records. Following the removal of duplicates, the remaining reports were imported to Rayyan. Rayyan is an online software that assists with meta-analysis projects. Abstracts were scanned to determine inclusion or exclusion in Rayyan. For reports identified from the AERA repository or backward and forward searches, the first author scanned the title and abstract to determine inclusion or exclusion.

For all reports included in the present meta-analysis, two trained coders utilized a coding scheme informed by the inclusion and exclusion criteria to tag each report with an inclusion decision or exclusion rationale (Appendix A). The coders overlapped on 24% of the reports for reliability. Cohen's Kappa (κ) was calculated as a measure of interrater reliability of the record assessment and indicated good agreement between the raters.

Data Extraction

A coding scheme was developed before the start of data extraction (see Appendix B). All records were coded by the first author and were confirmed by a trained graduate student. Any discrepancies were resolved through discussion among the coders. The coding scheme includes report-level information such as publication status and report year; experiment-level information such as sample characteristics, participant assignment, manipulation characteristics, and any validity and reliability information provided for the manipulation; and finally, statistics relevant for calculating effect sizes. Further, a copy of the manipulation used in each study was obtained (See Supplementary Materials). This collection of experimental manipulations can be used to further understand study effect

sizes and can also serve as a repository for future experimental study replications and extensions.

Sample and Treatment Sizes

The sample size for the overall study was extracted as the number of total participants. In addition, the cell sizes for each treatment were extracted. For between-subjects designs, cell size is the number of participants in each condition (e.g., the number of participants assigned to the White vignette). When cell sizes were not given and the report mentions that participants were randomly assigned, cell sizes were estimated by dividing the reported analytic sample size equally across conditions. For within-subjects designs, the cell size was extracted as the number of administered manipulations in each condition (i.e., the total number of vignettes administered to each participant multiplied by the number of participants in the analytic sample). This approach for handling within-subjects designs is suggested by the factorial vignette survey literature (Fish, 2015; Rossi & Nock, 1982) and has been employed in other meta-analyses that examine experimental research (see Bal et al., 2011; Batinovic et al., 2022; Dawtry et al., 2020).

Statistics

The respective means and standard deviations pertaining to all treatments were extracted. If standard errors were reported, they were transformed into standard deviations ($SE*\sqrt{n} = SD$) and recorded. In cases where means and standard deviations were not reported, test statistics were extracted pertaining to the effect or effects of interest (e.g., *t*-values or *F*-values and their respective degrees of freedom; Dawtry et al.,

2020). In cases where effect sizes could not be calculated due to missing statistics, the authors were contacted to obtain relevant statistics.

Computation of Effect Sizes

Effect sizes were calculated using Wilson's (2001) Practical Meta-Analysis Effect Size Calculator (Lipsey & Wilson, 2001). This calculator facilitates the computation of four effect-size types from various input data: the standardized mean difference (d), the correlation coefficient (r), the odds ratio (OR), and the risk ratio (RR). Given that studies may be inconsistent in presenting the findings, all effects were converted to Cohen's *d* before analysis (Causadias & Korous, 2021).

The direction of each effect size was coded based on theory and previous research so that a positive value indicated that the BIPOC target received harsher or less favorable outcomes and a negative value would indicate that the White target received harsher or more negative outcomes. A positive effect size indicates that racism was present. A negative effect size indicates the absence of racism toward BIPOC targets or a pro-BIPOC bias. Importantly, a negative effect size should not be interpreted as *reverse racism* – a "nonsensical construct" given racism is rooted in White power and dominance and a lack of power and dominance for BIPOC (Harrell, 2000; Rothenberg, 1988).

In many cases, multiple effect sizes were extracted from each study[2]. Some studies examined racism across more than one domain and afforded more than one effect size fitting our inclusion criteria (e.g., comparisons of teachers rating the social

[2] In meta-analysis, study effect size are the unit of analysis, not report effect sizes. Thus, a report with multiple studies (e.g., study 1, study 2, study 3) were coded as unique studies and do not contribute to nesting.

adjustment and providing a discipline decision for a Latine versus White target). In these cases, we extracted and calculated a unique effect size for each domain examined. Further, some studies implemented multiple manipulations, disaggregating the effect of race on the outcomes (e.g., the inclusion of Black boy, Black girl, White boy, and White girl targets). In these cases, we extracted and calculated a unique effect size for each matching manipulation to isolate the effect of race (i.e., White girl versus Black girl; White boy versus Black boy). Finally, some studies examined racism toward more than two racial groups. A study that contrasts racism toward Black, Asian, and Latino targets versus White targets, for instance, would produce three effect sizes. In these cases, we calculate an effect size for each BIPOC contrast with White, given our definition of racism (Harrell, 2000). A report-by-report account of the statistics used to calculate each effect size, the direction of each effect size, and the number of effect sizes extracted is available in Appendix C.

Data Analysis Plan

Before examining research questions, the nested nature of the effect sizes of the included studies were aggregated. The nested nature of the data includes effect sizes nested in an ethnic-racial contrast (e.g., outcomes observed for a Native versus White vignette manipulation) and ethnic-racial contrasts nested in studies (e.g., Native versus White and Black versus White manipulations included in a study). Effect sizes within ethnic-racial contrasts were aggregated before analysis to eliminate one layer of nesting. For example, in a given study, effect sizes for multiple dependent variables (e.g., behavior ratings and discipline decisions) observed for the manipulation of a Black versus White target were aggregated to create one overall effect size for the treatment of

Black versus White (a standardized mean difference; Cohen's *d*). Figure 1 depicts the handling of nesting within ethnic-racial contrasts. The Borenstein, Hedges, Higgins, and Rothstein (BHHR; 2009) procedure was used to aggregate effect sizes within ethnic-racial contrasts, given the aggregation was found to be the least biased and most precise (Hoyt & Del Re, 2015)[3].

Given that effect sizes for multiple ethnic-racial contrasts could exist in a given study, data are still nested. Procedures outlined by Assinka and Wibbelink (2006) were employed to assess if multi-level modeling was appropriate. After examining an intercept-only model and estimating the between, within, and across study variance, we estimated the overall effect size in a multi-level and random-effects model. A random-effects model did not assume that all studies tested in the analysis had a common effect size (Lipsey and Wilson, 2001; Raudenbush, 2009). The random-effects specification is recommended if the effect estimated varied because of uncontrollable design features (Lipsey and Wilson, 2001; Raudenbush, 2009). This was in contrast to a fixed-effects model that assumed a constant underlying effect across the whole population (Lipsey and Wilson, 2001; Raudenbush, 2009). Cohen's *d* was used as the measure of average effect size and confidence interval.

The level of heterogeneity between studies was determined by the I^2 index (e.g., level of heterogeneity as a percentage) and then tested for significance using Cochran's heterogeneity statistic (Q). Higgins and colleagues (2003) assign I^2 categories of 25%

[3] Estimates for the models were examined in a four level multi-level framework to test for the sensitivity of the aggregate function (see Appendix D). Level 1: studies, level 2: ethnic-racial comparisons, level 3: effect sizes, level 4: sampling error. The results follow the same trends as those presented in Table 3.

low, 50% medium, and 75% high heterogeneity.

Moderation

Moderators coded in the study were: 1) the type of experimental technique, 2) the specific perpetrator (e.g., peer, teacher, counselor, principal), 3) the BIPOC target (e.g., Black, Asian, Latina/o/e, Native), 4) the region of the US. Each moderator was separately examined (i.e., one predictor at a time), and all moderators were categorical. An effect size for each level of a categorical variable was estimated and compared. All the possible comparisons within each categorical variable were exhausted, and no continuous moderators were examined.

Protocol deviations for moderation analyses. The pre-registered protocol proposed to examine a fifth moderator, the type of elicited response (i.e., participant's direct response, participant's vicarious response). This moderator could not be examined because no experiments identified in pre-K-12th grade explored participant's vicarious response to racism. Thus, the estimated effect size is calculated only for studies that examine participant's direct racist response.

In addition, moderation analyses diverged from the pre-registered protocol by conducting two post-hoc explorations of the perpetrator of racism. The nine types of perpetrators were categorized into two groups: those who can enact the power of the educational institution versus perpetrators who cannot enact the power of the educational institution. This categorization allowed an effect size to be estimated for the prevalence of institutional racism and individual racism in U.S. education. Given study samples for experiments examining individual racism were required to be only White, a final analysis for perpetrator-type was conducted to examine sample composition as a moderator (i.e.,

White sample versus mixed BIPOC and White sample). This analysis was conducted as an attempt to tease apart the level of racism exhibited by the perpetrator from the effect of the perpetrator's own ethnic-racial identity.

Finally, moderation by study design was also examined, which diverged from the pre-registered protocol (i.e., within-subjects versus between-subjects). Previous meta-analyses suggest that this moderation would be significant (Bal et al., 2011; Finkelstein et al., 1995; Gordon & Arvey, 2004). A within-subjects design may impact the process by which participants make comparisons when compared to between-subjects designs, given participants are exposed to both or all of the conditions (e.g., exposed to both a White and Black target). This direct comparison might influence the ratings of the targets by creating a situation in which, all other characteristics being equal, race becomes especially salient.

Publication Bias

The possibility of publication bias was examined. First, a subgroup analysis was conducted to compare the overall effect size of published and unpublished studies. Second, a funnel plot was reviewed to provide a graphical depiction of the symmetry of effect sizes against the size of the sample (Sterne & Egger, 2001). When publication bias is present, the graph will look asymmetric (Sterne et al., 2005). The fail-safe N test was proposed to estimate the number of nonsignificant studies needed to yield a null overall effect size (Rosenthal 1979, 1984). Given models in this analysis are multi-level, the Fail-Safe N could not be performed and a traditional Egger's test could not be examined to test for asymmetry. Thus, a proxy for the Eggers test was conducted by exploring moderation by the study variance.

The syntax used to conduct analyses in the present study can be viewed in Appendix E.

Results

Characteristics of Studies Included

The systematic search of databases and registers yielded 2,585 records from the five databases after removing duplicate records (see Figure 2 for the Prisma Flow Diagram of this search procedure). After being screened for inclusion, 255 reports were sought for retrieval. Of the 253 reports retrieved, 218 reports were excluded because they did not focus on pre-K-12th grade contexts, did not employ an experimental method, were conducted outside of the U.S., did not focus on racism, or reported on data already included from a previous study. In addition, five reports that met the inclusion criteria were excluded because they did not provide sufficient statistics to calculate an effect size and the corresponding author did not reply to the inquiry.

The identification of new records via conferences and backward and forward searches yielded 414 reports. Of the 404 reports retrieved, 372 reports were excluded because they did not focus on pre-K-12th grade contexts, did not employ an experimental method, were conducted outside of the U.S., did not focus on racism, or reported on data already included from a previous report. In addition, ten reports that met the inclusion criteria were excluded because they did not provide sufficient statistics to calculate an effect size and the corresponding author did not reply to the inquiry.

A total of 53 reports were examined. Within these reports, 57 studies were included in the meta-analysis. Ten studies contained two or more ethnic-racial contrasts. The number of extracted effect sizes per ethnic-racial contrast ranged from one to 36. A

total of 343 effect sizes were extracted across all studies (see Appendix C). Effect sizes were aggregated to yield one effect size per ethnic-racial contrast in a study (BHHR, 2009). After aggregating, 71 effect sizes were included in the meta-analysis.

Table 2 provides the characteristics of each study and effect size after aggregation. The total sample in the present analysis is 50,402, and samples ranged from 38 to 19,726. Study samples were located in the West (10.53%, k=6, n=720), Midwest (10.53%, k=6, n=875), South (12.28%, k=7, n=1632), and Northeast (10.53%, k=6, n=775; Figure 3A). In addition, studies included samples from multiple or all regions of the U.S. (38.60%, k=22, n=32,295). The region was not reported in ten studies. Empirical, published peer-reviewed studies were 63.16% percent of the studies (k=36), 33.33% were unpublished dissertations or theses (k=19), and 3.51% were research reports that were not peer-reviewed (k=2; i.e., working and research reports; Figure 3B).

Studies conducted audit experiments (7.02%, k=4), vignette experiments (61.40%, k=35), experimental tasks (29.82%, k=17), and natural experiments (1.75%, k=1; Figure 3C). No confederate experiments were conducted that met our inclusion criteria. The majority of studies used a between-subjects design (68.42%, k=39; Figure 3D). Perpetrators examined by the experimental examinations included: 1) teachers (57.89%, k=33), 2) school counselors/psychologists (15.79%, k=9), 3) students (5.26%, k=3), 4) parents (10.53%, k=6), 5) principals (5.26%, k=3), 6) hiring committees (1.75%, k=1), and 7) school district office (1.75%, k=1).

Studies overwhelmingly included two fictitious ethnic-racial targets, which allowed for a single ethnic-racial comparison (e.g., Black target versus White target; 78.95%, k=45). Some studies included three (k=7), four (k=3), and five (k=2) fictitious

ethnic-racial target groups in their experimental design. However, only 17.54% of studies present multiple ethnic-racial comparisons (k=10), which were primarily Black-White (73.68%, k=42), followed by Latine-White (12.28%, k=7), Asian-White (5.26%, k=3), Native-White (1.75%, k=1), aggregate minority group-White (5.26%, k=3), and multi-racial-White (1.75%, k=1). See Figure 3E for a visual representation of ethnic-racial contrasts.

Average Effect Size

The overall effect of racism was estimated by fitting a three-level meta-analytic model to account for the three different sources of variance modeled in the meta-analytic model: 1) sampling variance at the first level, 2) within-study variance at the second level, and 3) between-study variance at the third level. The overall effect of racism was estimated by an intercept-only three-level meta-analytic model. The traditional interpretation of Cohen's *d* is "*the standardized effect size for the treatment group when compared to the control group"* (Frost, N.d). Thus, in the present examination, an exemplar interpretation could be "*the standardized effect for the disproportionate treatment of BIPOC when compared to White was d".* Alternatively, given the term racism implies the comparison of treatment toward BIPOC when compared to White, Cohen's *d* can also be interpreted as *"the standardized effect of racism toward BIPOC."* This interpretation will be used for clarity given the present examination is focused on racism. Importantly, the interpretation "*the standardized effect of racism toward BIPOC when compared to White was d"* is not an acceptable interpretation because it implies that Whites can experience racism which defies the guiding definition of racism (Harrell, 2000).

The overall estimated effect of racism toward BIPOC in U.S pre-K-12th grade education derived from the 57 studies was $d = 0.15$ (.05) and was positive and significant [t(70)=3.04, p<.01, 95% CI = 0.05, 0.25. The estimated effect indicates the presence of racism toward BIPOC in pre-K-12th grade educational contexts. The overall effect of 0.15 can be regarded as a small effect according to the criteria formulated by Cohen (1988; small $d = .2$, moderate $d = .5$, and large effect $d = .8$). The forest plot of effect sizes can be viewed in Figure 4.

Heterogeneity in Effect Sizes

The intercept-only model estimated that the value of variance between effect sizes within studies was .002 and between studies was .11. The test for heterogeneity revealed significant variation between all effect sizes in the data set [$Q(70) = 593.24$, $p < .001$]. The fit of the intercept-only model was compared to the fit of variance-reduced models. Likelihood ratio tests revealed significant variability between effect sizes within studies ($LRT = 4.07$, $p<.04$) and between studies ($LRT = 15.55$, $p<.001$].

Finally, the distribution of variance over the three levels was examined. The I^2 value was 98.07 percent; 94.37 percent of the total variance can be attributed to differences between studies at level 3 (i.e., between-study variance); 3.70 percent of the total variance can be attributed to differences between effect sizes within studies at level 2 (i.e., within-study variance); and 1.93 percent of the variance can be attributed to variance at level 1 (i.e., the within-study sampling variance). The model indicates heterogeneity between the effects, primarily at the between-studies level, and potential moderators of the effect were examined (Hunter & Schmidt, 1990).

Moderating Effects

All moderators were categorical, and levels of a categorical variable with at least three effect sizes were examined in analyses. The results of moderation analyses can be viewed in Table 3.

Experimental Technique

The experimental technique was examined as a moderator of the overall effect of racism in pre-K-12th grade contexts. Though hypothesized as an experimental technique used to examine racism in pre-K-12th grade contexts, no confederate experiments were identified via this search. Included in the model are vignette experiments (k = 45), audit experiments (k =7), task experiments (k =17), and natural experiments (k = 2). The omnibus test suggests that the effect size did not significantly differ among the experimental method [*Qregression:* F(3, 67) = 0.50, p = 0.68]: vignette experiments (0.13, 95% CI = 0.002, 0.254), task experiments (0.20, 95% CI = 0.02, 0.39), audit experiments (0.08, 95% CI = -0.27, 0.42), or Natural experiments (0.67, 95% CI = -0.44, 1.78). The effect size for natural experiments should be interpreted with caution, given only two studies contributed to the estimate.

Perpetrator

Nine types of perpetrators of racism in education were identified in studies included in this review: 1) teachers (k =36), 2) counselors/psychologists (k =15), 3) district office (k =1), 4) hiring committees (k =2), 5) parents/community members (k = 9), 6) educational policy (k = 2), 7) principals (k = 3), 8) students/pupils (k = 3), and 9) teachers (k = 33). Perpetrator type was not a significant moderator of effect sizes [*Qregression*: F(7, 63) = 1.22, p = 0.30]. The results suggest that effect sizes did not vary

by counselor/psychologist (0.05, 95% CI = -0.23, 0.32), parent/community member (0.21, 95% CI = -0.07, 0.49), principal (0.16, 95% CI = -0.25, 0.57), student/pupil (0.67, 95% CI = 0.26, 1.08), teacher (0.10, 95% CI = -0.03, 0.23), district offices (0.10, 95% CI = -0.57, 0.77), hiring committees (0.15, 95% CI = -0.52, 0.82), policy (0.67, 95% CI = -.43, 1.77). Given the limited number of effect sizes observed for racism from district offices, hiring, and policy, the estimates for these perpetrators should be interpreted with caution.

This moderation was further investigated by aggregating the nine types of perpetrators into two groups: those who can enact the power of the educational institution versus perpetrators who cannot enact the power of the educational institution. As defined in the method section, this dichotomy emerges from the theoretical importance of power in racism (Harrell, 2000). Based on this conceptualization, samples were allowed to include mixed BIPOC-White samples or were restricted to only White samples. The following perpetrators were coded into a single "institutional racism" category (k = 59), indicating their positions of power afford the ability to exhibit racism on behalf of the education system: 1) policies on student outcomes, 2) principals to students, 3) counselors to students,4) schools to parents (e.g., via the provision of information), 5) districts to parents, 6) schools/principals to teachers, and 7) teachers to students. The following perpetrators were coded as a single "individual racism" category (k = 12), indicating they do not have access to the power of the institution when exhibiting racism: 1) parents to schools and 2) students to peers. The moderation was significant [*Qregression*: F(1, 69) = 3.96, p = 0.05]. In particular, the individual racism group, composed of all-White participants, had a larger overall effect size (0.36, 95% CI=0.13,

0.58) when compared to the institutional racism category, composed of a mixed BIPOC-White (0.11, 95% CI = -0.001, 0.21).

To examine if the ethnic-racial composition of the sample was leading this moderation, as opposed to institutional power, the ethnic-racial composition of each study sample was coded as 1=White or 2=Mixed BIPOC-White. In addition to the 12 parent and peer studies examined above, three studies examining teachers and counselors were included in the White categorization. The moderation was not significant [*Qregression*: F(1, 69)= 3.601, p=.06] and did not indicate that studies with an all-White sample (0.31, 95% CI= 0.12, 0.50) exhibited a larger effect of racism than mixed BIPOC-White samples (0.10, 95% CI = -0.01, 0.21). Further, when comparing the effect size for the individual racism group examined previously (parents and students; 0.36) to the all-White sample examined in this moderation (0.31), the size of the effect decreased slightly, suggesting that the ethnic-racial composition of the sample may not be the underlying driver of the moderation effect.

Race of Fictitious Targets

Studies examined racism toward the following fictitious ethnic-racial targets: Black (k =52), Latine (k =10), Native American (k =1), Asian (k =4), Biracial (k =1), and aggregated BIPOC groups (k =3; e.g., Black and Latine). Each of the fictitious targets were compared to a fictitious White target. The omnibus test was not significant [*Qregression:* F(5, 65) = 0.53, p = 0.75]. The results suggest that effect sizes did not vary by Black-White comparisons (0.16, CI 95% = 0.06, 0.26), Latine-White comparisons (0.13, CI 95% = 0.002, 0.25), Native-White comparisons (0.04, 95% CI = -0.72, 0.79), Asian-White comparisons (0.12, CI 95% = -0.12, 0.35), Biracial-White (0.11, 95% CI = -

0.40, 0.61), or aggregated BIPOC group-White comparisons (0.05, CI 95% = -0.12, 0.22). Given the small number of effect sizes for Native American and Biracial targets, estimates should be interpreted with caution.

Region

Study samples were from the Midwest (k =6), the Northeast (k =8), the South (k =8), the West (k =6), and multiple regions (k =32). Region significantly moderated the effects [*Qregression:* F(4, 55) = 2.30, p =.05]. Effect sizes from samples in the South (0.52, 95% CI = 0.25, 0.80) significantly differed from effect sizes from samples in the Midwest (0.04, 95% CI =-0.31, 0.39), the Northeast 0.02, 95% CI = -0.29, 0.34), the West (0.04, 95% CI = -0.26, 0.34), and samples from multiple regions (0.11, 95% CI = -0.05, 0.27). The results revealed that the effect of racism was higher in the southern region.

Participant Assignment

Moderation by study design was examined: within-subjects (k = 24) and between-subjects (k=47). Effect sizes did not vary between studies that implemented multiple treatments to a single participant (within-subjects; 0.17, 95% CI=0.001, 0.34) versus studies that randomly assigned treatments across individuals (between-subjects; 0.14, 95% CI= 0.01, 0.26) [*Qregression:* F(1, 69) = 0.13, p = 0.72].

Publication Bias

Finally, publication status was explored. First, moderation of publication status was explored for published studies (k=47) and an aggregated group of unpublished studies (k=24) that included dissertations, theses, and unpublished research reports (e.g., working papers). Significant differences in the magnitude of effect size between

published and unpublished records were not found [*Qregression:* F(1, 69) = 0.063, p = 0.80; unpublished: 0.13, 95% CI=-0.03, 0.30) and published: 0.16, 95% CI=0.04, 0.28]. Publication bias was also examined via funnel plots of the overall model and a model per each significant moderator (Figure 5 and Figure 6). Though outliers exist, funnel plots were generally symmetrical, which suggests low publication bias in the effect size estimates. Finally, a proxy for the Egger's test was conducted to statistically test the symmetry of estimates. The sample variance was tested as a moderator to approximate this test. The results did not suggest evidence of funnel plot asymmetry (p=0.57), supporting the conclusion that effect size estimates in the present study have low publication bias.

Discussion

The current meta-analysis provides the first quantitative synthesis of multiple experimental methods that assess racism in pre-K-12th grade education. One of the strengths of this meta-analysis is the employment of a random-effects model, which allowed the effect sizes across studies to vary. This technique afforded inferences beyond the limited number of studies included in the analysis and had the potential to identify methods and designs that account for the heterogeneity of effect sizes.

The overall estimated effect size indicated that racism is indeed present in U.S. pre-K-12th grade contexts [Cohen's d = 0.15]. A strength of meta-analysis is that the overall effect size is more precise and generalizable than effect sizes from individual studies (Cohn & Becker, 2003). The increase in power associated with meta-analytic techniques results in a more accurate and reliable estimate of the prevalence of racism in U.S. pre-K-12th grade educational contexts, an estimate closer to the true effect. Further,

the accuracy and reliability of this estimated effect of racism is strengthened, given only experimental evidence of racism was synthesized. Thus, the findings of a positive and significant effect size objectively demonstrate that BIPOC are at risk of experiencing harmful oppression in educational contexts. The overall estimated effect of racism underscores findings from qualitative research and survey research that documents BIPOC personal accounts of experiencing racism from perpetrators in educational contexts (Estrada & Hondagneu-Sotelo, 2011; Hardie & Tyson, 2013; Seaton & Iida, 2019; Wang & Yip, 2020; Zeiders, 2017).

Importantly, there was considerable heterogeneity in the overall effect size estimate, with most of the variation occurring between studies (I^2 between = 94.37). Heterogeneity between studies indicates that there is dispersion in study outcomes (Higgins & Thompson, 2002; Higgins et al., 2003). Some study results were positive, which indicated the presence of racism toward BIPOC, and some study results were negative. In total, 21 of the 71 included effect sizes indicated a negative estimate (ranging from d = -0.002 to d = -0.43). The negative effect estimates found for individual studies indicated the absence of racism toward BIPOC. The negative effect estimates can also be interpreted as indicating a pro-BIPOC bias. Importantly, negative effect estimates do not indicate "reverse racism," a nonexistent construct given Whites cannot be the recipients of racism (Harrell, 2000).

Nevertheless, understanding the emergence of negative effect estimates from experiments examining racism represents an interesting and important area for future theoretical and empirical research. Interestingly, when the study characteristics of the 21 negative effect estimates were examined, no estimates were from experimental

examinations of individual racism or from samples of all White participants. That is, negative effect estimates came exclusively from experimental studies that examined institutional racism and from samples that were mixed BIPOC and White. As noted in the introduction, all individuals holding positions of power within the educational institution were conceptualized as exerting the racist power of the institution (e.g., principals, counselors, and teachers to students). Studies exploring racism from samples of educators in positions of power that are BIPOC and White may be examining a unique construct that is distinct from the current conceptualization of institutional racism. Perhaps a stricter definition of institutional racism is needed, and that racism from teachers, principals, and counselors is, in fact, individual racism that can only be exhibited by Whites. Though an important area for future research, this empirical question will be difficult to explore until researchers adopt the lens provided by Jones (1997) and furthered by Harrell (2000) that Whites are the only beneficiaries of racism, and thus samples of only White participants should be examined in experiments that explore the perpetrators of racism.

Alternatively, the emergence of negative effect estimates – findings of the absence of racism or a pro-BIPOC bias – could be due to compromised internal validity and/or external validity of the experimental manipulations. For example, studies producing negative effect size estimates may not include adequate cover stories for participants in treatment conditions. This may create unnatural experimental conditions or lead to participant suspicion that the study is examining racism. In addition, it may be that studies' experimental materials do not map onto the unique racism faced by the ethnic-racial targets examined. For example, the examination of racism in discipline

decisions is distinct for Black and Latine students versus Asian students. Similarly, the examination of racism in academic tracking and grading is also distinct for Black and Latine students versus Asian students. Finally, it may be that studies producing negative effect size estimates use experimental materials that lack validity and are misaligned with the measured outcome. For example, Neal and colleagues (2003) explored educators' perceptions of Black versus White students' academic potential after showing a video vignette of students walking down the hallway. It isn't clear how this vignette evidenced validity or how it was connected to the outcome of interest. Yet, this study demonstrated a negative effect size estimate.

A third explanation is that studies producing negative effect estimates may have indeed found evidence of a pro-BIPOC bias. For example, it could be possible that participants who exhibit a pro-BIPOC bias are more aware of the ways racism manifests in educational contexts and may intentionally exhibit a pro-BIPOC bias in an attempt to rectify the racism experienced by BIPOC students (e.g., teachers providing higher grades to BIPOC students than White students). The construct of pro-BIPOC bias indicates an area for future research. Understanding if pro-BIPOC biases exist in the education system should be explored by qualitative research to provide adequate nuance needed to understand this construct. Importantly, a pro-BIPOC bias in the education system is not a goal. Racial liberation, or equal rights and status for all ethnic-racial groups, should be the focus of all efforts in U.S. educational contexts (Love et al., 2007; Shivers & Janssen, 2022; Watkins & Shulman, 2008).

If a pro-BIPOC bias indeed exists in certain educational contexts or from specific perpetrators, the underlying motivations should be explored. White guilt has been

identified as a powerful motivator in White adults (Spanierman & Heppner, 2004; Swim & Miller, 1999). White guilt arises from the awareness of the unearned racial privilege afforded to Whites in a racially dominant society (Goodman, 2001; Spanierman & Heppner, 2004). Steele (1990) explained,

> "*Guilt makes us afraid for ourselves and thus generates as much self-preoccupation as concern for others. The nature of preoccupation is always the redemption of innocence, the re-establishment of good feeling about oneself*" (p. 501).

Thus, White guilt as a motivator for action is problematic because it is not truly connected to the abolition of racism. The underlying focus is whiteness (Martin, 2009). Thus, White guilt is unlikely to lead to lasting change for BIPOC or to the advancement of Whites' own racial awareness (Spanierman & Heppner, 2004; Swim & Miller, 1999).

White saviorism has also been identified as a powerful motivator in White adults and, thus, could be a potential underlying motivator for a pro-BIPOC bias in education. White saviorism arises from the need to "save" BIPOC because they are "less fortunate" or because they need saving from themselves (Aronson, 2013, 2017; Cole, 2012). White saviorism has been linked to U.S. White teachers' and parents' attempts to perform as *"good Whites in urban classrooms,"* and as *'heroic liberal warriors who will save students of color"* (Aronson, 2017, p. 37; Hagerman, 2016, 2016; Matias, 2016, p .9). Thus, similar to White guilt, White saviorism ignores institutional and structural racism and instead allows Whites to be in control of who receives support, in what capacity, and for how long. Further, White saviorism inhibits and disregards the long history of BIPOC leading resistance efforts for their own liberation (Aronson, 2017; Cammarota, 2011).

Though understanding the heterogeneity in the present meta-analysis indicates an interesting area for future theoretical and empirical research, it is important to contextualize the posturing about the negative effect estimates of these 21 studies with the overall estimated effect size found in this meta-analysis, which was positive. The overall estimated effect size indicates that when outcomes from individual studies were viewed in aggregate, they indicated the presence of racism in pre-K-12th grade contexts.

Moderation by Experimental Method

The type of experimental method was not found to significantly moderate the effect of racism in pre-K-12th grade educational contexts. This finding may suggest that each experimental method was equally effective at identifying racism. However, estimated effect sizes did show evidence of variability across the methods. Audit experiments were estimated to demonstrate the lowest effect size (d =.08), followed by vignettes (d = 0.13), tasks (d = 0.20), and natural experiments (d =.67). Further, t-tests revealed that the estimated effect size were significantly different from zero for vignette experiments and task experiments.

In contrast, the effect size estimates were not significantly different from zero for audit experiments or natural experiments, though the estimates were in the expected direction. Results from t-tests may suggest that audit and natural examinations are less likely to find evidence of racism in educational contexts than vignette and task experiments. Audit experiments operationalize racism as the presence or absence of a response, which may not be an effective measurement of racism, given racism is becoming more covert (Maron, 2019; Morrison, 2017). Further, natural experiments rely on pre-existing data collected before and after a naturally occurring event. Such data may

not accurately map onto the experience of racism. However, the very low number of experiments using the audit (k=7) and natural (k=2) techniques is likely impacting the significance of the effect size estimates. Audit experiments in other sectors indicate the consistent identification of racism (see Pager, 2007 and Gaddis, 2018). Further, the effect size estimate for natural experiments was actually the largest estimate of all the techniques.

Researchers should consider implementing the audit methodology in future research. Specifically, the audit correspondence technique given it represents a low-cost method of experimentally examining racist behaviors in the education system. Additional audit experiments are needed that probe teachers' responsiveness to students and families and administrative and district responsiveness to families. Audit experiments could also be used to probe racism in student extracurricular inclusion (e.g., student clubs). Further, researchers could extend the audit correspondence methodology to not only explore racism in response rates, but also to explore racism in the quality of responses (e.g., the qualitative coding of email responses for quality, length, and disposition of the response; see Giulietti et al., 2017). Finally, given the low cost of the audit correspondence technique, an audit experiment could be conducted in combination with survey and qualitative research to understand nuances in participants' rationale for their response rates.

In addition, though natural experiments are challenging to conduct, opportunities to utilize this method do exist and should be considered. Specifically, it may be possible for researchers to examine ethnic-racial disparities in discipline decisions that result from zero-tolerance policies and social-emotional learning policies (SEL) and ethnic-racial

disparities that emerge from district transitions to voucher and charter programs. Further, an emerging area for the implementation of natural experiments is to examine racial disparities that emerge from districts implementing color-evasive and racist policies that ban the teaching of curriculum that is explicit about racism (e.g., the true history of U.S. enslavement; Florida's Stop Woke Act; Arizona's CRT Hotline) and curriculum designed to be anti-racist.

Finally, the present meta-analysis proposed to examine the moderating effect of the confederate experimental method. This method is commonly used in collegiate samples to examine the perpetration of racism and to examine reactions to witnessing racism. However, no confederate experiments were identified in this meta-analysis. Understandably, confederate experiments become increasingly challenging with younger participants. Training children to implement a manipulation consistently across trials may not be possible. However, confederate experiments could be conducted in pre-K-12th grade educational conducts to examine racism toward educators, among staff, and in hiring decisions, which are all understudied areas in the experimental literature.

Moderation by Perpetrator

Moderation by perpetrator was examined in three ways. First, each of the eight types of perpetrators were investigated, and the results revealed that effect sizes did not significantly vary across types. However, effect size estimates did show variability, with the smallest effect size estimate emerging for counselors (d = 0.05) and the highest effect size estimates emerging for students (d = .67), an estimate that was also found to be significantly different from zero. Though t-tests for the other perpetrator types were not

significantly different from zero, every perpetrator effect size estimate was in the expected direction, indicating that each perpetrator exhibited some level of racism.

The lack of significance for the moderation analysis for perpetrators was surprising and prompted additional analyses not included in the pre-registered protocol. The eight perpetrator categories were collapsed by the level of racism exhibited: institutional racism or individual racism. This moderation did yield significant results and estimated that the effect of individual racism (d = .36) was more prevalent than institutional racism (d =.11). Racism was more prevalent when exhibited in relationships that could not enforce the power of the institution (e.g., peer-student, teacher-teacher). T-test suggested that both estimates may be different from zero, but only the effect estimate for individual racism was within the p<.05 threshold.

Findings for the level of racism, when situated within the previous discussion about the present study's operationalization of institutional racism, should be interpreted with caution. Indeed, the power dynamics between racism in a peer-student relationship is distinct from racism in a teacher-student or principal-student relationship. However, it could also be argued that racism in a teacher-student relationship is different from racism that results from educational policies. Thus, the present operationalization of individual versus institutional racism may not accurately capture the important distinctions between these levels of racism. If the present operationalization of institutional racism is indeed too lenient, experimental evidence from only white samples would be needed to understand the overall effect of racism in teacher-student, counselor-student, and principal-student relationships.

Finally, studies included in the present analysis primarily investigate the perpetration of racism toward students. Experimental investigations of how adults in the education system exhibit racism toward other adults are scant. An important next step for experimental research is to explore racism in additional educational relationships (e.g., racism among teachers or racism from administrators to teachers). Such understanding is vital to estimate the true overall effect of racism in pre-K-12th grade educational contexts and to understand the ways in which racism is used to exclude all BIPOC from educational attainment and advancement.

Moderation by Fictitious Target Race

Moderation by target race did not indicate significant differences in the overall effect size. The lack of moderation may indicate that the prevalence of racism does not significantly vary by ethnic-racial group in educational contexts. In fact, effect size estimates among the ethnic-racial targets did not vary as much as estimates among other moderators. Further, t-tests indicated that estimated effect sizes for Black targets and for Latine targets were significantly different than zero. Though estimates for Native, Asian, aggregated minority, and Biracial targets did not yield significant t-test results, very few studies examined these ethnic-racial targets. For example, only 1 study examined a Native-White comparison or a Biracial-White comparison. Further, only four studies examined an Asian-White comparison, and no studies individually explored racism toward Asian Americans (i.e., studies included Asian and Latine or Black targets). Given the unique ways in which Asian Americans are discriminated against in the U.S. via the model minority myth (Lee et al., 2020; Lee & Zhou, 2015; Rosenbloom et al., 2004; Thompson & Kiang, 2010), the lack of significance for the effect estimate of racism

toward Asian targets could be due to both the low sample size or to the inappropriate investigation of racism toward Asian Americans.

Overall, the lack of moderation and the relatively similar estimated effect sizes across the ethnic-racial targets has important implications for the common comparison of which ethnic-racial group experiences the *most* racism in the U.S., often referred to as the *oppression Olympics* (Hancock, 2011; Nordin et al., 2022; Yuval-Davis, 2012). The present findings indicate that each ethnic-racial group experiences racism in the education system and, as a result, experiences harmful barriers in their academic pursuits.

Moderation by U.S. Region

The results from moderation analyses imply that the overall effect of racism varied as a function of the U.S. region sampled. The overall effect size for the south demonstrated a more pronounced effect of racism in pre-K-12th education than any other region, with a moderate effect size of .52. That is, the effect estimate indicates racism is the most prevalent in southern educational contexts. Though findings in the literature are mixed, previous research has found higher levels of racism in southern educational institutions when compared to other regions. However, the findings were specifically for racism toward Black populations. Importantly, Black-White contrasts make up the majority of studies in this analysis. Thus, it may be that racism – whose legacy is historically rooted in the enslavement of Blacks in the southern U.S. – is more prevalent against Black individuals in educational institutions in the south than in any other region.

Moderation analysis by the U.S. region did not reveal any other significant differences. Thus, results imply that the effect of racism is similar in the remaining U.S. regions. Indeed, effect size estimates for the other regions were more similar, with the

smallest effect of racism found in samples from the Northeast (d = .02), followed by samples from the Midwest and the Northeast (d =.04), and samples from multiple U.S. regions (d = .11).

Though the present meta-analysis indicates racism may be the most prevalent in the southern region, the results do not suggest the remainder of the U.S. is devoid of racism. That is, the estimated effect size for each region was positive, indicating the presence of racism toward BIPOC across the U.S.. However, t-tests revealed that the effect size estimates for all other regions were not significantly different from zero. Thus, additional research may be needed to understand the unique ways in which racism in educational institutions varies across the U.S.

Further, though moderation by U.S. region provides understanding about the risk of racism in a given geographic area, future research may be more impactful if racism is examined by smaller geographic locales such as state, city, neighborhood, or school district. Though it might be expected that racial attitudes and behaviors would be more similar within a given region than across regions, it is more logical to expect that similarities exist within a given state, city, or neighborhood. Conducting more specific examinations of the relation between geographic location and racism is particularly important for educational contexts, given educational policies and practices are primarily controlled at the state and district level. That is, the officials in states and in individual school districts, who are elected by their constituents, hold most of the power for deciding school curricula, teaching methods, discipline policies, and whether to adopt educational reformations or not (Berliner & Glass, 2014). Thus, evidence that emerges from the examination of these more nuanced geographic locations would provide

important evidence for where racism is the most pervasive and could potentially help inform both teachers' and families' selection of schools. Importantly, the ability to conduct a more nuanced examination of racism by geographic location does require more detailed reporting by researchers. In the present analysis, it was found that many studies did not provide even the U.S. region in which their sample was collected.

Moderation by Participant Assignment

Moderation by treatment design was explored but was not proposed in the pre-registered protocol. During the coding of the included studies, both within-subjects and between-subjects treatment designs were identified. Though within-subject designs facilitate the ability to more clearly demonstrate differences in the treatment of ethnic-racial targets, it has also been suggested that within-subject designs pose a greater risk of suspicion from participants and may result in priming of participants depending on which ethnic-racial target is presented first (i.e., White target or BIPOC target).

Results from the moderation analysis did not reveal significant differences between studies that used a between-subjects versus a within-subjects treatment design. Thus, our results do not provide evidence that one design is superior at documenting racism in educational settings. In fact, effect size estimates were quite similar (i.e., between-subjects 0.14, within-subjects 0.17), and t-tests revealed that the effect size estimates for both designs were significantly different from zero; that both designs provided evidence of racism.

One important consideration for future experimental research is the necessary sample size that is needed for between-study designs. In a within-subjects design, participants experience two or more treatment conditions, which creates larger cell sizes

when analyzing differences in ethnic-racial targets (Bal et al., 2011; Fish, 2015; Rossi & Nock, 1982). In contrast, between-subjects designs administer only one treatment condition per participant (Bal et al., 2011; Fish, 2015; Rossi & Nock, 1982). Thus, a larger sample size is needed for between-subjects designs in order to conduct comparisons across conditions. Thus, obtaining large enough cell sizes per treatment can become increasingly challenging for more nuanced manipulations, a challenge noticed in multiple studies included in the present analysis (e.g., Briscoe-Juin, 2020; Pernell, 1984; Levy, 2004; Saunders, 2021; Golson, 2022; Woods, 2022; Ura, 2022; Okonofua, 2015). Researchers conducting experimental examinations of racism should be mindful of both their overall sample size and the cell size for each treatment condition when both designing their manipulation materials and when conducting analyses. Maintenance of statistical power should always be given priority.

Limitations

Several limitations of the current literature restrict understanding of the prevalence of racism in U.S. pre-K-12th grade contexts. First, studies primarily examined racism toward Black targets and Latine targets. In the nine studies that did examine other ethnic-racial groups, only one study designed their manipulation to investigate the unique oppression faced by that ethnic-racial group (Raymond et al., 1997 examination of Native American versus White). Further, three of the nine studies that explored racism beyond Black and Latine targets created an aggregated profile of the BIPOC targets instead of providing individual comparisons to the White target. The aggregation of ethnic-racial groups does not align with the histories of racial oppression in the U.S. or current theoretical or empirical understanding (Ladson-Billings, 2009; Lee et al., 2020;

Thompson & Kiang, 2010). When BIPOC are collapsed into a single "minority" category, nuances in their experiences when compared to Whites are understandably attenuated, and researchers should stop employing this practice. It is more prudent to examine a single ethnic-racial group and to employ an experimental manipulation that is designed to elicit the racism uniquely prescribed to that ethnic-racial group. This recommendation is especially salient for future experimental investigations of racism toward Asian, Native, Pacific Islander, and multiracial individuals, given the dearth of research that exists for these ethnic-racial groups in educational contexts.

There was also little variation in the type of perpetrator-target relationship examined in the literature. Overwhelmingly, racism in educator-student relationships was examined by experimental research in U.S. education. Though clearly important, the disproportionate empirical attention paid to student experiences leaves gaps in scientific understanding of BIPOC educators' experiences of racism from their colleagues, principals, other administrative committees, and policies. The absence of such experimental explorations severely limits the ability of the present meta-analysis to estimate the true overall effect of racism and limits the ability to fully illustrate the institution of racism that exists in pre-K-12th grade contexts that limits BIPOC progression through primary and secondary school but also limits BIPOC career trajectories in educational contexts.

Further, the present study is limited in that it did not propose to investigate moderation by the type of educational domain examined. Studies included in the present meta-analysis investigated racism via peer-liking, educational tracking, disciplining, cognitive and behavioral diagnoses, enrollment decisions, grading, school and classroom

selection, hiring, and policy endorsement. It is possible that heterogeneity among study effect size estimates is due to the various educational domains examined by experimental examinations of racism in pre-K-12th grade contexts. The present study did not explore this moderation because many studies explored racism across multiple educational domains for a single ethnic-racial contrast (i.e., Black-White). Given the present analysis employs the aggregate function to combine individual effect sizes within a given study, the effect of racism for unique educational domains were aggregated together in many studies. Thus, moderation could not be conducted. However, future meta-analyses could explore differences in racism across educational domains by employing a more advanced multilevel model. If conducted, results from the moderation could provide evidence for which educational domain is most impacted by racism and, therefore, indicate which domain is in most need of educational reform. However, it could also be argued that racism in any educational domain, no matter the size of the effect, is worth reformation.

Implications

The present study has implications for the theoretical understanding of racism in U.S. education. Findings contribute to the theoretical posturing that racism is perpetrated to limit BIPOC from educational power, esteem, and status and to benefit Whites (Harrell, 2000; Jones, 1997). Every effect size estimated in the present study was positive, indicating racism toward BIPOC. Thus, the present synthesis revealed no evidence that Whites were negatively impacted by their race in any U.S. region, from any perpetrator, or in comparison to any other ethnic-racial group. Thus, the findings reinforce theoretical work that states "reverse racism" is a non-existent phenomenon in pre-K-12th grade contexts (Harrell, 2000; Jones, 1997).

In addition, the findings from the present meta-analysis have implications for the reformation of the U.S. education system. At the beginning of this dissertation, I criticized legislators and leaders in the U.S. education system for discounting qualitative and quantitative evidence of racism and prescribing to color evasive and post-racial ideologies (Annamma et al., 2017; Bonilla-Silva, 2015). I argued that initiatives such as intergroup contact in diverse classroom contexts and professional development about multiculturalism were insufficient solutions to reducing racism in U.S. education. Instead, I named that the identification and subsequent modification of persons and policies that perpetrated racism were needed.

However, the landscape of educational policy has changed over the course of this dissertation project. The educational policies I once criticized have little comparison to the educational policies being implemented in many states today. Bans on Critical Race Theory in elementary, middle, and high school education have been implemented in over 44 states (Schwartz, 2023). State-funded and operated hotlines have been created to monitor classroom instruction and ensure racism is not included in lesson plans (e.g., Arizona, Virginia; Associated Press, 2023). Further, bans on curriculum in school districts and in state legislatures have been enforced to limit teachers' ability to discuss race and racism in their classrooms (e.g., Georgia, Florida, Mississippi; USA Facts, 2022). Not only do these initiatives do nothing to reduce the harmful effects of racism in educational contexts, but they make the creation of policies and initiatives that explicitly target the modification of persons and policies that perpetrate racism improbable and, in some states, illegal.

Thus, in the current educational climate, the primary implication of the present meta-analysis is that racism is indeed present in U.S. educational contexts, and it manifests to limit and harm BIPOC individuals. For residents of states like Florida and Arizona, where these harmful policies are being implemented, this finding should be used to inform decisions about whom to elect as school board members, whom to elect as state officials, and even whom to elect in the federal branches of government. Candidates who do not endorse the implementation of explicit, anti-racist curriculum and educator professional development should not be considered for election.

For states or districts unaffected by these policies, the findings from the present analysis provide evidence of the crucial need for racism reduction efforts and educational reform to create educational environments that are promotive for all. The present synthesis of experimental research demonstrates the presence of racism and can be used to prioritize which perpetrators to target for reform and in what locations to begin implementation. However, this study's focus on experimental methods was meant to triangulate scientific understanding already generated by qualitative and survey research. Educational reformation efforts should be inspired by qualitative and survey research. For example, the development of intensive professional development for White teachers, administrators, and staff could utilize BIPOC students' narrative accounts of experiencing educator racism or the subtle ways that educators' complicity with white supremacy manifests in their work. Further, qualitative and survey research could be used to explore racial privilege and racial power afforded to educators. Moreover, pre-K-12th grade curriculum and materials for White students could be created using quantitative and

qualitative research on White racial identity, the history of racism, and the ways in which White students perpetuate racism in educational contexts.

Conclusion

The experimental evidence summarized in the present meta-analysis demonstrates the overall presence of racism in pre-K-12th grade settings and illustrates nuances in the prevalence of racism across experimental techniques, perpetrator type, BIPOC target, and U.S. region. Though the size of the effect of racism varied across moderators, the estimates consistently revealed that harsher and more adverse outcomes were experienced by BIPOC when compared to White targets. These findings support theoretical work that delineates Whites as the sole beneficiaries of racism and provides emerging evidence of the ways in which multiple levels of racism intersect to restrict the academic attainment of BIPOC. Finally, findings from the present meta-analysis demonstrate the importance and utility of experimental methods in racism research. The experimental methodologies summarized in the present study and the experimental manipulations provided in the supplementary materials should be used by all researchers to support scientific explorations of the ways in which racism and White supremacy are perpetuated U.S. education.

Table 1

Database Search Strings and Exclusion Terms

Database	Search String	Exclusions	Reports Retrieved
Psych Info	("racial discrimination" OR "ethnic discrimination" OR "racial prejudice" OR "ethnic prejudice" OR "racial bias" OR "ethnic bias" OR racism OR "racial disparities" OR "ethnic disparities" OR "race-based" OR "ethnic-based" OR "ethnic/racial discrimination" OR "racial/ethnic discrimination" OR "ethnic/racial prejudice" OR "racial/ethnic prejudice" OR "ethnic/racial bias" OR "racial/ethnic bias" OR "ethnic/racial disparities" OR "racial/ethnic disparities" OR "racial-ethnic discrimination" OR "ethnic-racial discrimination" OR "racial-ethnic prejudice" OR "ethnic-racial prejudice" OR "racial-ethnic bias" OR "ethnic-racial bias" OR "racial-ethnic disparities" OR "ethnic-racial disparities") AND (school OR education OR "high school" OR elementary OR primary OR secondary OR "junior high" OR "middle school" OR "preschool" OR "pre-school") AND ("student*" OR "peer*" OR "teacher*" OR "preservice teacher" OR "pre-service teacher" OR "educator*" OR "principal*" OR "staff" OR "faculty" OR "advisor*" OR "participant*" OR "admission*" OR "counselor*") AND (experiment* OR "audit" OR "field study" OR "natural experiment" OR Vignette OR "association task" OR "scenario") NOT title(intervention)	NOT Animal English	656
Pub Med	(("racial discrimination" OR "ethnic discrimination" OR "racial prejudice" OR "ethnic prejudice" OR "racial bias" OR "ethnic bias" OR racism OR "racial disparities" OR "ethnic disparities" OR "race-based" OR "ethnic-based" OR "ethnic/racial discrimination" OR "racial/ethnic discrimination" OR "ethnic/racial prejudice" OR "racial/ethnic prejudice" OR "ethnic/racial bias" OR "racial/ethnic bias" OR "ethnic/racial disparities" OR "racial/ethnic disparities" OR "racial-ethnic discrimination" OR "ethnic-racial discrimination" OR "racial-ethnic prejudice" OR "ethnic-racial prejudice" OR "racial-ethnic bias" OR "ethnic-racial bias" OR "racial-ethnic disparities" OR "ethnic-racial disparities") AND (school OR education OR "high school" OR elementary OR primary OR secondary OR "junior high" OR "middle school" OR "preschool" OR "pre-school") AND ("student*" OR "peer*" OR "teacher*" OR "preservice teacher" OR "pre-service teacher" OR "educator*" OR "principal*" OR "staff" OR "faculty" OR "advisor*" OR "participant*" OR "admission*" OR "counselor*") AND (experiment* OR "audit" OR "field study" OR "natural experiment" OR Vignette OR "association task" OR "scenario")) NOT (intervention[Title])	Humans English	181

Eric	("racial discrimination" OR "ethnic discrimination" OR "racial prejudice" OR "ethnic prejudice" OR "racial bias" OR "ethnic bias" OR racism OR "racial disparities" OR "ethnic disparities" OR "race-based" OR "ethnic-based" OR "ethnic/racial discrimination" OR "racial/ethnic discrimination" OR "ethnic/racial prejudice" OR "racial/ethnic prejudice" OR "ethnic/racial bias" OR "racial/ethnic bias" OR "ethnic/racial disparities" OR "racial/ethnic disparities" OR "racial-ethnic discrimination" OR "ethnic-racial discrimination" OR "racial-ethnic prejudice" OR "ethnic-racial prejudice" OR "racial-ethnic bias" OR "ethnic-racial bias" OR "racial-ethnic disparities" OR "ethnic-racial disparities") AND (school OR education OR "high school" OR elementary OR primary OR secondary OR "junior high" OR "middle school" OR "preschool" OR "pre-school") AND ("student*" OR "peer*" OR "teacher*" OR "preservice teacher" OR "pre-service teacher" OR "educator*" OR "principal*" OR "staff" OR "faculty" OR "advisor*" OR "participant*" OR "admission*" OR "counselor*") AND (experiment* OR "audit" OR "field study" OR "natural experiment" OR Vignette OR "association task" OR "scenario") NOT title(intervention)	NOT Animal English	656
Web of Science	((((ALL=(("racial discrimination" OR "ethnic discrimination" OR "racial prejudice" OR "ethnic prejudice" OR "racial bias" OR "ethnic bias" OR racism OR "racial disparities" OR "ethnic disparities" OR "race-based" OR "ethnic-based" OR "ethnic/racial discrimination" OR "racial/ethnic discrimination" OR "ethnic/racial prejudice" OR "racial/ethnic prejudice" OR "ethnic/racial bias" OR "racial/ethnic bias" OR "ethnic/racial disparities" OR "racial/ethnic disparities" OR "racial-ethnic discrimination" OR "ethnic-racial discrimination" OR "racial-ethnic prejudice" OR "ethnic-racial prejudice" OR "racial-ethnic bias" OR "ethnic-racial bias" OR "racial-ethnic disparities" OR "ethnic-racial disparities"))) AND ALL=((school OR education OR "high school" OR elementary OR primary OR secondary OR "junior high" OR "middle school" OR "preschool" OR "pre-school"))) AND ALL=(("student*" OR "peer*" OR "teacher*" OR "preservice teacher" OR "pre-service teacher" OR "educator*" OR "principal*" OR "staff" OR "faculty" OR "advisor*" OR "participant*" OR "admission*" OR "counselor*"))) AND ALL=((experiment* OR "audit" OR "field study" OR "natural experiment" OR Vignette OR "association task" OR "scenario"))) NOT TI=(Intervention)	English USA	391
ProQuest Dissertations and Theses	abstract("racial discrimination" OR "ethnic discrimination" OR "racial prejudice" OR "ethnic prejudice" OR "racial bias" OR "ethnic bias" OR racism OR "racial disparities" OR "ethnic disparities" OR "race-based" OR "ethnic-based" OR "ethnic/racial discrimination" OR "racial/ethnic discrimination" OR "ethnic/racial prejudice" OR "racial/ethnic prejudice" OR "ethnic/racial bias" OR "racial/ethnic bias" OR "ethnic/racial disparities" OR "racial/ethnic disparities" OR "racial-ethnic discrimination" OR "ethnic-racial discrimination" OR "racial-ethnic prejudice" OR "ethnic-racial prejudice" OR "racial-ethnic bias" OR "ethnic-racial bias" OR "racial-ethnic disparities" OR "ethnic-racial disparities") AND(school OR education OR "high school" OR elementary OR "primary school" OR "secondary school" OR "junior high" OR "middle school" OR "preschool" OR "pre-school") AND ("student*" OR "peer*" OR "teacher*" OR "preservice teacher*" OR "pre-service teacher*" OR "educator*" OR "principal*" OR "staff" OR "faculty" OR "advisor*" OR "participant*" OR "admission*" OR "counselor*") AND abstract(experiment* OR "audit" OR "field study" OR "natural experiment" OR Vignette OR "association task" OR "scenario") NOT title(intervention)	English	701

Table 2

Characteristics of effect sizes and included studies

Study	N	Effect Size (d)	Perpetrator	Target Race[a]	Region	Study Design	Experimental Method	Publication Status
Armstrong (2021)	120	-0.43	Teacher	Black	Northeast	Between-subjects	Vignette	Dissertation
Ash (2023)	179	0.01	Teacher	Black	Multiple	Between-subjects	Vignette	Peer-reviewed
Barry (2020)	179	0.05	Teacher	Black	Multiple	Between-subjects	Vignette	Dissertation
Boyd-Swan (2019) [1]	2720	0.21	Hiring	Black	Multiple	Within-subjects	Audit	Peer-reviewed
Boyd-Swan (2019) [2]	2720	0.09	Hiring	Latine	Multiple	Within-subjects	Audit	Peer-reviewed
Brinkman (2022)	185	0.05	Parent	Black	Missing	Within-subjects	Task	Peer-reviewed
Briscoe-Juin (2020)	53	-0.22	Teacher	Black	Midwest	Between-subjects	Vignette	Thesis
Copur-Gencturk (2022)	989	0.02	Teacher	Black	Multiple	Within-subjects	Task	Peer-reviewed
Copur-Gencturk (2020) [1]	390	-0.02	Teacher	Black	South	Within-subjects	Vignette	Peer-reviewed
Copur-Gencturk (2020) [2]	390	0.00	Teacher	Latine	South	Within-subjects	Vignette	Peer-reviewed
Cox (1996)	288	1.62	Student	Black	South	Within-subjects	Task	Peer-reviewed
Dameron (2018)	334	-0.20	Counselor	Black	Multiple	Between-subjects	Vignette	Dissertation
DeMeis (1978)	68	0.41	Teacher	Black	Missing	Within-subjects	Task	Peer-reviewed
Dunbar (2022)	561	-0.01	Parent	Black	Multiple	Between-subjects	Vignette	Peer-reviewed
Elhoweris (2005)	207	0.34	Teacher	Black	Midwest	Between-subjects	Vignette	Peer-reviewed
Fisher (2019)	138	0.07	Teacher	Black	Missing	Between-subjects	Vignette	Dissertation
Francis (2019)	152	0.01	Counselor	Black	Multiple	Within-subjects	Task	Peer-reviewed
Gilliam (2016)	135	0.37	Teacher	Black	Multiple	Between-subjects	Task	Brief
Giulietti (2019)	19726	0.10	District	Black	Multiple	Between-subjects	Audit	Peer-reviewed
Golson (2022) [1]	229	0.18	Counselor	Black	Multiple	Between-subjects	Vignette	Peer-reviewed
Golson (2022) [2]	229	0.18	Counselor	Latine	Multiple	Between-subjects	Vignette	Peer-reviewed
Golson (2022) [3]	229	-0.22	Counselor	Asian	Multiple	Between-subjects	Vignette	Peer-reviewed
Griffiths (2020) Study One	128	0.06	Teacher	Black	West	Between-subjects	Task	Dissertation

Griffiths (2020) Study Two	97	-0.03	Teacher	Black	Missing	Between-subjects	Task	Dissertation
Hailey (2022) [1]	156	0.32	Parent	Black	Northeast	Within-subjects	Vignette	Peer-reviewed
Hailey (2022) [2]	156	0.26	Parent	Latine	Northeast	Within-subjects	Vignette	Peer-reviewed
Hailey (2022) [3]	156	0.19	Parent	Ag. Minority	Northeast	Within-subjects	Vignette	Peer-reviewed
Harris (2013)	409	1.68	Teacher	Black	South	Between-subjects	Vignette	Dissertation
Hoffman (2014) [1]	12790	1.76	Policy	Black	Missing	Between-subjects	Natural	Peer-reviewed
Hoffman (2014) [2]	12790	-0.21	Policy	Latine	Missing	Between-subjects	Natural	Peer-reviewed
Janssen (2022) [1]	251	-0.06	Counselor	Black	Multiple	Within-subjects	Audit	Peer-reviewed
Janssen (2022) [2]	251	0.06	Counselor	Latine	Multiple	Within-subjects	Audit	Peer-reviewed
Janssen (2022) [3]	251	-0.03	Counselor	Asian	Multiple	Within-subjects	Audit	Peer-reviewed
Jarvis (2020) Study One	85	0.52	Principal	Black	South	Between-subjects	Vignette	Peer-reviewed
Jarvis (2021) Study Two	234	-0.04	Principal	Black	Multiple	Between-subjects	Vignette	Peer-reviewed
Kindaichi (2010) [1]	256	0.32	Counselor	Black	Multiple	Between-subjects	Vignette	Dissertation
Kindaichi (2010) [2]	256	0.46	Counselor	Asian	Multiple	Between-subjects	Vignette	Dissertation
Kindaichi (2010) [3]	256	0.31	Counselor	Biracial	Multiple	Between-subjects	Vignette	Dissertation
King (2004)	77	-0.07	Teacher	Black	Northeast	Within-subjects	Vignette	Dissertation
Kunesh (2019)	98	0.28	Teacher	Black	Midwest	Between-subjects	Vignette	Peer-reviewed
Levy (2004)	38	0.19	Counselor	Black	Missing	Between-subjects	Vignette	Dissertation
Lorenzetti (2021)	233	0.17	Teacher	Black	Northeast	Within-subjects	Vignette	Dissertation
Marcucci (2020)	287	-0.22	Teacher	Black	Missing	Between-subjects	Vignette	Peer-reviewed
Mead (2011)	54	-0.32	Teacher	Black	West	Within-subjects	Vignette	Dissertation
Neal (2003)	136	-0.12	Teacher	Black	South	Between-subjects	Vignette	Peer-reviewed
Oberfield (2021)	3260	0.06	Principal	Black	Multiple	Between-subjects	Audit	Peer-reviewed
Okonofua (2015) Study One	204	0.30	Teacher	Black	Multiple	Between-subjects	Vignette	Peer-reviewed
Okonofua (2015) Study Two	57	0.46	Teacher	Black	Multiple	Between-subjects	Vignette	Peer-reviewed
Perez (2022)	192	-0.12	Teacher	Black	West	Between-subjects	Vignette	Peer-reviewed
Pernell (1984)	275	-0.13	Teacher	Black	Midwest	Between-subjects	Vignette	Brief

Petts (2020)	228	0.07	Parent	Ag. Minority	Multiple	Within-subjects	Task	Peer-reviewed
Quinn (2020)	1549	0.04	Teacher	Black	Multiple	Between-subjects	Task	Peer-reviewed
Raymond (1997)	129	0.04	Teacher	Native	Northeast	Between-subjects	Vignette	Thesis
Rollins (2007)	160	0.02	Teacher	Black	South	Within-subjects	Vignette	Dissertation
Saunders (2021) [1]	88	0.00	Teacher	Black	Multiple	Between-subjects	Vignette	Dissertation
Saunders (2021) [2]	88	-0.18	Teacher	Latine	Multiple	Between-subjects	Vignette	Dissertation
Sedlacek (2021) Study One	70	0.02	Teacher	Latine	Missing	Between-subjects	Task	Peer-reviewed
Sedlacek (2019) Study Two	70	0.21	Teacher	Latine	West	Between-subjects	Task	Dissertation
Shepherd (2016)	128	0.08	Teacher	Ag. Minority	West	Within-subjects	Task	Peer-reviewed
Small (2012)	304	0.25	Parent	Black	Missing	Within-subjects	Task	Peer-reviewed
Steinberg (1980)	128	0.46	Student	Black	Missing	Within-subjects	Vignette	Dissertation
Sullivan (2019) Study One	60	0.27	Counselor	Black	Northeast	Between-subjects	Task	Peer-reviewed
Sullivan (2019) Study Two	106	0.01	Counselor	Black	Midwest	Between-subjects	Task	Peer-reviewed
Sullivan (2019) Study Three	136	-0.19	Counselor	Black	Midwest	Between-subjects	Task	Peer-reviewed
Ura (2022)	164	-0.03	Teacher	Black	South	Between-subjects	Vignette	Peer-reviewed
Valant (2016) [1]	680	0.56	Parent	Black	Multiple	Between-subjects	Vignette	Peer-reviewed
Valant (2016) [2]	680	0.69	Parent	Latine	Multiple	Between-subjects	Vignette	Peer-reviewed
Wang (2020)	190	0.00	Student	Black	Multiple	Within-subjects	Vignette	Peer-reviewed
Woods (2022) [1]	94	0.05	Teacher	Black	Multiple	Between-subjects	Vignette	Peer-reviewed
Woods (2022) [2]	94	0.18	Teacher	Asian	Multiple	Between-subjects	Vignette	Peer-reviewed
Xie (2015)	148	0.34	Teacher	Black	West	Between-subjects	Vignette	Dissertation

Note. Target Race is the ethnic-racial target. All studies use White as the comparison target. [1, 2, 3] denotes multiple ethnic-racial contrast within a study.

Table 3
Summary of Moderators of the Effect of Racism in Pre-K-12th Grades.

Moderator	*k*	*d*	[95% CI]	*t*	$Q_{residual}$	*F*
Experimental Method						
Audit	7	0.08	[-0.27 , 0.42]	0.44	578.17***	0.50
Task	17	0.20	[0.02 , 0.39]	2.17*		
Vignette	45	0.13	[0.00 , 0.25]	2.02*		
Natural	2	0.67	[-0.44 , 1.78]	1.21		
Perpetrator Type						
Counselor	15	0.05	[-0.23 , 0.32]	0.34	541.68***	1.22
District Office	1	0.10	[-0.57 , 0.77]	0.30		
Hiring	2	0.15	[-0.52 , 0.77]	0.46		
Parent/Community	9	0.21	[-0.07 , 0.49]	1.48		
Policy	2	0.67	[-0.43 , 1.77]	0.82		
Principal	3	0.16	[-0.25 , 0.57]	0.79		
Student/pupil	3	0.67	[0.26 , 1.08]	3.26**		
Teacher	36	0.10	[-0.03 , 0.23]	1.61		
Level of Racism						
Individual Racism	12	0.36	[0.13 , 0.58]	3.12**	582.25***	3.96*
Institutional Racism	59	0.11	[-0.01 , 0.21]	1.98+		
Sample Composition						
White	16	0.31	[0.12 , 0.50]	3.20**	579.56***	3.60+
Mixed	55	0.10	[-0.01 , 0.21]	1.74+		
Target Race						
Black-White	52	0.16	[0.06 , 0.26]	3.18**	590.08***	0.54
Latine-White	10	0.13	[0.00 , 0.25]	2.03*		
Native-White	1	0.05	[-0.72 , 0.79]	0.10		
Asian-White	4	0.12	[-0.12 , 0.35]	0.98		
Minority-White	3	0.05	[-0.12 , 0.22]	0.60		
Biracial-White	1	0.11	[-0.40 , 0.61]	0.42		
Region						
Midwest ^{S}	6	0.04	[-0.31 , 0.39]	0.24	478.44***	2.30*
All/Multiple ^{S}	32	0.11	[-0.05 , 0.27]	1.41		
Northeast ^{S}	8	0.02	[-0.29 , 0.34]	0.15		
South ^{M,A,N,W}	8	0.53	[0.24 , 0.80]	3.79***		
West ^{S}	6	0.04	[-0.26 , 0.34]	0.25		
Study Design						
Between-Subjects	47	0.14	[0.01 , 0.26]	2.21*	593.07***	0.13
Within-Subjects	24	0.17	[0.01 , 0.34]	2.08*		
Publication Status						
Published	47	0.16	[0.04 , 0.28]	2.55*	585.54***	0.06
Unpublished	24	0.13	[-0.03 , 0.30]	1.62		

Note. *k* = number of studies; *d* = the overall association between the moderator and the effect of racism; 95% CI = 95% confidence interval; F = value of the test of moderators; $Q_{residual}$ = residual heterogeneity σ^2_1 and σ^2_2 = variance within- and between-study, respectively. Positive effects indicate that greater effect of racism.

Figure 1
Exemplar aggregations of effect sizes within a single report

Figure 2

PRISMA Flow Diagram of Search Procedure

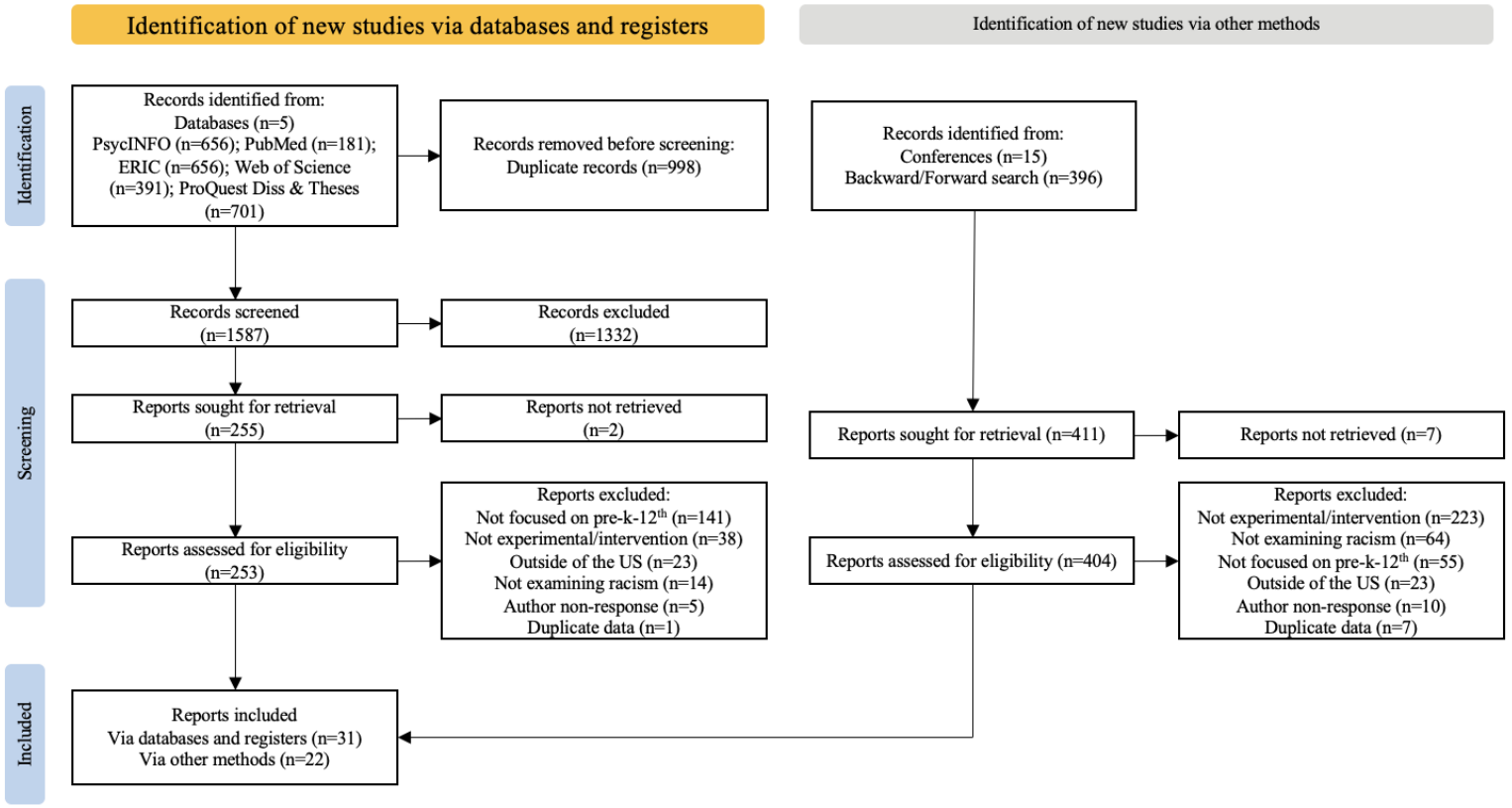


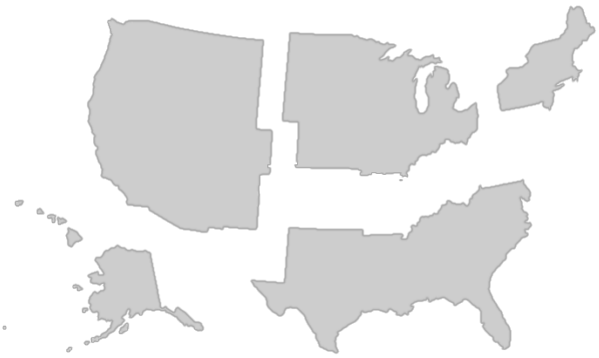
Figure 3

Study Characteristics

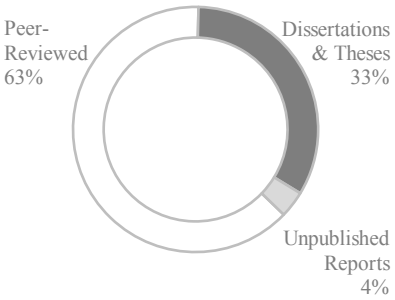
A. Distribution of Studies Across U.S. Regions

U.S. Region	% of Studies	Sample Size
West	11% (k=6)	720
Midwest	11% (k=6)	875
South	12% (k=7)	1632
Northeast	11% (k=5)	775
Multiple regions	39% (k=22)	32,295

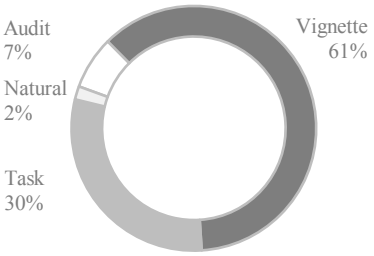
Note. Ten studies did not report the U.S. region of sampling



B. Distribution of Studies Across Publication Status

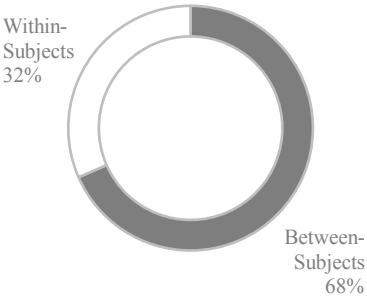


C. Distribution of Studies Across Experiment Type

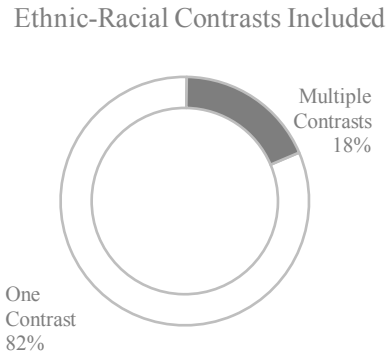


Note. No confederate experiments were included.

D. Distribution of Studies Across Treatment Design



E. Distribution of Studies Across Ethnic-Racial Contrasts



Type of Ethnic-Racial Contrasts Included	% of Studies
White versus Black Target	74%
White versus Latine Target	12%
White versus Asian Target	5%
White versus Native Target	2%
White versus Multiracial Target	2%
White versus an Aggregated Group of BIPOC Targets	5%

Figure 4

Forrest Plot of Effect Sizes

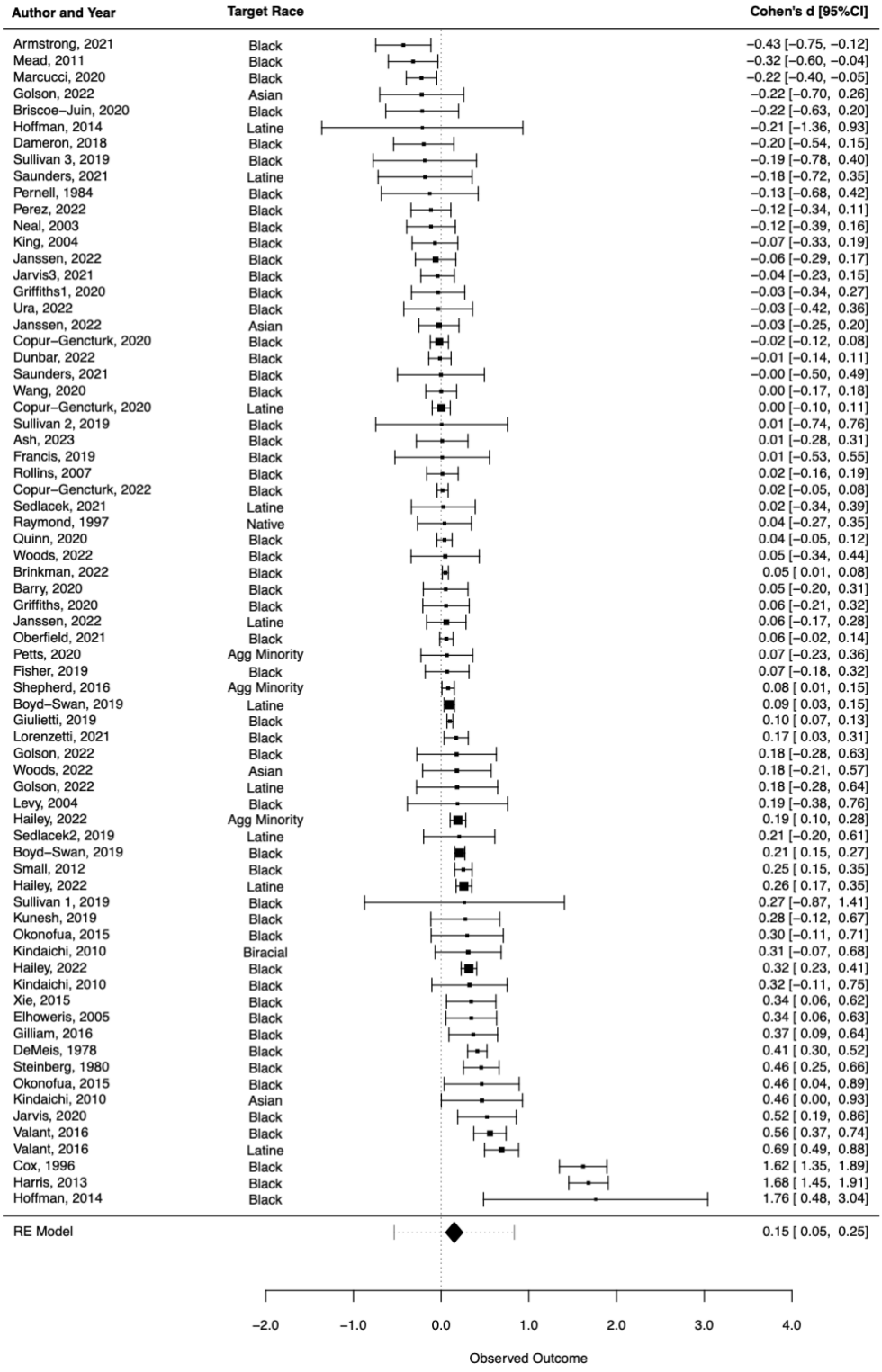


Figure 5

Funnel Plot of Cohen's d Effect Sizes in the Overall Model

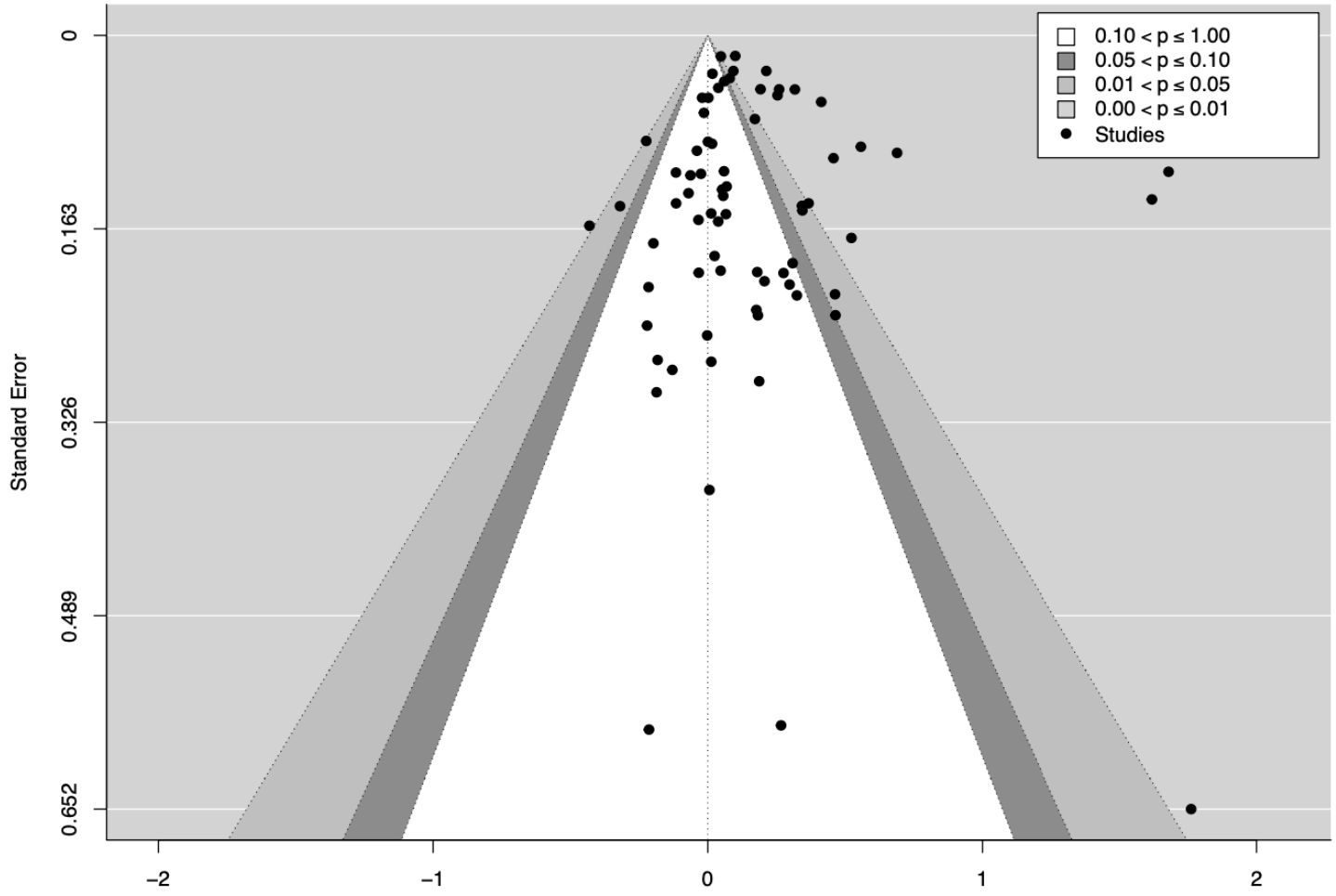
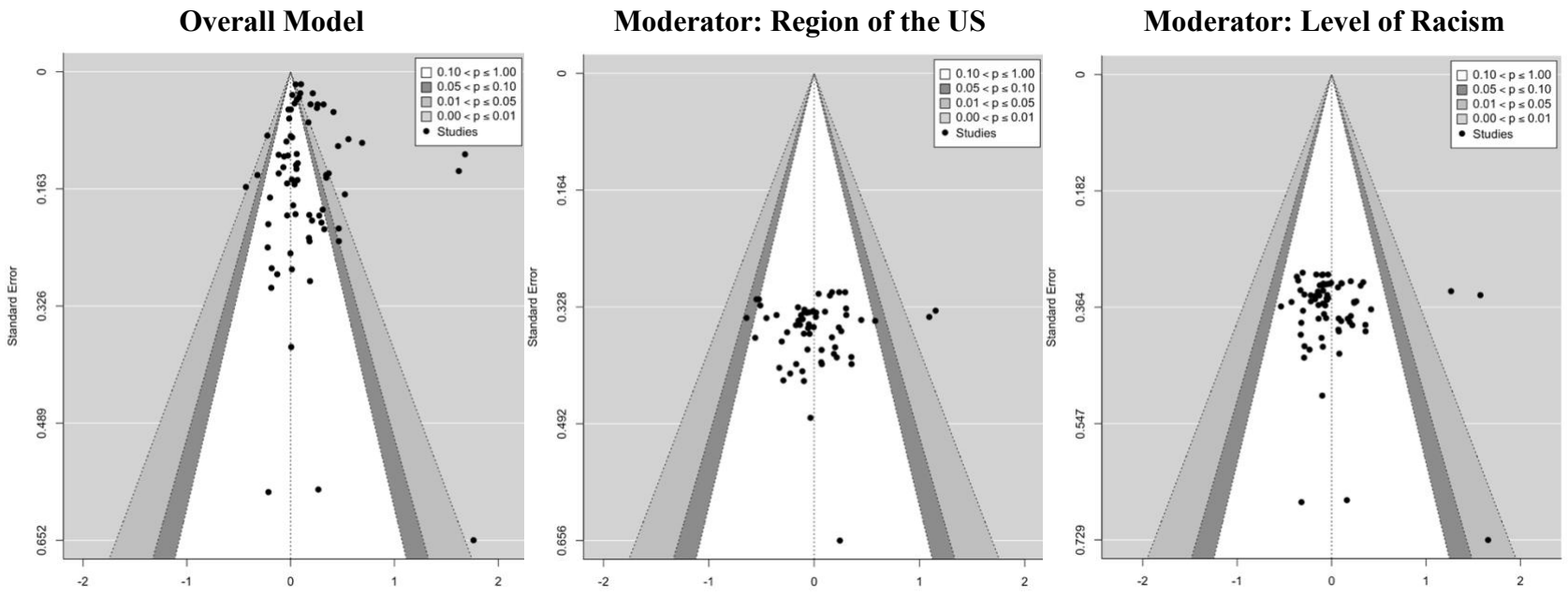


Figure 6

Funnel Plots of Cohen's d Effect Sizes Across Models



REFERENCES

Adams, G. R. (1978). Racial membership and physical attractiveness effects on preschool teachers' expectations. *Child Study Journal*, 8(1), 29–41.

Alonso-Marsden, S. (2017). *Teacher decision-making about student mental health: The role of race and gender* [Doctoral dissertation, Duke University]. ProQuest Dissertations & Theses Global.

Annamma, S. A., Jackson, D. D., & Morrison, D. (2017). Conceptualizing color-evasiveness: Using disability critical race theory to expand a color-blind racial ideology in education and society. *Race Ethnicity and Education*, 20(2), 147–162. https://doi.org/10.1080/13613324.2016.1248837

Armstrong Jr, J. P. (2021). *Hidden Giftedness, Racial Inequity, and Underidentification in Gifted Programming Across a Large Northeastern Metropolitan Area* [Doctoral dissertation, Temple University]. ProQuest Dissertations & Theses Global.

Aronson, B. A. (2017). The white savior industrial complex: A cultural studies analysis of a teacher educator, savior film, and future teachers. *Journal of Critical Thought and Praxis*, *6*(3). https://doi.org/10.31274/jctp-180810-83

Ash, E., & Cranmer, G. A. (2020). Transcontextual Effects of Racialized Sports Media Framing: Students' Off the Field Perceptions of Student-Athletes. *Communication & Sport*, *8*(3), 389-408. https://doi.org/10.1177/2167479519830621

Ash, T. L., Helminen, E. C., Yamashita, S., & Felver, J. C. (2023). Teachers' anti-Black biases in disciplinary decisions: the role of mindfulness. *Journal of school psychology*, *96*, 75-87. https://doi.org/10.1016/j.jsp.2022.11.003

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, 12(3), 154-174. https://doi.org/10.20982/tqmp.12.3.p154

Axt, J. R. (2017). An unintentional pro-Black bias in judgement among educators. *British Journal of Educational Psychology*, *87*(3), 408-421. https://doi.org/10.1111/bjep.12156

Bal, A. C., Reiss, A. E., Rudolph, C. W., & Baltes, B. B. (2011). Examining positive and negative perceptions of older workers: A meta-analysis. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 66(6), 687-698. https://doi.org/10.1093/geronb/gbr056

Barry, N. S. (2020). *Evaluating the Effects of Perceived Student Race on Preservice Teachers' Perceptions of Situation Severity and Requests for Student Assistance*.

[Doctoral dissertation, Mississippi State University]. ProQuest Dissertations & Theses Global.

Barry, N. S., McCleon, T. E., & Stratton-Gadke, K. K. (2020). *Evaluating the Effects of Perceived Student Race on Preservice Teachers' Perceptions of Situation Severity and Requests for Student Assistance* (Publication Number 28255632) [Doctoral dissertation, Mississippi State University]. ProQuest Dissertations & Theses Global.

Batinovic, L., Howe, M., Sinclair, S., & Carlsson, R. (2022). Ageism in hiring: A systematic review and meta-analysis of age discrimination. https://doi.org/10.31234/osf.io/sbzmv

Bell, A. C., Burkley, M., & Bock, J. (2019). Examining the asymmetry in judgments of racism in self and others. *The Journal of Social Psychology*, *159*(5), 611-627. https://doi.org/10.1080/00224545.2018.1538930

Benner, A. D., & Graham, S. (2013). The antecedents and consequences of racial/ethnic discrimination during adolescence: Does the source of discrimination matter? *Developmental Psychology*, *49*(8), 1602–1613. https://doi.org/10.1037/a0030557

Berliner, D. C., & Glass, G. V. (Eds.). (2014). *50 myths and lies that threaten America's public schools: The real crisis in education*. Teachers College Press.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review*, *94*(4), 991-1013. https://doi.org/10.1257/0002828042002561

Biefeld, S. D. (2020). Elementary school children's understanding of disproportionate discipline. [Doctoral dissertation, University of Kentucky]. ProQuest Dissertations and Theses Global.

Billingham, C. M., & Hunt, M. O. (2016). School racial composition and parental choice: New evidence on the preferences of White parents in the United States. *Sociology of Education*, *89*(2), 99-117. https://doi.org/10.1177/0038040716635718

Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology*, 79(6), 993-997. https://psycnet.apa.org/doi/10.1037/0021-9010.79.6.993

Bonilla-Silva, E. (2001). *White supremacy and racism in the post-civil rights era*. Lynne Rienner Publishers.

Bonilla-Silva, E. (2015). The structure of racism in color-blind, "Post-Racial" America. *American Behavioral Scientist*, 59(11), 1358–1376. https://doi.org/10.1177/0002764215586826

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Boyd-Swan, C., & Herbst, C. M. (2019). Racial and ethnic discrimination in the labor market for child care teachers. *Educational Researcher*, 48(7), 394-406. https://doi.org/10.3102/0013189X19867941

Brinkman, C. S., Gabriel, S., Kimelberg, S. M., Poulin, M. J., Valenti, J., & Hugenberg, K. The psychological determinants of classroom preferences: Implicit and explicit racial attitudes and racial composition of classrooms. *Journal of Community & Applied Social Psychology*. https://doi.org/10.1002/casp.2680

Briscoe-Juin, D. G. (2020). *Pre-service Teacher's Implicit Biases and the School-to-Prison Pipeline* (Publication Number 28108224) [Doctoral dissertation, Miami University]. ProQuest Dissertations & Theses Global.

Briscoe-Juin, D. G. (2020). *Pre-service Teacher's Implicit Biases and the School-to-Prison Pipeline* [Doctoral dissertation, Miami University]. ProQuest Dissertations & Theses Global.

Brown, C. S. Bias at school: Perceptions of racial/ethnic discrimination among Latino and European American children. *Cognitive Development*, 21(4), 401-419. https://psycnet.apa.org/doi/10.1016/j.cogdev.2006.06.006

Buckley, E. J. (2019). *The impact of positive behavior interventions and support, counseling, and mentoring on the behavior and achievement of African American males.* [Doctoral dissertation, University of Mississippi]. ProQuest Dissertations & Theses Global.

Calhoun, M. L. (1975). Teachers' causal attributions for a child's hyperactivity: Race, socioeconomic status, and typicalness. *Perceptual and Motor Skills*, *41*(1), 195-198. https://doi.org/10.2466/pms.1975.41.1.195

Cammarota, J. (2011). Blindsided by the avatar: White saviors and allies out of Hollywood and in education. *Review of Education, Pedagogy, Cultural Studies, 33*(3), 242-259. https://doi.org/10.1080/10714413.2011.585287

Card, N. A. (2017). Advances in meta-analysis methodologies contribute to advances in research accumulation: comments on Cheung & Hong and Johnson et al. *Health psychology review*, *11*(3), 302-305. https://doi.org/10.1080/17437199.2017.1345646

Chang, D. F., & Sue, S. (2003). The effects of race and problem type on teachers' assessments of student behavior. *Journal of consulting and clinical psychology*, *71*(2), 235. https://psycnet.apa.org/doi/10.1037/0022-006X.71.2.235

Chatterji, R. (2020). Fighting Systemic Racism in K-12 Education: Helping Allies Move From the Keyboard to the School Board. *American Progress*. https://www.americanprogress.org/article/fighting-systemic-racism-k-12-education-helping-allies-move-keyboard-school-board/

Chavez, F. R. (2021). *The anti-racist writing workshop: How to decolonize the creative classroom*. Haymarket Books.

Chavous, T. M., Rivas-Drake, D., Smalls, C., Griffin, T., & Cogburn, C. (2008). Gender matters, too: The influences of school racial discrimination and racial identity on academic engagement outcomes among African American adolescents. *Developmental Psychology*, 44(3), 637–654. https://doi.org/10.1037/0012-1649.44.3.637

Chen, B., & Weseley, A. J. (2011). The effect of student ethnicity on teacher perceptions. *Journal of Research in Education*, *21*(2), 55-65. https://doi.org/10.3102/0034654319853545

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological methods*, *8*(3), 243. https://doi.org/10.1037/1082-989x.8.3.243

Cole, T. (2012). The white-savior industrial complex. *The Atlantic*, *21*(1). Retrieved from http://media.guidebook.com.s3.amazonaws.com/service/GDt64dhpSUYQb5WXvrAFgc8JTeK2ajEq/TheWhiteSaviorIndustrialComplexGlobalTheAtlantic.pdf

Comeaux, E. (2013). Faculty perceptions of high-achieving male collegians: A critical race theory analysis. *Journal of College Student Development*, *54*(5), 453-465. https://psycnet.apa.org/doi/10.1353/csd.2013.0071

Copur-Gencturk, Y., Cimpian, J. R., Lubienski, S. T., & Thacker, I. (2020). Teachers' bias against the mathematical ability of female, Black, and Hispanic students. *Educational Researcher*, *49*(1), 30-43. https://psycnet.apa.org/doi/10.3102/0013189X19890577

Copur-Gencturk, Y., Thacker, I., & Cimpian, J. R. (2022). Teacher bias in the virtual classroom. *Computers & Education*, *191*, 104627. https://doi.org/10.1016/j.compedu.2022.104627

Cox, C. L., Smith, S. L., & Insko, C. A. (1996). Categorical race versus individuating belief as determinants of discrimination: A study of southern adolescents in 1966,

1979, and 1993. *Journal of Experimental Social Psychology*, *32*(1), 39-70. https://doi.org/10.1006/jesp.1996.0003

Crengle, S., Robinson, E., Ameratunga, S., Clark, T., & Raphael, D. (2012). Ethnic discrimination prevalence and associations with health outcomes: Data from a nationally representative cross-sectional survey of secondary school students in New Zealand. *BMC Public Health*, *12*(1), 45. https://doi.org/10.1186/1471-2458-12-45

Crosby, J. R., & Monin, B. (2007). Failure to warn: How student race affects warnings of potential academic difficulty. *Journal of Experimental Social Psychology*, *43*(4), 663-670. https://doi.org/10.1016/j.jesp.2006.06.007

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Cunningham, G. B., Ferreira, M., & Fink, J. S. (2009). Reactions to prejudicial statements: The influence of statement content and characteristics of the commenter. *Group Dynamics: Theory, Research, and Practice*, 13(1), 59-73. https://psycnet.apa.org/doi/10.1037/a0012967

Dameron, M. L. (2018). *An Examination of the Impact of Student Race, Gender, and Socioeconomic Status on School Counselors' Decisions to Place Students in Alternative Learning Programs* [Doctoral dissertation, The University of North Carolina at Charlotte]. ProQuest Dissertations & Theses Global.

Dameron, M. L., Foxx, S. P., & Flowers, C. (2019). The impact of race, gender, and socioeconomic status on school counselors' alternative learning program placement decisions: An experimental study. *The Urban Review*, *51*, 699-723. https://doi.org/10.1007/s11256-019-00502-9

Dawtry, R. J., Callan, M. J., Harvey, A. J., & Gheorghiu, A. I. (2020). Victims, vignettes, and videos: Meta-analytic and experimental evidence that emotional impact enhances the derogation of innocent victims. *Personality and Social Psychology Review*, *24*(3), 233-259. https://doi.org/10.1177/1088868320914208

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86(1), 195–210. https://doi.org/10.1162/003465304323023750

DeMeis, D. K., & Turner, R. R. (1978). Effects of students' race, physical attractiveness, and dialect on teachers' evaluations. *Contemporary Educational Psychology*, *3*(1), 77-86. https://psycnet.apa.org/doi/10.1016/0361-476X(78)90012-7

Dickter, C. L., Kittel, J. A., & Gyurovski, I. I. (2012). Perceptions of non-target confronters in response to racist and heterosexist remarks. *European Journal of Social Psychology*, 42(1), 112-119. https://doi.org/10.1002/ejsp.855

Donovan, B. M. (2014). Playing with fire? The impact of the hidden curriculum in school genetics on essentialist conceptions of race. *Journal of Research in Science Teaching*, 51(4), 462-496. https://doi.org/10.1002/tea.21138

Donovan, B. M. (2016). Framing the genetics curriculum for social justice: An experimental exploration of how the biology curriculum influences beliefs about racial difference. *Science Education*, 100(3), 586-616. https://doi.org/10.1002/sce.21221

Dunbar, A., Kupchik, A., Hughes, C., & Lewis, R. (2022). Fear of a black (and poor) school: Race, class, and school safety policy preferences. *Race and Justice*, *12*(2), 344-367. https://doi.org/10.1177/2153368719881679

Durm, J. C., & Barrios, B. (1999). *Race and receptivity: An analysis of the convergent and discriminant properties of the scripted dyadic paradigm* (Publication Number 9956066) [Doctoral dissertation, The University of Mississippi]. ProQuest Dissertations & Theses Global.

Elhoweris, H., Mutua, K., Alsheikh, N., & Holloway, P. (2005). Effect of children's ethnicity on teachers' referral and recommendation decisions in gifted and talented programs. *Remedial and Special Education*, *26*(1), 25-31. https://doi.org/10.1177/07419325050260010401

Estrada, E., & Hondagneu-Sotelo, P. (2011). Intersectional dignities: Latino immigrant street vendor youth in Los Angeles. *Journal of Contemporary Ethnography*, 40(1), 102–131. https://doi.org/10.1177/0891241610387926

Finkelstein, L. M., Burke, M. J., & Raju, M. S. (1995). Age discrimination in simulated employment contexts: An integrative analysis. *Journal of applied psychology*, *80*(6), 652. https://psycnet.apa.org/doi/10.1037/0021-9010.80.6.652

Fish, R. E., Ferree, M., & Gamoran, A. (2015). *The Intersection of Race/Ethnicity, Gender and Exceptionality: The Racialized Construction of Educational Disabilities and Giftedness* (Publication Number 10187218) [Doctoral dissertation, The University of Wisconsin - Madison]. ProQuest Dissertations & Theses Global.

Fisher, A. E. (2019). *Students Identities and Teacher Expectations: A Factorial Experiment at the Intersection of Race, Gender, and Ability.* [Doctoral dissertation, University of Kentucky]. ProQuest Dissertations and Theses Global.

Ford, T. E. (1997). Effects of stereotypical television portrayals of African-Americans on person perception. *Social Psychology Quarterly*, *60*(3), 266-275. https://doi.org/10.2307/2787086

Francis, D. V., De Oliveira, A. C., & Dimmitt, C. (2019). Do school counselors exhibit bias in recommending students for advanced coursework?. *The BE Journal of Economic Analysis & Policy*, *19*(4). https://doi.org/10.1515/bejeap-2018-0189

Fredrickson, G. M. (1981). *White supremacy: A comparative study of American and South African history*. OUP USA.

Fryer Jr, R. G., & Levitt, S. D. (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, *119*(3), 767-805. https://doi.org/10.1162/0033553041502180

Gaddis, S. M. (2018). *An introduction to audit studies in the social sciences* (pp. 3-44). Springer International Publishing.

Gee, G. C., & Hicken, M. T. (2021). Structural racism: the rules and relations of inequity. *Ethnicity & Disease*, *31*(Suppl 1), 293. https://doi.org/10.18865%2Fed.31.S1.293

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions. *Yale University Child Study Center*, *9*(28), 1-16. https://marylandfamiliesengage.org/wp-content/uploads/2019/07/Preschool-Implicit-Bias-Policy-Brief.pdf

Giulietti, C., Tonin, M., & Vlassopoulos, M. (2019). Racial discrimination in local public services: A field experiment in the United States. *Journal of the European Economic Association*, *17*(1), 165–204. https://doi.org/10.1093/jeea/jvx045

Glock, S., Krolak-Schwerdt, S., & Pit-ten Cate, I. M. (2015). Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory. *European Journal of Psychology of Education*, *30*(2), 169-188. https://doi.org/10.1007/s10212-014-0237-2

Golson, M. E., Haverkamp, C. R., McClain, M. B., Schwartz, S. E., Ha, J., Harris, B., & Benallie, K. J. (2022). Influences of student race/ethnicity and gender on autism special education classification considerations. *Autism*, *26*(6), 1423-1435. https://doi.org/10.1177/13623613211050440

Goodman, D. J. (2001). *Promoting diversity and social justice: Educating people from privileged groups.* Thousand Oaks, CA: Sage.

Gordon, R. A., & Arvey, R. D. (2004). Age bias in laboratory and field settings: A meta-analytic investigation 1. *Journal of applied social psychology*, *34*(3), 468-492. https://doi.org/10.1111/j.1559-1816.2004.tb02557.x

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

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, *555*(7695), 175-182. https://doi.org/10.1038/nature25753

Hackney, A. A., & Glaser, J. (2013). Reverse deterrence in racial profiling: Increased transgressions by non-profiled Whites. *Law and Human Behavior*, 37(5), 348-353. https://doi.org/10.1037/lhb0000038

Hagerman, M. A. (2016). Reproducing and reworking colorblind racial ideology: Acknowledging children's agency in the white habitus. *Sociology of Race and Ethnicity*, *2*(1), 58-71. https://doi.org/10.1177/2332649215594817

Hagerman, M. A. (2017). White racial socialization: Progressive fathers on raising "antiracist" children. *Journal of Marriage and Family*, *79*(1), 60-74. https://psycnet.apa.org/doi/10.1111/jomf.12325

Hailey, C. A. (2022). Racialized Perceptions of Anticipated School Belonging. *Educational Policy*, *36*(4), 879-910. https://doi.org/10.1177/13623613211050440

Halberstadt, A. G., Castro, V. L., Chu, Q., Lozada, F. T., & Sims, C. M. (2018). Preservice teachers' racialized emotion recognition, anger bias, and hostility attributions. *Contemporary Educational Psychology*, *54*, 125-138. https://doi.org/10.1016/j.cedpsych.2018.06.004

Hancock, A. (2011). *Solidarity politics for Millennials: A guide to ending the oppression olympics*. Springer.

Harber, K. D. (1998). Feedback to minorities: Evidence of a positive bias. *Journal of personality and social psychology*, *74*(3), 622. https://psycnet.apa.org/doi/10.1037/0022-3514.74.3.622

Harber, K. D. (2004). The positive feedback bias as a response to out-group unfriendliness. *Journal of Applied Social Psychology*, *34*(11), 2272-2297. https://doi.org/10.1111/j.1559-1816.2004.tb01977.x

Hardie, J. H., & Tyson, K. (2013). Other people's racism: Race, rednecks, and riots in a Southern high school. *Sociology of Education*, 86(1), 83–102. https://doi.org/10.1177/0038040712456554

Harrell, S. P. (2000). A multidimensional conceptualization of racism-related stress: Implications for the well-being of people of color. *American Journal of Orthopsychiatry*, 70(1), 42–57. https://doi.org/10.1037/h0087722

Harris, J. K. (2013). *Ethnic disparities in teachers' disciplinary approaches towards students with Attention-Deficit Hyperactivity Disorder (ADHD): A quantitative study* [Doctoral dissertation, Capella University]. ProQuest Dissertations & Theses Global.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539-1558. https://doi.org/10.1002/sim.1186

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Highhouse, S. (2009). Designing experiments that generalize. *Organizational Research Methods*, *12*(3), 554-566. https://doi.org/10.1177/1094428107300396

Hoffman, S. (2014). Zero benefit: Estimating the effect of zero tolerance discipline polices on racial disparities in school discipline. *Educational Policy*, *28*(1), 69-95. https://doi.org/10.1177/0895904812453999

Hoffman, M. (2016). How is information valued? Evidence from framed field experiments. *The Economic Journal*, 126(595), 1884-1911. https://doi.org/10.1111/ecoj.12401

Hoffman, S. M., & Cancelli, A. (2019). *Understanding Implicit Racial Bias and Disciplinary Decision-Making in School Mental Health Professionals* (Publication Number 13884893) [Doctoral Dissertation, Fordham University]. ProQuest Dissertations & Theses Global.

Hoyt, W. T., & Del Re, A. C. (2018). Effect size calculation in meta-analyses of psychotherapy outcome research. *Psychotherapy Research*, *28*(3), 379-388. https://doi.org/10.1080/10503307.2017.1405171

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. Jo*urnal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2), 481–502. https://doi.org/10.1111/j.1467-985X.2007.00527.x

Inman, M. L., Huerta, J., & Oh, S. (1998). Perceiving discrimination: The role of prototypes and norm violation. *Social Cognition*, 16(4), 418-450. https://psycnet.apa.org/doi/10.1521/soco.1998.16.4.418

Janssen, J., Seaton, E., Jager, J., & Miller, C. F. (2022). Guidance or Gatekeeping: An Audit Examination of Racial Discrimination in Leading STEM High Schools: Systems that Marginalize BIPOC Youth Special Issue: Dismantling Oppression Series. *Journal of Research on Adolescence*, *32*(2), 625-635. https://doi.org/10.1111/jora.12722

Janssen, J., Seaton, E., Miller, C. F., Causadias, J., & Jager, J. (2022). *Critiquing Experimental Examinations of Interpersonal and Institutional Racism in U.S. Educational Institutions: A Systematic Review.*[Paper presentation]. American Educational Research Association 2022 Annual Conference, San Diego, CA, United States. https://doi.org/10.3102/1895305

Jarvis, S. N., & Okonofua, J. A. (2020). School deferred: When bias affects school leaders. *Social Psychological and Personality Science*, *11*(4), 492-498. https://doi.org/10.1177/1948550619875150

Jones, J. M. (1997). *Prejudice and racism*. McGraw-Hill Humanities, Social Sciences & World Languages.

Killen, M., & Stangor, C. (2001). Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child development*, *72*(1), 174-186. https://doi.org/10.1111/1467-8624.00272

Killen, M., Lee-Kim, J., McGlothlin, H., Stangor, C., & Helwig, C. C. (2002). How children and adolescents evaluate gender and racial exclusion. *Monographs of the society for research in child development*, i-129. https://doi/10.1111/1540-5834.00218

Kindaichi, M. M. (2010). *School counselors' perceptions of biracial students' functioning*. [Doctoral dissertation, Columbia University]. ProQuest Dissertations and Theses Global.

King, M. C. (2004). *The Impact of Race, Culture, and Attitudes on Teacher Judgment of Problem Behaviors* [Doctoral dissertation, Howard University]. ProQuest Dissertations and Theses Global.

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic bulletin & review*, *20*, 54-72. https://doi.org/10.3758/s13423-012-0341-8

Kunesh, C. E., & Noltemeyer, A. (2019). Understanding disciplinary disproportionality: Stereotypes shape pre-service teachers' beliefs about black boys' behavior. *Urban Education*, *54*(4), 471-498. https://doi.org/10.1177/0042085915623337

Ladson-Billings, G. (2009). Race still matters: Critical race theory in education. In *The Routledge International Handbook of Critical Education* (pp. 110-122). Routledge.

Lahey, J. N., & Beasley, R. A. (2009). Computerizing audit studies. *Journal of Economic Behavior & Organization*, *70*(3), 508-514. https://doi.org/10.1016/j.jebo.2008.02.009

Lalani, M., & Bhutta, S. M. (2019). Reducing attitudes of prejudice and discrimination through literature: A quasi-experiment on the upper key-stage 2 students of the British schools in Karachi, Pakistan. *Intercultural Education*, *30*(1), 83–100. https://doi.org/10.1080/14675986.2018.1528772

Lawrence, K., & Keleher, T. (2004,). *Structural racism* [Paper presentation]. Race and Public Policy Conference.

Lee, J., & Zhou, M. (2015). *The Asian American Achievement Paradox*. Russell Sage Foundation.

Lee, T., & Outten, H. R. (2020). The effect of model minority myth salience on White Americans' perceptions of Black Americans and their support for affirmative action. *Psi Chi Journal of Psychological Research*, *25*(1), 2-13. https://psycnet.apa.org/doi/10.24839/2325-7342.JN25.1.2

Levy, M. A. (2004). *The effects of a child's race and socioeconomic status on psychologists' observations, diagnosis and recommendations*. [Doctoral dissertation, Hofstra University]. ProQuest Dissertations and Theses Global.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.

Littleford, L. N., Wright, M. O. D., & Sayoc-Parial, M. (2005). White students' intergroup anxiety during same-race and interracial interactions: A multimethod approach. *Basic and Applied Social Psychology*, *27*(1), 85-94. https://doi.org/10.1207/s15324834basp2701_9

Lorenzetti, N. L. (2021). *Teachers' Implicit Racial Attitudes and Classroom Discipline Referral* [Doctoral dissertation, City University of New York]. ProQuest Dissertations & Theses Global.

Lott, B. E., & Maluso, D. E. (1995). *The social psychology of interpersonal discrimination*. Guilford Press.

Love, B. J., DeJon, K., Hughbanks, C. (2007). *Critical Liberation Theory.* UMASS, Amherst. Retrieved from https://drive.google.com/file/d/1jX-QJArxryKLCzmiVgnI00JHsG9h7uMP/view

MacLin, M. K. (2020). *Experimental Design in Psychology: A Case Approach*. Routledge.

Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education, 60*(3), 245-256. https://doi.org/10.1177/0004944116664618

Marcucci, O. (2020). Implicit bias in the era of social desirability: Understanding antiblackness in rehabilitative and punitive school discipline. *The Urban Review*, *52*(1), 47-74. https://doi.org/10.1007/s11256-019-00512-7

Martin, C. E. (2015). *Disproportionality of African American students in special education: The influence of aversive racism on referrals*. University of Iowa. https://doil.org/10.17077/etd.uxgi3rde

Martin, D. (2009). The anatomy of white guilt. Retrieved from https://www.racialequitytools.org/resources/fundamentals/core-concepts/system-of-white-supremacy-and-white-privilege

Matias, C. E. (2016a). "Why do you make me hate myself?": Re-teaching Whiteness, abuse, and love in urban teacher education. *Teaching Education*, *27*(2), 194-211. https://doi.org/10.1080/10476210.2015.1068749

Matias, C. E. (2016b). *Feeling white: Whiteness, emotionality, and education*. Rotterdam, The Netherlands: Sense Publishers.

McGlothlin, H., Killen, M., & Edmonds, C. (2005). European-American children's intergroup attitudes about peer relationships. *British Journal of Developmental Psychology*, *23*(2), 227-250. https://doi.org/10.1348/026151005x26101

Mead, H. (2010). *The Effects of Aversive Racism on Teachers' Assessments of Attention Deficit/Hyperactivity Disorder among African American Youth*. [Doctoral dissertation, University of Washington]. ProQuest Dissertations & Theses Global.

Meckler, L., & Natanson, H. (2021). As schools expand racial equity work, conservatives see a new threat in critical race theory. *The Washington Post*, *3*. *https://www.allsides.com/news/2021-05-04-0957/schools-expand-racial-equity-work-conservatives-see-new-threat-critical-race*

Milkman, K. L., Akinola, M., & Chugh, D. (2012). Temporal distance and discrimination: An audit study in academia. *Psychological Science*, 23(7), 710–717. https://doi.org/10.1177/0956797611434539

Milkman, K. L., Akinola, M., & Chugh, D. (2015). What Happens Before? A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway Into Organizations. *Journal of Applied Psychology*, 100(6), 1678-1712. http://dx.doi.org/10.1037/apl0000022

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*, *4*(1), 1-9. https://doi.org/10.1186/2046-4053-4-1

Morris, K. S., Janssen, J. A., & Seaton, E. K. (2023). Level Up: Recommendations for Measuring Racial Discrimination. In *Diversity and Developmental Science: Bridging the Gaps Between Research, Practice, and Policy* (pp. 19-38). Cham: Springer International Publishing.

Morrison, T. G., & Kiss, M. (2017). Modern racism scale. Encyclopedia of personality and individual differences, 1-3. https://doi.org/10.1007/978-3-319-28099-8_1251-1

Moskowitz, R. L., & Druckman, J. N. (2016). *Race, equality, and community in local education: How complex beliefs and values shape educational attitudes, votes, and school policy* [Doctoral dissertation, Northwestern University]. ProQuest Dissertations & Theses Global.

Moss, A. J., & O'Brien, L. T. (2019). *A step before the first step? Social norms and admitting implicit racial prejudice.* [Doctoral dissertation, Tulane University School of Science and Engineering]. ProQuest Dissertations & Theses Global.

Neal, L. V. I., McCray, A. D., Webb-Johnson, G., & Bridgest, S. T. (2003). The effects of African American movement styles on teachers' perceptions and reactions. *The Journal of Special Education*, *37*(1), 49-57. https://doi.org/10.1177/00224669030370010501

Newton, V. A., Dickter, C. L., & Gyurovski, I. I. (2011). The effects of stereotypical cues on the social categorization and judgment of ambiguous-race targets. *Journal of Interpersonal Relations, Intergroup Relations and Identity*, *4*, 31-45.

Nichols, A. H., & Reason, R. D. (2009). *Campus racism and White stereotype threat: Implications for campus racial climates and interracial interactions among students* (Publication Number 3374528) [Doctoral dissertation, The Pennsylvania State University]. ProQuest Dissertations & Theses Global.

No author (2022). *Which States Passed Laws Restricting School Curriculum*? USA Facts. Retrieved from https://usafacts.org/articles/which-states-passed-laws-restricting-school-curriculum/

No Author (2023). Arizona schools chief sets up critical race theory hotline. Associated Press. Retrieved from https://apnews.com/article/arizona-critical-race-theory-hotline-5466b192f8209decb12c3650e749aaff

Noltemeyer, A.L., Mujic, J., & McLoughlin, C.S. (2012). The history of inequality in education. In A.L.Noltemeyer & C.S. McLoughlin (Eds.), Disproportionality in Education and Special Education. Springfield, IL: Charles C. Thomas Publisher Ltd.

Nordin, A., & Brown, K. (2022). Chinese identity and intersectionality in an age of anti-Asian racism. *Journal of the British Association for Chinese Studies*, 124. https://doi.org/10.51661/bjocs.v12i1.174

Oberfield, Z. W., & Incantalupo, M. B. (2021). Racial discrimination and street-level managers: Performance, publicness, and group bias. *Public Administration Review*, *81*(6), 1055-1070. https://doi.org/10.1111/puar.13376

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, *26*(5), 617-624. https://doi.org/10.1177/0956797615570365

Pager, D. (2007). The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *The Annals of the American Academy of Political and Social Science*, *609*(1), 104-133. https://doi.org/10.1177/0002716206294796

Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, *34*, 181-209. https://doi.org/10.1146/annurev.soc.33.040406.131740

Parks, F. R., & Kennedy, J. H. (2007). The impact of race, physical attractiveness, and gender on education majors' and teachers' perceptions of student competence. *Journal of Black Studies*, *37*(6), 936-943. https://psycnet.apa.org/doi/10.1177/0021934705285955

Perez, A. D., & Okonofua, J. A. (2022). The good and bad of a reputation: Race and punishment in K-12 schools. *Journal of Experimental Social Psychology*, *100*, 104287. https://doi.org/10.1016/j.jesp.2022.104287

Pernell, E. (1984). The Influence of Race and Social Behavior on Teacher Recommendation for Special Education Services. Michigan State University

Petts, A. L. (2020, June). It's all in the definition: Color-blind interpretations of school diversity. In *Sociological Forum* (Vol. 35, No. 2, pp. 465-487).

Priest, N., Paradies, Y., Trenerry, B., Truong, M., Karlsen, S., & Kelly, Y. (2013). A systematic review of studies examining the relationship between reported racism and health and wellbeing for children and young people. *Social Science & Medicine*, 95, 115–127. https://doi.org/10.1016/j.socscimed.2012.11.031

Quinn, D. M. (2020a). Experimental effects of "achievement gap" news reporting on viewers' racial stereotypes, inequality explanations, and inequality prioritization. *Educational Researcher*, 49(7), 482-492. https://doi.org/10.3102/0013189X20932469

Quinn, D. M. (2020b). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, 42(3), 375-392. https://doi.org/10.3102/0162373720932188

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. *The handbook of research synthesis and meta-analysis*, *2*, 295-316. https://doi.org/10.7758/9781610448864

Raymond, R. P., & Salvia, J. (1997). *Teacher perceptions of Native American students* (Publication Number 9817562) [Doctoral dissertation, The Pennsylvania State University]. ProQuest Dissertations & Theses global.

Reichardt, C. S. (2019). Quasi-experimentation: A guide to design and analysis. The Guilford Press.

Rhee-Worobec, C. H. (2000). *Dispositional empathy, racial attitudes, and affective responding to racial minorities*. [Doctoral dissertation, University of Massachusetts Boston]. ProQuest Dissertations & Theses global.

Richardson, J. D. (2005). Switching social identities: The influence of editorial framing on reader attitudes toward affirmative action and African Americans. *Communication Research*, 32(4), 503-528. https://psycnet.apa.org/doi/10.1177/0093650205277321

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, 40(3), 417–423. https://doi.org/10.1016/j.jesp.2003.09.002

Robinson, J. D., Dieckmann, N., Withers, E., Hassouneh, D., & Thomas, C. R., Jr. (2018). Does racial bias affect NCI-Funded PIs' willingness to mentor prospective graduate students? *Cancer Research*, 78(17), 4809-4811. https://doi.org/10.1158/0008-5472.can-18-0269

Rollins, D. A. (2005). *Gender and ethnicity referral bias for ADHD: the school's view*. [Doctoral dissertation, Texas A&M University]. ProQuest Dissertations and Theses Global.

Rosenbloom, S. R., & Way, N. (2004). Experiences of discrimination among African American, Asian American, and Latino adolescents in an urban high school. *Youth & Society*, *35*(4), 420-451. https://doi.org/10.1177/0044118X03261479

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638. https://psycnet.apa.org/doi/10.1037/0033-2909.86.3.638

Rosenthal, R., & Rosnow, R. L. (1984). Applying Hamlet's question to the ethical conduct of research: A conceptual addendum. *American Psychologist*, *39*(5), 561. https://psycnet.apa.org/doi/10.1037/0003-066X.39.5.561

Rossi, P. H., & Nock, S. L. (1982). *Measuring social judgments: The factorial survey approach* (pp. 15-67). Beverly Hills, CA: Sage.

Rothenberg, P.S. (Ed.). (1988). *Racism and sexism: An integrated study*. New York: St. Martin's Press.

Ruscher, J. B., Wallace, D. L., Walker, K. M., & Bell, L. H. (2010). Constructive feedback in cross-race interactions. *Group Processes & Intergroup Relations*, 13(5), 603-619. https://doi.org/10.1177/1368430210364629

Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence-combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology*, *97*(1), 43. https://psycnet.apa.org/doi/10.1037/0022-0663.97.1.43

Sadeh, S. S., & Sullivan, A. L. (2017). *School psychologists' decision making in evaluations for emotional disturbance* [Doctoral dissertation, University of Minnesota]. ProQuest Dissertations & Theses Global.

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, *69*, 371-380. https://doi.org/10.1016/j.chb.2016.12.033

Saunders, R. M. (2020). *Investigating the Relationship between Student Race and Academic Expectations Among High School Counselors and Teachers* (Doctoral dissertation, The University of North Carolina at Charlotte). ProQuest Dissertations & Theses Global.

Schwartz, S., & Harris, E. P. (2021). Map: Where critical race theory is under attack. *Education Week.* Retrieved from https://www.edweek.org/policy-politics/map-where-critical-race-theory-is-under-attack/2021/06

Seaton, E. K., & Iida, M. (2019). Racial discrimination and racial identity: Daily moderation among Black youth. *American Psychologist*, 74(1), 117–127. https://doi.org/10.1037/amp0000367

Seaton, E. K., & Yip, T. (2009). School and neighborhood contexts, perceptions of racial discrimination, and psychological well-being among African American adolescents. *Journal of Youth and Adolescence*, 38(2), 153–163. https://doi.org/10.1007/s10964-008-9356-x

Sedlacek, Q. C. (2019). *What Counts as Science Writing and Who Gets to Be" Good" at It?: Pedagogy, Ideology, and Inequality in High School Biology*. [Doctoral dissertation, Stanford University]. ProQuest Dissertations & Theses Global.

Shepherd, M. A. (2020). Effects of race/ethnicity, gender, and intonation on secondary science teachers' evaluation of spoken responses. *Urban Education*, *55*(5), 730-752. https://doi.org/10.1177/0042085916660346

Shivers, E. M., Janssen, J. A., Subramaniam, A., Parker, A. L., Noroña, C. R., Lara, C. … Fitzgibbons, S. (2022). Digging Deeper: De-Colonizing Our Understanding and Practice of Reflective Supervision Through a Racial Equity Lens: First-Wave Findings. Prepared by Indigo Cultural Center for the Alliance for the Advancement of Infant Mental Health. With funding from Perigee Fund. Retreived from https://drive.google.com/file/d/13gXn_7l5qx-th3-TgQqZ7mKzsV2FoiQo/view?usp=share_link

Small, D. A., Pope, D. G., & Norton, M. I. (2012). An age penalty in racial preferences. *Social Psychological and Personality Science*, *3*(6), 730-737. https://doi.org/10.1177/194855061243822

Spanierman, L. B., & Heppner, M. J. (2004). Psychosocial Costs of Racism to Whites scale (PCRW): Construction and initial validation. *Journal of counseling psychology*, *51*(2), 249. https://psycnet.apa.org/doi/10.1037/0022-0167.51.2.249

Spinrad, T. L., Eisenberg, N., Xiao, S. X., Xu, J., Berger, R. H., Pierotti, S. L., ... & Lopez, J. (2023). White children's empathy-related responding and prosocial behavior toward White and Black children. *Child Development*. https://doi.org/10.1111/cdev.13841

Starcke, M., & Porter, S. R. (2019). Do student conduct administrators discriminate against Black students? An analysis of drug sanctions using vignettes. *Review of Higher Education*, 42(2), 765-792. https://doi.org/10.1353/rhe.2019.0014

Steele, S. (1990). *The content of our character.* New York: Harper Peren- nial.

Steinberg, J. A. (1980). *Effects of observed positive, negative and neutral social interactions on perceived liking in children* (Publication Number 8026384)

[Doctoral dissertation, Syracuse University]. ProQuest Dissertations & Theses global.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046-1055. https://doi.org/10.1016/s0895-4356(01)00377-8

Stewart, R., & Uggen, C. (2020). Criminal records and college admissions: A modified experimental audit. *Criminology*, 58(1), 156–188. https://doi.org/10.1111/1745-9125.12229

Sullivan, A. L., Sadeh, S., & Houri, A. K. (2019). Are school psychologists' special education eligibility decisions reliable and unbiased?: A multi-study experimental investigation. *Journal of School Psychology*, 77, 90-109. https://doi.org/10.1016/j.jsp.2019.10.006

Swim, J. K., & Miller, D. L. (1999). White guilt: Its antecedents and consequences for attitudes toward affirmative action. *Personality and Social Psychology Bulletin, 25,* 500–514. https://doi.org/10.1177/0146167299025004008

Takaki, R. (2008). *A different mirror: A history of multicultural America (Revised edition)*. eBookIt. com.

Teeger, C. (2015). Ruptures in the rainbow nation: How desegregated South African schools deal with interpersonal and structural racism. *Sociology of Education*, 88(3), 226-243. https://doi.org/10.1177/0038040715591285

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of educational psychology*, *99*(2), 253. https://psycnet.apa.org/doi/10.1037/0022-0663.99.2.253

Thapar, A., & Rutter, M. (2019). Do natural experiments have an important future in the study of mental disorders? *Psychological Medicine*, *49*(7), 1079-1088. https://doi.org/10.1017/S0033291718003896

Thompson, T. L., & Kiang, L. (2010). The model minority stereotype: Adolescent experiences and links with adjustment. *Asian American Journal of Psychology*, *1*(2), 119. https://doi.org/10.1037/a0019966

Thornhill, T. (2019). We want Black students, just not you: How white admissions counselors screen black prospective students. *Sociology of Race and Ethnicity*, 5(4), 456–470. https://doi.org/10.1177/2332649218792579

Tittler, M. V., & Wade, N. G. (2019). Engaging White participants in racial dialogues: Group composition and dialogue structure. *Group Dynamics: Theory, Research, and Practice*, 23(2), 75-90. https://doi.org/10.1037/gdn0000099

Townsend, B. L. (2000). The disproportionate discipline of African American learners: Reducing school suspensions and expulsions. *Exceptional children*, *66*(3), 381-391. https://doi.org/10.1177/001440290006600308

Travers, J. C., Tincani, M., & Krezmien, M. P. (2013). A multiyear national profile of racial disparity in autism identification. *The Journal of Special Education*, *47*(1), 41-49. https://doi.org/10.1177/0022466911416247

Tuttle, C. N., & Kearney, M. (2020). *Essays on Government Responses to Crime and Racial Inequality* (Publication Number 27995023) [Doctoral dissertation, University of Maryland, College Park]. ProQuest Dissertations & Theses global.

Ura, S. K., & d'Abreu, A. (2022). Racial Bias, Social–Emotional Competence, and Teachers' Evaluation of Student Behavior. *Children & Schools*, *44*(1), 17-26. https://doi.org/10.1093/cs/cdab028

Valant, J., & Newark, D. A. (2016). The politics of achievement gaps: US public opinion on race-based and wealth-based differences in test scores. *Educational Researcher*, *45*(6), 331-346. https://doi.org/10.3102/0013189X16658447

Vanman, E. J., Paul, B. Y., Ito, T. A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology*, 73(5), 941-959. https://psycnet.apa.org/doi/10.1037/0022-3514.73.5.941

Vanman, E. J., Ryan, J. P., Pedersen, W. C., & Ito, T. A. (2013). Probing prejudice with startle eyeblink modification: A marker of attention, emotion, or both? *International Journal of Psychological Research*, 6, 30-41.https://doi.org/10.21500/20112084.717

Wang, W., Spinrad, T. L., Gal-Szabo, D. E., Laible, D., Xiao, S. X., Xu, J., ... & Carlo, G. (2020). The relations of White parents' implicit racial attitudes to their children's differential empathic concern toward White and Black victims. *Journal of Experimental Child Psychology*, *199*, 104928. https://doi.org/10.1016/j.jecp.2020.104928

Wang, Y., & Yip, T. (2020). Sleep facilitates coping: Moderated mediation of daily sleep, ethnic/racial discrimination, stress responses, and adolescent well-being. *Child Development*, 91(4). https://doi.org/10.1111/cdev.13324

Watkins, M and Shulman, H. (2008). Toward Psychologies of Liberation. New York, NY: Palgrave McMillan.

Wegmann, K. M. (2017). “His skin doesn’t match what he wants to do”: Children’s perceptions of stereotype threat. *American Journal of Orthopsychiatry*, 87(6), 615-625. https://psycnet.apa.org/doi/10.1037/ort0000238

Westmoreland, A. S., Mahalingam, R., & Lee, F. (2018). Not Your Model Minority: Workplace Outcomes among Asian Pacific Islander Americans (Publication Number 10902968)

Wolsko, C., Park, B., Judd, C. M., & Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. *Journal of Personality and Social Psychology*, 78(4), 635-654. https://psycnet.apa.org/doi/10.1037/0022-3514.78.4.635

Woods Jr, I. L. (2022). Teachers’ Racial Bias: Can Teachers Rate an African American Boy Similarly to His Racial Counterparts on Problem Behaviors for a Broad Rating Scale?. *School Psychology Review*, 1-15. https://doi.org/10.1080/2372966X.2022.2041212

Xie, S. (2015). *The role of implicit bias in the overrepresentation of African American males within the public school system* [Doctoral dissertation, University of Washington Libraries]. ProQuest Dissertations & Theses Global.

Xie, S., & Cook, C. R. (2016). *The role of implicit bias in the overrepresentation of African American males within the public school system* [Doctoral dissertation, University of Washington]. ProQuest Dissertations & Theses Global.

Yuval-Davis, N. (2012). Dialogical epistemology—an intersectional resistance to the “oppression Olympics.” *Gender & Society*, *26*(1), 46-54. https://doi.org/10.1177/0891243211427701

Zabel, K. L., Olson, M. A., Johnson, C. S., & Phillips, J. E. (2015). What we talk about matters: Content moderates cognitive depletion in interracial interactions. *The Journal of Social Psychology*, 155(6), 545-552. https://doi.org/10.1080/00224545.2015.1032197

Zeiders, K. H. (2017). Discrimination, daily stress, sleep, and Mexican-origin adolescents’ internalizing symptoms. *Cultural Diversity and Ethnic Minority Psychology*, 23(4), 570–575. https://doi.org/10.1037/cdp0000159

Zhu, Y., & Bresnahan, M. (2018). "They Make No Contribution!" versus "We Should Make Friends with Them!" American domestic students' perception of Chinese international students' reticence and face. *Journal of International Students*, 8(4), 1614-1635. https://doi.org/10.32674/jis.v8i4.221

APPENDIX A

INCLUSION AND EXCLUSION CODING SCHEME

Code	Description
1 Not US	Study was not conducted in the US.
2 Not Experimental	Study did not use an experimental method (vignette, audit, association task, field experiment, natural experiment).
3 Intervention	Study purpose was to examine a change in racism (pre/post).
4 Not examining racism	Study did not examine racial discrimination, racial bias, cultural bias, multiculturalism, etc.
5 Racism toward White	Study examined racism toward White participants or fictitious White targets.
6 No White in sample	Study did not include Whites in sample (e.g., BIPOC responses to racism; a sample of ethnic-racial minorities, only).
7 Not Primary ED	Study was missing an education focus for the manipulation (e.g., examining students/other adults with no connection to pre-K-12th grade).

APPENDIX B

META-ANALYTIC CODEBOOK

Code	Description
Report Number (ID_A)	Record the report number.
Report First Author's Last Name (A_NAME)	Record the first author's last name.
Year of Report (YEAR)	Record the year of the publication or presentation.
Journal Title (JTITLE)	Record the journal title if published. Record the institution if dissertation or thesis. Record the conference name if a presentation.
Publication Status (PUB_STS)	1 = Publication, peer reviewed 2 = Dissertation 3 = Thesis 4 = Conference 5 = Other unpublished report
Sample Size (N)	Record the sample size.
Experiment Location (LOCAT)	Record the city, state, region of the experimental location(s). Separate multiple locations with a colon.
Region (REGION)	Northeast - Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania Midwest - Ohio, Michigan, Indiana, Wisconsin, Illinois, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, and Kansas South - Delaware, Maryland, Virginia, West Virginia, Kentucky, North Carolina, South Carolina, Tennessee, Georgia, Florida, Alabama, Mississippi, Arkansas, Louisiana, Texas, and Oklahoma West - Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, California, Oregon, Washington, Alaska, and Hawaii
Education level Examined (ED_LVL)	1 = Pre K 2 = Elementary (e.g., 3 = Middle/Jr. High (e.g., 4 = High School 6 = K=12 school in general 7 = Pre-service educator
Participant Age (AGE)	mean, sd
Educational Domain	Open ended

Participant Gender (pGEND)	1 = Female 2 = Male 3 = Non-binary 4 = Other 5 = Unspecified
Participant Gender Other (pGEND_O)	Specify other gender included
Participant SES (pSES)	Note any characteristics about participant SES.
Participant Races Included (pRACE)	Check all that apply 1 = Arab 2 = Asian 3 = Native American 4 = Black 5 = Latina 6 = Hawaiian 7 = Pacific Islander 8 = White 9 = Multiracial 10 = Aggregated Minority Group 11 = Other Race 12 = No report
Participant Race Other (pRACE_O)	Specify
Participant Race Multiracial (pRACE_M)	Specify
Participant Assignment (ASSIGN)	1 = Random 2 = Matched 3 = Other (specify) 4 = Within-subjects
Participant Assignment Other (ASSIGN _O)	Specify
Experimental Method (EXPM)	1 = Audit/Correspondence Study 2 = Natural Experiment 3 = Vignette Experiment 4 = Task 5 = Confederate 6 = Other
Experimental Method (EXPM_O)	Specify
Lab (LAB)	1 = Yes 2 = No

Participant Perpetrator Type (PERP)	1 = Peer 2 = Teacher/Professor 3 = Guidance/Academic Counselor 4 = Principal 5 = District Office 6 = Admissions 7 = Hiring of school positions 8 = Student/pupil 9 = Policy/curriculum 10 =Parent 11 = Other (specify)
Racism Perpetration Level (RACELEVEL)	1 = Interpersonal 2 = Power of the institution 3 = Institutional
Participant Perpetrator Examined Other (PERP_O)	Specify
Target Names (T_NAME)	Record the names and gender of the fictitious auditors used in the study. Separate names with a comma.
Target Name Citation (T_NAME_C)	Record the citation or source that justifies the selection of fictitious names
Treatment Pilot (T_PILOT)	Record information about a pilot
Treatment Validation (T_VALIDATE)	Record information about the validity of the manipulation
Target Races Included (T_RACE)	Check all that apply 1 = Arab 2 = Asian 3 = Native American 4 = Black 5 = Latina 6 = Hawaiian 7 = Pacific Islander 8 = White 9 = Multiracial 10 = Aggregated Minority Group 11 = Other Race 12 = No report
Target Race Other (T_RACE_O)	Specify
Target Race Multiracial (T_RACE_M)	Specify

Target Type (T_TYPE)	1 = Peer 2 = Teacher/Professor 3 = Guidance/Academic Counselor 4 = Principal 5 = District Office Employee 6 = Admissions 7 = Hiring of Personnel 8 = Student/pupil 9 = Legislator/Policy 10 =Parent 11 = Other (specify)
Target Type Other (T_TYPE_O)	Specify
Effect for Target(s) (EF_PEARSON.R)	R = N = 95% CI [,] FISHCER'S ZR = OTHER =
Effect for Target(s) (EF_COHEN.HEDGES)	MEAN X1 = SD X1 = N X1 = MEAN X2 = SD X2 = N X2 = D = 95% C.I. [,] OTHER =
Effect for Target(s) (EF_ODDS)	OUTCOME X1 = NO OUTCOME X1 = N X1 = OUTCOME X2 = NO OUTCOME X2 = N X2 = OR = 95% C.I.[X,X] ORLOGGED = VORLOGGED = OTHER =
Final Notes (NOTES)	Open ended

APPENDIX C

EFFECT SIZE COMPUTATION FOR EACH RECORD

This table describes the process by which each effect size included in the present meta-analysis was calculated. All effect sizes were calculated using the Practical Meta-Analysis Effect Size Calculator by David B. Wilson, Ph.D., George Mason University (2001).

This calculator is a companion to the 2001 book by Mark W. Lipsey and David B. Wilson, Practical Meta-analysis, published by Sage and is freely available at - https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php.

Authors	Year	Notes on Effect Size Calculation	Notes on Effect Size Direction	Page Number
Armstrong1	2021	Giftedness Rating (WvB): Effect size based on group n's (listed), M's, ad SD's.	Outcome rating is higher values indicates more giftedness. Whites used as treatment group.	p.66
Armstrong2	2021	Referral Likelihood(WvB): Effect size based on group n's (listed), M's, ad SD's.	Outcome rating is higher values indicate more likelihood to refer to giftedness. Whites used as treatment group.	p.69
Barry1	2020	Situation Severity Rating (WvB): Effect size based on group n's (listed), M's, and SD's.	Outcome rating is higher value indicates perception that behavior is more severe. Black used as treatment group	p.41
Barry2	2020	Likelihood to Seek Assistance (WvB): Effect size based on group n's (listed), M's, ad SD's.	Outcome rating is higher value indicates more likely to seek assistance. Black used as treatment group.	p.41
Sedlacek1	2019	Overall writing score (out of 10; WvL): Effect size based on group n's (listed), M's, and SD's.	Outcome rating is higher values indicates better writing score. Whites used as treatment group.	p.125
Sedlacek2	2019	Feedback length total sentences (WvL): Effect size based on M's, group n's (listed), and t-value statistic	Outcome rating is higher values indicates more quality feedback. White used as treatment group.	p.127

Harris1	2013	Discipline decision (WvB): Effect size based on group n's (listed), and t-value for White versus Black; correlation (r) could also be used.	Outcome rating is discipline. Higher value indicates more discipline. Black used as treatment group.	p.56
Copur-Gencturk1-6	2022	Correctness rating (WvB). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SE were transformed into SD by hand. Effect sizes compare White boy v black boy; White girl v black girl on the incorrect, partially, and fully correction solution types.	Outcome rating is higher values is more correctness. White is used as the treatment group.	p.8
Copur-Gencturk7-12	2022	Estimate of mathematical ability (WvB): Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SE were transformed into SD by hand. Effect sizes compare White boy v black boy; White girl v black girl on the incorrect, partially correct, and fully correction solution types.	Outcome rating is higher values is more math ability. White is used as the treatment group.	p.9
Copur-Gencturk13-18	2022	Recommendation for giftedness (WvB): Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SE were transformed into SD by hand. Effect sizes compare White boy v black boy; White girl v black girl on the incorrect, partially correct, and fully correction solution types.	Outcome rating is higher values is more giftedness. White is used as the treatment group.	p.10

Copur-Gencturk19-24	2022	Recommendation for IEP (WvB): Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SE were transformed into SD by hand. Effect sizes compare White boy v black boy; White girl v black girl on the incorrect, partially correct, and fully correction solution types.	Outcome rating is higher likeliness to recommend to IEP. Black is used as the treatment group given black is disproportionately represented in special education.	p.10
Dunbar1	2022	Investment in Security (WvB). Effect size based on M's, group n's (not listed, sample size was divided into equal groups per treatment), and SD.	Outcome rating is higher investment of security (punitive). Black is used as the treatment group.	T2
Dunbar2	2022	Investment in Suspensions (WvB). Effect size based on M's, group n's (not listed, sample size was divided into equal groups per treatment), and SD.	Outcome rating is higher suspensions of suspensions (punitive). Black is used as the treatment group.	T2
Dunbar3	2022	Investment in After school programming (WvB). Effect size based on M's, group n's (not listed, sample size was divided into equal groups per treatment), and SD.	Outcome rating is higher investment in afterschool programs (supportive). White is used as the treatment group.	T2
Dunbar4	2022	Investment in mental health services (WvB). Effect size based on M's, group n's (not listed, sample size was divided into equal groups per treatment), and SD.	Outcome rating is higher investment in mental health (supportive). White is used as the treatment group.	T2
Dunbar5	2022	Contribution of own money (WvB). Effect size based on M's, group n's (not listed, sample size was divided into equal groups per treatment), and SD.	Outcome rating is higher contribution of personal money. White is used as the treatment group.	T2

Hailey1	2022	Perceptions of welcomeness (WvL school). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SE. *White participants only were included in effect size calculation given child toward peer (e.g., no power of institution to exert)	Outcome is increased feeling of welcomeness. White is treatment group.	p.894
Hailey2	2022	Perceptions of welcomeness (WvB school). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SE . *White participants only were included in effect size calculation given child toward peer (e.g., no power of institution to excert)	Outcome is increased feeling of welcomeness. White is treatment group.	p.894
Hailey3	2022	Perceptions of welcomeness (WvM school). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SE. *White participants only were included in effect size calculation given child toward peer (e.g., no power of institution to excert)	Outcome is increased feeling of welcomeness. White is treatment group.	p.894
Hoffman1	2014	*Article specifies that sample sizes for native and Asian who were expelled is too small to include. Proportion change in expulsion recommendations (WvB). Effect size based on proportion change from 2008-2005 calculated for Black and for White, N's is total number of secondary students in 2008+2005 (listed)	Outcome is expulsion recommendation from not likely to likely. Black is treatment group.	p.80

Hoffman2	2014	Proportion change in expulsion recommendations (WvL). Effect size based on proportion change from 2008-2005 calculated for Black and for White, N's is total number of secondary students in 2008+2005 (listed)	Outcome is expulsion recommendation from not likely to likely. Latine is treatment group.	p.80
Janssen1	2022	Response rate (WvB). Effect size based on outcome frequency (response, no response was obtained from author for each racial group), Cell n's calculated as the number of vignette manipulations delivered.	Outcome is responsiveness from not likely to likely. White is treatment group.	Author's data
Janssen2	2022	Response rate (WvA). Effect size based on outcome frequency (response, no response was obtained from author for each racial group), Cell n's calculated as the number of vignette manipulations delivered.	Outcome is responsiveness from not likely to likely. White is treatment group.	Author's data
Janssen3	2022	Response rate (WvL). Effect size based on outcome frequency (response, no response was obtained from author for each racial group), Cell n's calculated as the number of vignette manipulations delivered.	Outcome is responsiveness from not likely to likely. White is treatment group.	Author's data
Jarvis1	2020	Discipline severity (WvB). Effect size based on F-test, Group n's (not listed, sample size was divided into equal groups per treatment).	Outcome is discipline severity from low to high. Black is treatment group.	p.494
Jarvis2	2020	Feeling Troubled (WvB). Effect size based on F-test, Group n's (not listed, sample size was divided into equal groups per treatment).	Outcome is feeling troubled from low to high. Black is treatment group.	p.495

Jarvis 3	2020	Rating of troublemakers (BvW). Effect size based on group n's (not listed, sample size was divided into equal groups per treatment), M's, and SD's.	Outcome is target as troublemaker from low to high. Black is treatment group.	p.495
Jarvis 4	2020	Rating of behavior as a pattern (BvW). Effect size based on group n's (not listed, sample size was divided into equal groups per treatment), M's, and SD's.	Outcome is target behavior is a pattern from not likley to likley. Black is treatment group.	p.495
Jarvis5	2020	Rating of days put in detention (BvW). Effect size based on group n's (not listed, sample size was divided into equal groups per treatment), M's, and SD's.	Outcome is number of days recommended for detention. Black is treatment group.	p.495
Jarvis6	2020	Rating of recommend suspension (BvW). Effect size based on group n's (not listed, sample size was divided into equal groups per treatment), M's, and SD's.	Outcome is recommended for suspension from low to high. Black is treatment group.	p.495
Lorenzetti 1-4	2021	Behaviors due to choices (Causality; BvW). Effect sizes based on M's, SD's, and cell size (total vignettes divided by treatment). Calculated for ADHD symptoms, internalizing sympotms, disruptive behaviors, disorganized behaviors.	Outcome is rating the symptoms as due to the student's personal choices. Black is treatment group.	p.49
Lorenzetti 5-8	2021	Behavior outside of control ; BvW). Effect sizes based on M's, SD's, and cell size (total vignettes divided by treatment). Calculated for ADHD symptoms, internalizing sympotms, disruptive behaviors, disorganized behaviors.	Outcome is rating the symptoms are due to things outside of student's control. White is treatment group.	p.49

Martin1	2008	White child selected neighbor(BvW). Gender was collapsed so that effect size is calculated for Black versus White target only. Effect size based on calculated mean coded as 0=frown, 1=neutral, 2=smile, Cell n's calculated as the number of vignette manipulations delivered, and SD (calculated by hand).	Outcome is assignment happiness rating to have a neighbor. White is treatment group.	p.73
Mead1-3	2011	Hyperactivity (BvW). Effect size based on M's, Cell size calculated as the number of vignette manipulations delivered and SD's. Effect sizes are calculated for White low, med, high.	Outcome is higher levels of hyperactivity. Black is treatment group.	p.93
Mead4-6	2011	Inattention (BvW). Effect size based on M's, Cell size calculated as the number of vignette manipulations delivered and SD's. Effect sizes are calculated for White low, med, high.	Outcome is higher levels of inattention. Black is treatment group.	p.93
Mead7-9	2011	Aggression (BvW). Effect size based on M's, Cell size calculated as the number of vignette manipulations delivered and SD's. Effect sizes are calculated for White low, med, high.	Outcome is higher levels of aggression. Black is treatment group.	p.93
Mead10-12	2011	Positive Impression (BvW). Effect size based on M's, Cell size calculated as the number of vignette manipulations delivered and SD's. Effect sizes are calculated for White low, med, high.	Outcome is higher levels of positive impression. White is treatment group.	p.93
Mead13-15	2011	Negative Impression (BvW). Effect size based on M's, Cell size calculated as the number of vignette manipulations delivered and SD's. Effect sizes are calculated for White low, med, high.	Outcome is higher levels of negative impression. Black is treatment group.	p.93

Mead16-18	2011	Referral (BvW). Effect size based on M's, Cell size calculated as the number of vignette manipulations delivered and SD's. Effect sizes are calculated for White low, med, high.	Outcome is higher likelihood of referral. Black is treatment group.	p.93
Copur-Gencturk1-3	2020	Correctness rating (BvW). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SD. Effect size for Black versus White comparison for boys on incorrect, partially correct, and fully correct solutions.	Outcome is level of correctness. White is the treatment group.	p.43
Copur-Gencturk4-6	2020	Correctness rating (BvW). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SD. Effect size for Black versus White comparison for girls on incorrect, partially correct, and fully correct solutions.	Outcome is level of correctness. White is the treatment group.	p.43
Copur-Gencturk7-9	2020	Correctness rating (LvW). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SD. Effect size for Latine versus White comparison for boys on incorrect, partially correct, and fully correct solutions.	Outcome is level of correctness. White is the treatment group.	p.43
Copur-Gencturk10-12	2020	Correctness rating (LvW). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SD. Effect size for Latine versus White comparison for girls on incorrect, partially correct, and fully correct solutions.	Outcome is level of correctness. White is the treatment group.	p.43

Copur-Gencturk13-15	2020	Math ability rating (BvW). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SD. Effect size for Black versus White comparison for boys on incorrect, partially correct, and fully correct solutions.	Outcome is level of math ability. White is the treatment group.	p.43
Copur-Gencturk16-18	2020	Math ability rating (BvW). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SD. Effect size for Black versus White comparison for girls on incorrect, partially correct, and fully correct solutions.	Outcome is level of math ability. White is the treatment group.	p.43
Copur-Gencturk19-21	2020	Math ability rating (LvW). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SD. Effect size for Latine versus White comparison for boys on incorrect, partially correct, and fully correct solutions.	Outcome is level of math ability. White is the treatment group.	p.43
Copur-Gencturk22-24	2020	Math ability rating (LvW). Effect size based on M's, Cell n's calculated as the number of vignette manipulations delivered, and SD. Effect size for Latine versus White comparison for girls on incorrect, partially correct, and fully correct solutions.	Outcome is level of math ability. White is the treatment group.	p.43
Study One Sullivan1-3	2019	Rating qualify for emotional disability (BvW). Effect size based on 2x2 frequency table. N (listed, divided equally for black and White). Effect size calculated for comparison on ineligible, not ambiguous, and ambiguous vignette. *Confidence about diagnoses was not coded.	Outcome is likelihood to be diagnosed with an emotional disorder. Black is treatment group because overrepresented in emotionality.	p.97

Study two Sullivan1-3	2019	Rating qualify for intellectual disability (BvW). Effect size based on 2x2 frequency table. N (listed, divided equally for black and White). Effect size calculated for comparison on ineligible, not ambiguous, and ambiguous vignette. *Frequency is between intellectual disability and none. *Confidence about diagnoses was not coded.	Outcome is likelihood to be diagnosed with an educational disability. Black is treatment group because overrepresented in special education.	p.97
Study three Sullivan1-3	2019	Rating qualify for autism (BvW). Effect size based on 2x2 frequency table. N (listed, divided equally for black and White). Effect size calculated for comparison on 2, 3, and 3 + ID symptoms. *Frequency is between autism and none. *Confidence about diagnoses was not coded.	Outcome is likelihood to qualify for autism. White is treatment group given black are underdiagnosed with autism.	p.97
Study three Sullivan4-6	2019	Refer for evaluation (BvW). Effect size based on 2x2 frequency table. N (listed, divided equally for black and White). Effect size calculated for comparison on 2, 3, and 3 + ID symptoms.	Outcome is likelihood to refer for evaluation for autism. White is treatment group given black are underdiagnosed with autism.	p.97
Xie1	2015	Total punitive discipline (BvW): Effect size based on group n's (not listed, sample size was divided into equal groups per treatment), M's, ad SD's.	Outcome is more severe discipline. Black is treatment group.	p.88
Xie2	2015	Total Special ed referral (BvW): Effect size based on group n's (not listed, sample size was divided into equal groups per treatment), M's, ad SD's.	Outcome is more severe referral. Black is treatment group given overrepresented in special ed.	p.88

Steinberg	1980	White child peer liking (BvW). Utilized presented frequencies to calculate mean scores and standard deviations for White children's liking to black versus White peer (collapsed behavior categories of Positive, negative, neutral). Cell n's calculated as the number of vignette manipulations delivered, and SD.	Outcome is disliking of peer. Black is used as treatment group.	p.33
Small1	2012	*study one is not experimental and is not coded. Charitable behavior to classroom (BvW). Cell n's calculated as the number of vignette manipulations delivered, and SD. T-value.	Outcome is increased donations. White is reference group.	p.6
Small2	2012	Stereotype (BvW). Cell n's calculated as the number of vignette manipulations delivered, and SD. T-value.	Outcome is stereotype. Black is reference group.	p.6
Sedlacek1	2021	Rating of score (LvW). Effect size based on group n's (listed), M's, ad SD's.	Outcome is rating score from low to high. White used as treatment.	p.2359
Sedlacek2	2021	Clarity score (LvW) Effect size based on group n's (listed), M's, ad SD's.	Outcome is clarity score from low to high. White used as treatment.	p.2359
Sedlacek3	2021	Detail score (LvW) Effect size based on group n's (listed), M's, ad SD's.	Outcome is detail score from low to high. White used as treatment.	p.2359
Sedlacek4	2021	Accuracy (LvW) Effect size based on group n's (listed), M's, ad SD's.	Outcome is accuracy score from low to high. White used as treatment.	p.2359
Sedlacek5	2021	Feedback in sentences (LvW) Effect size based on group n's (listed), M's, ad SD's.	Outcome is number of sentences in increasing order. White is used as treatment.	p.2361

Saunders1	2021	Academic expectation (BvW). Effect size based on group n's (listed), M's, ad SD's.	Outcome is increased academic expectations. White used as treatment.	p.64
Saunders2	2021	Academic expectation (LvW). Effect size based on group n's (listed), M's, ad SD's.	Outcome is increased academic expectations. White used as treatment.	p.64
Perez1	2022	Feeling troubled (BvW) Effect size based on N (listed), M's, and SD's. Statistics were calculated from available data on OSF. Categories of good versus bad reputation and offence 1 versus 2 were combined.	Outcome is Feeling troubled from low to high. Black is treatment group.	OSF dat
Perez2	2022	Discipline Severity (BvW) Effect size based on N (listed), M's, and SD's. Statistics were calculated from available data on OSF. Categories of good versus bad reputation and offence 1 versus 2 were combined.	Outcome is Discipline severity from low to high. Black is treatment group.	OSF dat
Perez3	2022	Detention. (BvW) Effect size based on N (listed), M's, and SD's. Statistics were calculated from available data on OSF. Categories of good versus bad reputation were combined	Outcome is Detention likelihood from low to high. Black is treatment group	OSF dat
Perez4	2022	Suspension(BvW) Effect size based on N (listed), M's, and SD's. Statistics were calculated from available data on OSF. Categories of good versus bad reputation were combined	Outcome is Suspension likelihood from. low to high. Black is treatment group	OSF dat
Perez5	2022	Troublemaker(BvW) Effect size based on N (listed), M's, and SD's. Statistics were calculated from available data on OSF. Categories of good versus bad reputation were combined	Outcome is Troublemaker from low to high. Black is treatment group.	OSF dat

Owens1	2022	Effect size could not be calculated. No statistics were provided.		Emailed
Okonofua1	2015	Feeling troubled First infraction (BvW). T-test statistic, N's (listed via supplemental t-tests).	Outcome is Feeling troubled from low to high. Black is treatment group.	p.619
Okonofua2	2015	Feeling troubled second infraction (BvW). Effect size based on N (listed via supplemental t-tests), M's, and SD's.	Outcome is Feeling troubled from low to high. Black is treatment group.	p.619
Okonofua3	2015	Discipline rating First infraction (BvW). T-test statistic, N's (listed via supplemental t-tests).	Outcome is Discipline severity from low to high. Black is treatment group.	p.619
Okonofua4	2015	Discipline rating second infraction (BvW). Effect size based on N (listed via supplemental t-tests), M's, and SD's.	Outcome is Discipline severity from low to high. Black is treatment group.	p.619
Okonofua5	2015	Troublemaker rating (BvW). Effect size based on N (listed via supplemental t-tests), M's, and SD's.	Outcome is Troublemaker from low to high. Black is treatment group.	p.620
Study 2 Okonofua 1-2	2015	Feeling troubled (BvW). Effect size based on N (listed via supplemental t-tests), M's, and SD's. Effect size calculated for first infraction and second infraction	Outcome is Feeling troubled from low to high. Black is treatment group.	p.621
Study 2 Okonofua 3-4	2015	Discipline rating (BvW). Effect size based on N (listed via supplemental t-tests), M's, and SD's. Effect size calculated for first infraction and second infraction	Outcome is Discipline severity from low to high. Black is treatment group.	p.621
Study 2 Okonofua 5	2015	Troublemaker rating (BvW) Effect size based on N (listed via supplemental t-tests), M's, and SD's.	Outcome is Troublemaker from low to high. Black is treatment group.	p.621

Study 2 Okonofua 6	2015	Pattern behavior (BvW). Effect size based on N (listed via supplemental t-tests), M's, and SD's.	Outcome is behavior is a pattern from not likely to likely. Black is treatment group.	p.621
Study 2 Okonofua7	2015	Suspension likelihood (BvW). Effect size based on N (listed via supplemental t-tests), M's, and SD's.	Outcome is Suspension likelihood from. low to high. Black is treatment group	p.622
Cox1	1996	Peer liking. Effect size based on F statistics. No aggregate M's, SD's presented	Outcome is dislike. Black is Treatment group.	p.50f
Quinn1	2020	Grade rating (WvB). Effect size based on N (listed), M's, SD.	Outcome is Grade bad to good. White is treatment	p.381
Quinn2	2020	Grade rubric (WvB). Effect size based on N(listed). M's, SD.	Outcome is rubric event from bad to good. White is treatment	p.381
Rennels1	2015	* positive flexibility reciprocation (PFR) not coded *Coded White children only. Forced Choice. positive bias reciprocation. Effect size based on t-test statistic, reported effect size, and N (reported), Mean differences were also provided.	Outcome is reciprocation from low to high. White is treatment	T2
Rennels2	2015	*Coded White children only. Non-forced chice. positive bias reciprocation. Effect size based on t-test statistic, reported effect size, and N (reported), Mean differences were also provided.	Outcome is reciprocation from low to high. White is treatment	T2

Valant1	2016	*Coded White parents only *Wealth gap used as control group Prioritization to close the gap (control v Black).. Effect size not calculated. M's, N (listed in appendix), SD obtained from the author.	Outcome is likelihood to support closing the gap. Wealth gap is treatment group.	p.326
Valant2	2016	*Coded White parents only *Wealth gap used as control group Prioritization to close the gap (control v Latine).. Effect size not calculated. M's, N (listed in appendix), SD obtained from the author.	Outcome is likelihood to support closing the gap. Wealth gap is treatment group.	p.326
Wang1	2020	Sympathy(WvB). Effect size based on N (Listed), M's, SD's.	Outcome is sympathy from low to high. White is treatment	T1
Wang2	2020	Distress (WvB) Effect size based on N (Listed), M's, SD's.	Outcome is personal distress from low to high. White is treatment because White children expected to be more distressed when White child is being bullied.	T1
Francis 1-4	2020	Preparation Score (WvB). Effect size based on M, SD's, N's (not listed, evenly divided by race and academic profile of vignette). Effect size calculated for White boy versus Black boy for S ac/S b, B ac/S b, S a/B b, B a/B b.	Outcome is preparation from low to high. White is treatment	FA7-14

Francis 5-8	2020	Preparation Score (WvB). Effect size based on M, SD's, N's (not listed, evenly divided by race and academic profile of vignette). Effect size calculated for White girl versus Black girl for S ac/S b, B ac/S b, S a/B b, B a/B b.	Outcome is preparation from low to high. White is treatment	FA7-14
Francis 9-12	2020	Advance placement recommendation (WvB). Effect size based 2X2 frequency table. N's (not listed, evenly divided by race and academic profile of vignette). Effect size calculated for White boy versus Black boy for S ac/S b, B ac/S b, S a/B b, B a/B b.	Outcome is Ap rec from low to high. White is treatment	FA7-14
Francis 13-16	2020	Advance placement recommendation (WvB). Effect size based 2X2 frequency table. N's (not listed, evenly divided by race and academic profile of vignette). Effect size calculated for White girl versus Black girl for S ac/S b, B ac/S b, S a/B b, B a/B b.	Outcome is Ap rec from low to high. White is treatment	FA7-14
Levy1	2004	BASC part A (WvB). Effect size based on total column, N(listed), M's, SDs.	Outcome is ADHD behaviors. White is treatment because more likely to be diagnosed.	p.63
Levy2	2004	Referral (WvB). Effect size based on total column, N(listed), M's, SDs.	Outcome is service referral for ADHD. White is treatment group	p.73
Levy3	2004	Diagnosis of ADHD (WvB). Effect size based on total column, N(listed), M's, SDs. *Confidence not coded.	Outcome is ADHD diagnosis when ADHD is present. White is treatment group.	p.75

Oberfield1	2021	Responsiveness (WvB). Effect size based on outcome frequency yes, outcome frequency no, total N (not listed, overall sample divided equally across treatments).	Outcome is Response rate from no to response. White is treatment group.	p.1061
Boyd-Swan1	2017	Responsiveness (WvB). Effect size based on outcome frequency yes, outcome frequency no, Cell size calculated as the number of emails across manipulated races.	Outcome is Response rate from no to response. White is treatment group.	p.45
Boyd-Swan1	2017	Responsiveness (WvL). Effect size based on outcome frequency yes, outcome frequency no, Cell size calculated as the number of emails across manipulated races.	Outcome is Response rate from no to response. White is treatment group.	p.45
Briscoe-Juin1	2020	Inappropriateness of behavior (BvW). Effect size based off frequency table and hand calculated M's, SD's, N's (listed)	Outcome is Behavior from appropriate to inappropriate. Black treatment group	p.17
Briscoe-Juin2	2020	Referral to office unlikelihood (BvW). Effect size based off frequency table and hand calculated M's, SD's, N's (listed).	Outcome is Referral from likely to unlikely. White treatment group	p.17
Briscoe-Juin3	2020	Referral to counselor unlikelihood (BvW). Effect size based off frequency table and hand calculated M's, SD's, N's (listed).	Outcome is Referral from likely to unlikely. White treatment group	p.17
Briscoe-Juin4	2020	Consequence severity (BvW). Effect size based off frequency table and hand calculated M's, SD's, N's (listed)	Outcome is Consequence from less to more severe. Black treatment group	p.17
Briscoe-Juin5	2020	Agency to impact behavior (BvW). Effect size based off frequency table and hand calculated M's, SD's, N's (listed)	Outcome is Agency from agree to disagree. Black treatment group	p.17
Briscoe-Juin6	2020	Behavior pattern unlikelihood (BvW). Effect size based off frequency table and hand calculated M's, SD's, N's (listed)	Outcome is Suspected behavior pattern from likely to unlikely. White treatment group	p.17

Briscoe-Juin7	2020	Likelihood enter criminal justice (BvW). Effect size based off frequency table and hand calculated M's, SD's, N's (listed)	Outcome is Perception to enter the criminal justice system from likely to unlikely. White treatment group	p.17
Briscoe-Juin8	2020	Likelihood monitor in the future(BvW). Effect size based off frequency table and hand calculated M's, SD's, N's (listed)	Outcome is monitoring from closely to not closely. White treatment group	p.17
Briscoe-Juin9	2020	Risk of school violence (BvW). Effect size based off frequency table and hand calculated M's, SD's, N's (listed)	Outcome is risk for violence from high to low. White treatment group	p.17
Study 1 Griffiths1		Writing score. (WvB) Effect sized based on N's. (listed), t-test values.	Outcome is Rubric score from low to high. White is reference group	p.28
Study 1 Griffiths2		Writing quality. (WvB) Effect sized based on N's. (listed), t-test values.	Outcome is Quality from low to high. White is reference group.	p.28
Study 1 Griffiths3		Student potential for second draft. (WvB) Effect sized based on N's. (listed), t-test values.	Outcome is Potential from low to high. White is reference group.	p.28
Study 1 Griffiths4		Student potential (WvB) Effect sized based on N's. (listed), t-test values.	Outcome is Potential from low to high. White is reference group.	p.28
Study 1 Griffiths5		Grade projections. (WvB) Effect sized based on N's. (listed), F-test values.	Outcome is Grade projections from low to high. White is reference group.	p.28
Study 1 Griffiths6		Student effort (WvB) Effect sized based on N's. (listed), t-test values.	Outcome is Effort from low to high. White is reference group.	p.30
Study 2 Griffiths1		Writing score (WvB) Effect sized based on N's. (listed), t-test values.	Outcome is Rubric from low to high. White is reference group	p.42

Study 2 Griffiths2		Essay quality. (WvB) Effect sized based on N's. (listed), t-test values.	Outcome is Quality from low to high. White is reference group.	p.43
Study 2 Griffiths3		Student potential for second draft (WvB) Effect sized based on N's. (listed), t-test values.	Outcome is Potential from low to high. White is reference group.	p.43
Study 2 Griffiths4		Grade predicition (WvB) Effect sized based on N's. (listed), F-test values.	Outcome is Grade projections from low to high. White is reference group.	p.43
Study 2 Griffiths5		Student effort (WvB) Effect sized based on N's. (listed), F-test values.	Outcome is Effort from low to high. White is reference group.	p.46
Study 2 Griffiths6		Student attributions(WvB) Effect sized based on N's. (listed), F-test values.	Outcome is Student attributions	p.46
Jarvis1	2021	Discipline severity T1. (WvB) Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Discipline severity from low to high. Black treatment group	p.12
Jarvis2	2021	Discipline severity T2 (WvB) Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Discipline severity from low to high. Black treatment group	p.12
Jarvis3	2021	Troublemaker rating (BvW). Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Troublemaker rating from low to high. Black treatment group	p.12
Jarvis4	2021	Detention Days. (WvB). Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Days in detention from low to high. Black treatment group.	p.12
Jarvis5	2021	Feeling troubled (WvB). F-test, N (divided evenly among vignettes).	Outcome is Feeling troubled from low to high. Black treatment group.	p.12

Jarvis6	2021	Behavior as a pattern (WvB). Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Behavior is a pattern from low to high. Black treatment group.	p.13
Jarvis7	2021	Misbehavior spread (WvB). Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Behavior as a pattern for other students from low to high. Black treatment group.	p.13
Jarvis8	2021	Right course. (BvW). Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Difficulty rating of putting child on the right course from low to high. Black treatment group.	p.13
Jarvis9	2021	Future suspension. (BvW). Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Likelihood to suspend in the future from low to high. Black treatment group.	p.13
Jarvis10	2021	Parental involvement (BvW). Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Likelihood of parent being involved from low to high. White is treatment group.	p.13
Jarvis11	2021	Extracurricular benefit (BvW). Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Child would benefit from extracurricular activities from low to high. White is treatment	p.13
Jarvis12	2021	Benefit from Counselor (WvB). Effect size based on M's, SD, and N (listed in t-test and divided evenly among vignettes).	Outcome is Child would benefit from talking to counselor. White is treatment.	p.13
Dameron1	2018	Likelihood to place in alternative learning program (BvW). Effect size calculated by M, SE's, and N (not listed, divided the sample evenly by vignette)	Outcome is Likelihood to place in ALP. Black as treatment group.	p.83
Gilliam1		*Gazing at children.* Black children, $F(1, 3405)=9.64$, $p=.002$, $d=.57$,	Increases in gaze means more suspicion. Black used as treatment group	p.7

Gilliam2		Child that requires the most attention (BvW). Effect size based on the proportion of selection, N's (listed)	Outcome is Increased attention needed. Black used as treatment group	p.7
Gilliam3		*Behavior severity* participants rated White children's behavior as more severe than Black children's, $F(1, 124)=3.39$, $p=.068$, $\eta^2=.03$, $d=.33$.	Outcome is Increased behavior severity rating. Black used as treatment group	p.9
Elhoweris1	2005	Referral to gifted. (BvW) Effect size based on M's, SD's and N (listed).	Outcome is Referral rated from no to yes. White as treatment group	p.28
Elhoweris1	2005	Personal placement in gifted. (BvW) Effect size based on M's, SD's and N (listed).	Outcome is Personal placement from no to yes. White as treatment group	p.28
Shepherd	2016	Question response grading (BvW) Effect size based on M's, SD's, and Cell n's calculated as the number of vignette manipulations delivered.	Outcome is Response graded from not well to very well. White is treatment group.	p.741
DeMeis1	1978	Personality (BvW) Cell n's calculated as the number of vignette manipulations delivered, F statistic from univariate analysis of variance.	Outcome is Personality rated from low to high. White treatment group.	p.81
DeMeis2	1978	Response. (BvW) Cell n's calculated as the number of vignette manipulations delivered, F statistic from univariate analysis of variance.	Outcome is Response rated from low to high. White treatment group.	p.81
DeMeis3	1978	Future academic ability. (BvW) Cell n's calculated as the number of vignette manipulations delivered, F statistic from univariate analysis of variance.	Outcome is Current academic abilities rated from low to high. White treatment group.	p.81

DeMeis4	1978	Current academic ability. (BvW) Cell n's calculated as the number of vignette manipulations delivered, F statistic from univariate analysis of variance.	Outcome is Future academic abilities rated from low to high. White treatment group.	p.81
Rollins	2007	*Behavioral unusualness not coded Seriousness. (WvB). Effect size based on M's, SD, and cell size (listed for number of vignettes administered)	Outcome is Outcome is the seriousness of ADHD when ADHD is present. White is treatment group.	p.71
Rollins	2007	*support seeking is not coded given when aggregated is also a measure of support. Assistance needed.(WvB). Effect size based on M's, SD, and cell size (listed for number of vignettes administered)	Outcome is Assistance needed from a little to a lot. Black as treatment group.	p.71
Rollins	2007	Timing of action. (WvB). Effect size based on M's, SD, and cell size (listed for number of vignettes administered)	Outcome is Timing of response from immediate to delayed. Black as treatment group.	p.71
Marcucci1	2020	Punitive discipline. Effect size based on M's, SD's, and N's (listed).	Outcome is Discipline from low to high punitive. Black as treatment group.	p.63
Marcucci2	2020	Punitive discipline. Effect size based on M's, SD's, and N's (listed).	Outcome is Discipline from low to high punitive. Black as treatment group.	p.63
Marcucci3	2020	Punitive discipline. Effect size based on M's, SD's, and N's (listed).	Outcome is Discipline from low to high punitive. Black as treatment group.	p.63
Marcucci4	2020	Punitive discipline. Effect size based on M's, SD's, and N's (listed).	Outcome is Discipline from low to high punitive. Black as treatment group.	p.63

Marcucci5	2020	Restorative discipline. Effect size based on M's, SD's, and N's (listed).	Outcome is Restorative discipline from restorative to high punitive. White as treatment group.	p.63
Marcucci6	2020	Restorative discipline. Effect size based on M's, SD's, and N's (listed).	Outcome is Restorative discipline from restorative to high punitive. White as treatment group.	p.63
Marcucci7	2020	Restorative discipline. Effect size based on M's, SD's, and N's (listed).	Outcome is Restorative discipline from restorative to high punitive. White as treatment group.	p.63
Marcucci8	2020	Support for exclusion/philosophy. Effect size based on M's, SD's, and N's (listed).	Outcome is low to high support of exclusion. Black as treatment group.	p.63
Golson1-2	2022	Likelihood to identify autism (WvB). Effect size based on from M's, SD's, and N (listed) from appendix. Effect size calculated for White versus black comparison for males and for females. *Confidence not rated.	Outcome is Likelihood to identify autism from low to high. White is treatment given Whites are more frequently diagnosed with autism.	SM p.12
Golson3-4	2022	Likelihood to identify disability (WvL)). Effect size based on from M's, SD's, and N (listed) from appendix. Effect size calculated for White versus black comparison for males and for females. *Confidence not rated.	Outcome is Likelihood to identify autism from low to high. White is treatment given Whites are more frequently diagnosed with autism.	SM p.12

Golson5-6	2022	Likelihood to identify disability (WvA)). Effect size based on from M's, SD's, and N (listed) from appendix. Effect size calculated for White versus black comparison for males and for females. *Confidence not rated. *Note article mentions support in method but does not provide findings for this outcome.	Outcome is Likelihood to identify autism from low to high. White is treatment given Whites are more frequently diagnosed with autism.	SM p.12
Giulietti1	2017	Response rate. (BvW). Proportion of response, N (listed and divided between treatment)	Outcome is Likelihood of receiving a response, White is treatment	p.33
Fisher1	2019	Behavior reoccurrence (BVW). Effect size based on M's, SD's, and N's (listed).	Outcome is likelihood for behavior to reoccur from unlikely to likely. Black is treatment.	p.28
Fisher2	2019	Behavior pattern (BVW). Effect size based on M's, SD's, and N's (listed).	Outcome is behavior is a pattern from disagree to agree. Black is treatment.	p.28
Fisher3	2019	Office discipline referral (BVW). Effect size based on M's, SD's, and N's (listed).	Office referral from unlikely to likely. Black is treatment	p.28
Fisher4	2019	Suspension (BVW). Effect size based on M's, SD's, and N's (listed).	Suspension from unlikely to likely. Black is treatment group	p.28
Fisher5	2019	Seriousness of punishment (BVW). Effect size based on M's, SD's, and N's (listed).	Outcome is seriousness of punishment from low to high. Black is treatment group	p.28
Fisher6	2019	Low academic(BVW). Effect size based on M's, SD's, and N's (listed).	Academic ability from high to low. Black is treatment group.	p.28

Fisher7	2019	Low cognitive (BVW). Effect size based on M's, SD's, and N's (listed).	Cognitive ability from high to low. Black is treatment group.	p.28
Fisher8	2019	Dropout (BVW). Effect size based on M's, SD's, and N's (listed).	Dropout likelihood from likely to unlikely. White is treatment.	p.28
Fisher9	2019	Academic Help (BVW). Effect size based on M's, SD's, and N's (listed).	Likelihood to experience academic help from unlikely to likely. White is treatment group.	p.28
Ash	2023	Discipline file ratings (BvW). Effect size based on M's, SD's, and N's (not listed, listed in t-test and divided evenly among treatments).	Outcome is Discipline ratings from low to high. Black is treatment	p.81
Neal1	2003	Achievement (BvW). M's, SD, N (listed)	Outcome is Achievement was rated from low to high. White is treatment.	P 53
Neal2	2003	Aggression (BvW). M's, SD, N (listed)	Outcome is Aggression is from low to high. Black is treatment group	p. 54
Neal3	2003	Referral for special education (BvW). M's, SD, N (listed)	Outcome is Referral from low to high when no evidence of disability is present. Black is treatment group.	p.54
King1	2004	Disruptive behavior index (hyperactive, inattentive, oppositional). Effect size based on M's, SD's, and cell size divided equally across vignettes.	Outcome is disruption index from low to high. Black is treatment.	p.144
King2	2004	Adaptability (BvW). Effect size based on M's, SD's, and cell size divided equally across vignettes.	Outcome is Adaptability rating from low to high. White is treatment.	p.144

King3	2004	Social skills (BvW). Effect size based on M's, SD's, and cell size divided equally across vignettes.	Outcome is Social skills rating from low to high. White is treatment.	p.144
King4	2004	Leadership (BvW). Effect size based on M's, SD's, and cell size divided equally across vignettes.	Outcome is Leadership rating from low to high. White is treatment.	p.144
Pernell1-4	1984	Need for special services for aggressive. (BvW). Effect size based on M's, SD's, and N's divided by photo yes/no and then equally across treatments. Effect sizes for light black, medium black, and dark black, and race named calculated separately.	Outcome is need for Special services from "Extremely likely" to "Extremely unlikely". White is treatment.	p.18, p.19
Pernell 5-8	1984	Need for special services for Manipulative. (BvW). Effect size based on M's, SD's, and N's divided by photo yes/no and then equally across treatments. Effect sizes for light black, medium black, and dark black, and race named calculated separately.	Outcome is Special services from "Extremely likely" to "Extremely unlikely". White is treatment.	p.18, p.19
Pernell 9-12	1984	Need for special services for None. (BvW). Effect size based on M's, SD's, and N's divided by photo yes/no and then equally across treatments. Effect sizes for light black, medium black, and dark black, and race named calculated separately.	Outcome is Special services from "Extremely likely" to "Extremely unlikely". White is treatment.	p.18, p.19
Pernell 13-16	1984	Reading ability (BvW). Effect size based on M's, SD's, and N's divided by photo yes/no and then equally across treatments. Effect sizes for light black, medium black, and dark black, and race named calculated separately.	Outcome is reading ability from low to high. White is treatment	p.20, p.21

Pernell 17-20	1984	Reading ability (BvW). Effect size based on M's, SD's, and N's divided by photo yes/no and then equally across treatments. Effect sizes for light black, medium black, and dark black, and race named calculated separately.	Outcome is reading ability from low to high. White is treatment	p.20, p.21
Pernell 21-24	1984	Reading ability (BvW). Effect size based on M's, SD's, and N's divided by photo yes/no and then equally across treatments. Effect sizes for light black, medium black, and dark black, and race named calculated separately.	Outcome is reading ability from low to high. White is treatment	p.20, p.21
Pernell 25-28	1984	Social adjustment. Effect size based on M's, SD's, and N's divided by photo yes/no and then equally across treatments. Effect sizes for light black, medium black, and dark black, and race named calculated separately.	Outcome is social adjustment from poor to well. White is treatment	p.22
Pernell 29-32	1984	Social adjustment. Effect size based on M's, SD's, and N's divided by photo yes/no and then equally across treatments. Effect sizes for light black, medium black, and dark black, and race named calculated separately.	Outcome is social adjustment from poor to well. White is treatment	p.22
Pernell 33-36	1984	Social adjustment. Effect size based on M's, SD's, and N's divided by photo yes/no and then equally across treatments. Effect sizes for light black, medium black, and dark black, and race named calculated separately.	Outcome is social adjustment from poor to well. White is treatment	p.22

Petts	2020	Schools supported (WvAgg Minority). Frequency selected minority school versus frequency selected majority school, N's (listed; the results are presented by participant average school viewed, thus vignette manipulations were not used as the cell size).	Outcome is Selection frequency from no to yes. White is treatment group.	T1
Kindaichi1	2010	Rating of Functioning. (WvB). Effect size calculated as N's (listed), M's, SD's	Outcome is Functioning from low to high. White is control.	p.75
Kindaichi2	2010	Functioning. (WvA). Effect size calculated as N's (listed), M's, SD's	Outcome is Functioning from low to high when functioning should be rated as low. White is control. (notice reverse scale)	p.75
Kindaichi3	2010	Functioning. (WvMultiracial black-White). Effect size calculated as N's (listed), M's, SD's	Outcome is Functioning from low to high when functioning should be rated as low. White is control. (notice reverse scale)	p.75
Kindaichi4	2010	Functioning. (WvMultiracial black-asian). Effect size calculated as N's (listed), M's, SD's	Outcome is Functioning from low to high when functioning should be rated as low. White is control. (notice reverse scale)	p.75
Kindaichi5	2010	Functioning. (WvMultiracial asian-White). Effect size calculated as N's (listed), M's, SD's	Outcome is Functioning from low to high when functioning should be rated as low. White is control. (notice reverse scale)	p.75

Kunesh 1 & 2	2019	Behavior inappropriateness (BvW). Effect size calculated based on M's, SD's, and N's (listed). Separate effect sizes were calculated for Explicit White versus Explicit Black and Explicit White versus Implicit Black	Outcome is Behavior as extremely inappropriate to not inappropriate. White is treatment.	p.486
Kunesh 3 & 4	2019	Behavior pattern. (BvW). Effect size calculated based on M's, SD's, and N's (listed). Separate effect sizes were calculated for Explicit White versus Explicit Black and Explicit White versus Implicit Black	Outcome is Behavior as likely to reoccur unlikely to likely. Black is treatment group.	p.486
Brinkman	2022	*study 2 not coded given sample is college students. White parents ratings of Classroom selection. (BvW). Effect size based on Pearson r correlation between % Black in classroom and rating of likelihood of selecting classroom for pre-school child. Calculated using OSF available data	Outcome is Likelihood to enroll. White is treatment group.	OSF data
Woods 1	2022	Externalizing (BvW). Effect size based on M's, SD's, and N (listed).	Outcome is Externalizing from low to high. Black is treatment group.	p.10
Woods 2	2022	Internalizing (BvW). Effect size based on M's, SD's, and N (listed).	Outcome is Internalizing from low to high. Black is treatment group.	p.10
Woods 3	2022	Academic Functioning (BvW). Effect size based on M's, SD's, and N (listed).	Outcome is Academic functioning from low to high. White is treatment group.	p.10
Woods 4	2022	Graduation (BvW). Effect size based on M's, SD's, and N (listed).	Outcome is Likelihood to graduate from low to high. White is treatment group.	p.10

Woods 5	2022	Quality of Home life (BvW). Effect size based on M's, SD's, and N (listed).	Outcome is Quality of home life from low to high. White is treatment group.	p.10
Woods 6	2022	Refer to special education (BvW). Effect size based on M's, SD's, and N (listed).	Outcome is referral from refer to not refer when referral is needed. White is control group because of reverse coding.	p.10
Woods 7	2022	Externalizing (AvW). Effect size based on M's, SD's, and N (listed).	Outcome is Externalizing from low to high. Asian is treatment group.	p.10
Woods 8	2022	Internalizing (AvW). Effect size based on M's, SD's, and N (listed).	Outcome is Internalizing from low to high. Asian is treatment group.	p.10
Woods 9	2022	Academic Functioning (AvW). Effect size based on M's, SD's, and N (listed).	Outcome is Academic functioning from low to high. White is treatment group.	p.10
Woods 10	2022	Graduation (AvW). Effect size based on M's, SD's, and N (listed).	Outcome is Likelihood to graduate from low to high. White is treatment group.	p.10
Woods 11	2022	Quality of Home life (AvW). Effect size based on M's, SD's, and N (listed).	Outcome is Quality of home life from low to high. White is treatment group.	p.10
Woods 12	2022	Refer to special education (AvW). Effect size based on M's, SD's, and N (listed).	Outcome is referral from refer to not refer when referral is needed. White is control group because of reverse coding.	p.10

Raymond 1 & 11	1997	Has a learning disability (NvW). Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is disability from disagree to agree. Native is treatment because BIPOC overrepresented in special education	p.42
Raymond 2 & 12	1997	Oppositional (NvW) Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is oppositional from disagree to agree. Native is treatment	p.42
Raymond 3 & 13	1997	ADHD (NvW) Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is ADHD from disagree to agree. White is treatment because White is overrepresented in ADHD diagnosis.	p.42
Raymond 4 & 14	1997	Average achiever (NvW) Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is average achiever without interventions from disagree to agree. White is treatment	p.42
Raymond 5 & 15	1997	Suspended (NvW) Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is suspended from from disagree to agree. Native is treatment	p.42
Raymond 6 & 16	1997	Special education placement (NvW) Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is special education from disagree to agree. Native is treatment.	p.42
Raymond 7 & 17	1997	Social emotional disturb (NvW) Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is socially emotionally disturbed from disagree to agree. Native is treatment	p.42

Raymond 8 & 18	1997	Graduate high school (NvW) Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is graduate high school from disagree to agree. White is treatment	p.42
Raymond 9 & 19	1997	Graduate college (NvW) Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is graduate college from disagree to agree. White is treatment	p.42
Raymond 10 & 20	1997	Taught in regular classroom (NvW) Effect size based on M's, SD, N (not listed, sample size divided equally across treatment groups). Native traditional and native assimilated.	Outcome is being taught in a regular classroom from disagree to agree. White is treatment	p.42
Ura 1-3	2022	Referral to counselor for compliant behavior (BvW). Effect size based on M's, N's. (listed) and SD's (obtained from author). Effect size calculated for high, medium and low SEC teacher.	Outcome is referral to counselor. Black is treatment group.	p.22
Ura 4-6	2022	Referral to counselor for aggressive behavior (BvW). Effect size based on M's, N's. (listed) and SD's (obtained from author). Effect size calculated for high, medium and low SEC teacher.	Outcome is referral to counselor from unlikely to likely. Black is treatment group.	p.22
Ura 7-9	2022	Referral to counselor for withdrawn behavior (BvW). Effect size based on M's, N's. (listed) and SD's (obtained from author). Effect size calculated for high, medium and low SEC teacher.	Outcome is referral to counselor from unlikely to likely. Black is treatment group.	p.22

Ura 10-12	2022	Referral to counselor for inattentive behavior (BvW). Effect size based on M's, N's. (listed) and SD's (obtained from author). Effect size calculated for high, medium and low SEC teacher.	Outcome is referral to counselor from unlikely to likely. Black is treatment group.	p.22
Ura 13-15	2022	Referral to counselor for hyperactive behavior (BvW). Effect size based on M's, N's. (listed) and SD's (obtained from author). Effect size calculated for high, medium and low SEC teacher.	Outcome is referral to counselor from unlikely to likely. Black is treatment group.	p.22

APPENDIX D

SENSITIVITY EXPLORATION OF THE AGGREGATE FUNCTION

Moderator	*k*	*d*	[95% CI]		*t*	$Q_{residual}$	*F*
Experimental Method							
Audit	7	0.08	[-0.26 ,	0.41]	0.45	1430.49***	1.76
Task	85	0.28	[0.12 ,	0.44]	3.39***		
Vignette	249	0.08	[-.03 ,	0.19]	1.40		
Natural	2	—	——	——	—		
Perpetrator Type							
Counselor	46	0.8	[-0.15 ,	0.31]	0.66	1336.48***	1.29
District Office	1	—	——	——	—		
Hiring	2	—	——	——	—		
Parent/Community	14	0.21	[-0.06 ,	0.47]	1.55		
Policy	2	—	——	——	—		
Principal	19	0.18	[-0.19 ,	0.54]	0.94		
Student/pupil	4	0.65	[0.26 ,	1.04]	3.30***		
Teacher	255	0.09	[-0.02 ,	0.20]	1.56		
Level of Racism							
Individual Racism	18	0.35	[0.13	0.56]	3.19**	1402.39***	4.35*
Institutional Racism	325	0.10	[0.01	0.19]	2.10*		
Sample Composition							
White	47	0.17	[-0.00 ,	0.34]	1.97*	1348.99***	0.17
Mixed	296	0.13	[0.03 ,	0.23]	.59**		
Target Race							
Black-White	280	0.14	[0.05 ,	0.25]	3.10**	1441.60***	0.14
Latine-White	27	0.16	[0.04 ,	0.27]	2.65**		
Native-White	10	—	——	——	—		
Asian-White	20	0.14	[-0.07 ,	0.36]	1.35		
Minority-White	3	0.06	[-0.18 ,	0.30]	0.52		
Biracial-White	3	—	——	——	—		
Region							
Midwest ^{S}	60	0.05	[-0.24 ,	0.34]	0.35	1076.46***	2.20+
All/Multiple ^{S}	115	-0.11	[-0.04 ,	0.25]	1.42		
Northeast ^{S}	40	0.04	[-0.24 ,	0.33]	0.30		
South ^{M,A,N,W}	53	0.50	[0.23 ,	0.75]	3.74***		
West ^{S}	34	0.03	[-0.25 ,	0.31]	0.22		
Study Design							
Between-Subjects	224	0.17	[0.06 ,	0.28]	3.09**	1450.70***	0.98
Within-Subjects	119	0.08	[-0.07 ,	0.23]	1.07		
Publication Status							
Published	200	0.15	[0.05 ,	0.27]	2.79**	1451.30***	0.28
Unpublished	143	0.11	[-0.04 ,	0.26]	1.45		

Note. *k* = number of effect sizes; *d* = the overall association between the moderator and the effect of racism; 95% CI = 95% confidence interval; F = value of the test of moderators; $Q_{residual}$ = residual heterogeneity σ^2_1 and σ^2_2 = variance within- and between-study, respectively. Positive effects indicate that greater effect of racism.

APPENDIX E

SYNTAX FOR ANALYSES IN R

```
install.packages("metafor")
install.packages("meta")
install.packages("RCurl")
install.packages("bitops")
install.packages("Formula")
install.packages('forestplot')
install.packages('MAd')
library("RCurl")
library("metafor")
library("bitops")
library("Formula")
library('forestplot')
library('dplyr')
library('MAd')
library('tidyverse')
library('readxl')

#################################################################
##########descriptive statistics
#################################################################
Study <- read_excel("Desktop/Study.xlsx")
aggStudy<- agg(id = id, es = D ,var = V, n.1 = Nx1, n.2 =Nx2, cor =0.5, method =
"BHHR", data = Study)
Study1 <- merge(aggStudy, Study[, c(1,2:105)], by='id')
Study0<- Study1[!duplicated(Study1$id),]
Study2<-escalc(yi=es, vi=var, data=Study0)
datT<- escalc(yi=D, vi=V, data=dat)
aggregate(Study2$White, list(Study2$PubStatus), FUN=sum)
aggregate(Study2$White, list(Study2$N), FUN=sum)
sum(Study2$N)
aggregate(Study2$White, list(Study2$C.Region), FUN=sum)
aggregate(Study2$N, list(Study2$C.Region), FUN=sum)
aggregate(Study2$White, list(Study2$ExMethod), FUN=sum)
aggregate(Study2$White, list(Study2$Assignment), FUN=sum)
aggregate(Study2$White, list(Study2$Perpetrator), FUN=sum)
aggregate(Study2$White, list(Study2$RaceComp), FUN=sum)

#################################################################
####################Main analyses################################
#################################################################################
dat <- read_excel("Desktop/SubgroupSensitivity.xlsx")
class(dat)
agg<- agg(id = id, es = D ,var = V, n.1 = Nx1, n.2 =Nx2, cor =0.5, method = "BHHR",
data = dat)
```

```
dat1 <- merge(agg, dat[, c(1,2:106)], by='id')
dat0<- dat1[!duplicated(dat1$id),]
dat2<-escalc(yi=es, vi=var, data=dat0)
datT<- escalc(yi=D, vi=V, data=dat)

#BHHR procedure was found to be the least biased and most precise (Hoyt & Del Re,
2015).
#Note the imputation of r = 0.5. This value was chosen because it is a conservative (and
#typical) starting value for aggregating psychologically-based ESs (e.g., Wampold,
Mondin, Moody, et al.,
#1997). Availability of between-measure correlations
#within each study are often not available and such
#starting imputation values are reasonable, although
#sensitivity analyses with several values (e.g., ranging
#perhaps from r=.3 to r=.7, although these values may
#differ depending on the particular substantive area
#under investigation) are recommended prior to
#running omnibus or meta-regression models.

# Estimate the overall effect by fitting an intercept-only model.
overall <- rma.mv(yi, vi, random = list(~ 1 | id, ~ 1 | MLID),
        tdist= TRUE, data=dat2)
summary(overall, digits=3)

#Plots and figures
mlabfun <- function(text, overall) {
 list(bquote(paste(.(text),
          " (Q = ", .(formatC(overall$QE, digits=2, format="f")),
          ", df = ", .(overall$k - overall$p),
          ", p ", .(metafor:::.pval(overall$QEp, digits=2, showeq=TRUE, sep=" ")), "; ",
          I^2, " = ", .(formatC(overall$I2, digits=1, format="f")), "%, ",
          tau^2, " = ", .(formatC(overall$tau2, digits=2, format="f")), ")")))}

forest(overall, slab = paste(dat2$Author, as.character(dat2$Year), sep = ", "),
   xlim=c(-5,5), digits=c(2,1), addpred=TRUE, order="obs",
   ilab=(BIPOC), ilab.xpos=c(-2),
   cex=.75)
text((-4.3), overall$k+2, ("Author and Year"), cex=.75, font=2)
text((-2), overall$k+2, ("Target Race"), cex=.75, font=2)
text((4.2), overall$k+2, ("Cohen's d [95%CI]"), cex=.75, font=2)

# Build a two-level model without within-study variance.
modelnovar2 <- rma.mv(yi, vi, random = list(~ 1 | id, ~ 1 | MLID),
         sigma2=c(0,NA), tdist=TRUE, data=dat2)
```

```
# Perform a likelihood-ratio-test to determine the
# significance of the within-study variance.
anova(overall,modelnovar2)

# Build a two-level model without between-study variance;
# Perform a likelihood-ratio-test to determine the
# significance of the between-study variance.
modelnovar3 <- rma.mv(yi, vi, random = list(~ 1 | id, ~ 1 | MLID),
        sigma2=c(NA,0), tdist=TRUE, data=dat2)
anova(overall,modelnovar3)

# Determining how the total variance is distributed over the
# three levels of the meta-analytic model;
# Print the results in percentages on screen.
n <- length(dat2$vi)
list.inverse.variances <- 1 / (dat2$vi)
sum.inverse.variances <- sum(list.inverse.variances)
squared.sum.inverse.variances <- (sum.inverse.variances) ^ 2
list.inverse.variances.square <- 1 / (dat2$vi^2)
sum.inverse.variances.square <-
 sum(list.inverse.variances.square)
numerator <- (n - 1) * sum.inverse.variances
denominator <- squared.sum.inverse.variances -
 sum.inverse.variances.square
estimated.sampling.variance <- numerator / denominator
I2_1 <- (estimated.sampling.variance) / (overall$sigma2[1]
              + overall$sigma2[2] + estimated.sampling.variance)
I2_2 <- (overall$sigma2[1]) / (overall$sigma2[1]
            + overall$sigma2[2] + estimated.sampling.variance)
I2_3 <- (overall$sigma2[2]) / (overall$sigma2[1]
            + overall$sigma2[2] + estimated.sampling.variance)
amountvariancelevel1 <- I2_1 * 100
amountvariancelevel2 <- I2_2 * 100
amountvariancelevel3 <- I2_3 * 100
amountvariancelevel1
amountvariancelevel2
amountvariancelevel3

#############################################################################
##############################Moderation
#############################################################################

########Experimental Method
```

```
# Determine the potential moderating effect of Experimental technique.
aggregate(dat2$White, list(dat2$ExMethod), FUN=sum)
dat2$ExMethod
dat2$Task <- car::recode(dat2$ExMethod,"
            'Task' = 1;
            'Vignette' = 0;
            'Audit'=0;
            'Natural'=0")
dat2$Vignette <- car::recode(dat2$ExMethod,"
            'Task' = 0;
            'Vignette' = 1;
            'Audit'=0;
            'Natural'=0")
dat2$Audit <- car::recode(dat2$ExMethod,"
            'Task' = 0;
            'Vignette' = 0;
            'Audit'=1;
            'Natural'=0")
dat2$Natural <- car::recode(dat2$ExMethod,"
            'Task' = 0;
            'Vignette' = 0;
            'Audit'=0;
            'Natural'=1")

mTask<- rma.mv(yi, vi, mods = ~ Vignette + Audit + Natural, random = list(~ 1 | id, ~1 |
MLID), tdist=TRUE, data=dat2)
summary(mTask, digits=3)

mVignette<- rma.mv(yi, vi, mods = ~ Task + Audit + Natural, random = list(~ 1 | id, ~1 |
MLID), tdist=TRUE, data=dat2)
summary(mVignette, digits=3)

mAudit<- rma.mv(yi, vi, mods = ~ Task + Vignette + Natural, random = list(~ 1 | id, ~1 |
MLID), tdist=TRUE, data=dat2)
summary(mAudit, digits=3)

########Perpetrator type
# Determine the potential moderating effect of the perpetrator.
aggregate(dat2$White, list(dat2$Perpetrator), FUN=sum)
dat2$Counselor <- car::recode(dat2$Perpetrator, "
                'Counselor/psychologist'=1;
                'District office'=0;
                'Hiring'=0;
                'Parent/community'=0;
```

```
                'Policy'=0;
                'Principal'=0;
                'Student/pupil'=0;
                'Teacher'=0")
dat2$District <- car::recode(dat2$Perpetrator, "
                'Counselor/psychologist'=0;
                'District office'=1;
                'Hiring'=0;
                'Parent/community'=0;
                'Policy'=0;
                'Principal'=0;
                'Student/pupil'=0;
                'Teacher'=0")
dat2$Hiring <- car::recode(dat2$Perpetrator, "
                'Counselor/psychologist'=0;
                'District office'=0;
                'Hiring'=1;
                'Parent/community'=0;
                'Policy'=0;
                'Principal'=0;
                'Student/pupil'=0;
                'Teacher'=0")
dat2$System <- car::recode(dat2$Perpetrator, "
                'Counselor/psychologist'=0;
                'District office'=.333333;
                'Hiring'=.333333;
                'Parent/community'=0;
                'Policy'=.333333;
                'Principal'=0;
                'Student/pupil'=0;
                'Teacher'=0")
dat2$Policy <- car::recode(dat2$Perpetrator, "
                'Counselor/psychologist'=0;
                'District office'=0;
                'Hiring'=0;
                'Parent/community'=0;
                'Policy'=1;
                'Principal'=0;
                'Student/pupil'=0;
                'Teacher'=0")
dat2$Parent <- car::recode(dat2$Perpetrator, "
                'Counselor/psychologist'=0;
                'District office'=0;
                'Hiring'=0;
```

```
                'Parent/community'=1;
                'Policy'=0;
                'Principal'=0;
                'Student/pupil'=0;
                'Teacher'=0")
dat2$Principal <- car::recode(dat2$Perpetrator, "
                'Counselor/psychologist'=0;
                'District office'=0;
                'Hiring'=0;
                'Parent/community'=0;
                'Policy'=0;
                'Principal'=1;
                'Student/pupil'=0;
                'Teacher'=0")
dat2$Student <- car::recode(dat2$Perpetrator, "
                'Counselor/psychologist'=0;
                'District office'=0;
                'Hiring'=0;
                'Parent/community'=0;
                'Policy'=0;
                'Principal'=0;
                'Student/pupil'=1;
                'Teacher'=0")
dat2$Teacher <- car::recode(dat2$Perpetrator, "
                'Counselor/psychologist'=0;
                'District office'=0;
                'Hiring'=0;
                'Parent/community'=0;
                'Policy'=0;
                'Principal'=0;
                'Student/pupil'=0;
                'Teacher'=1")
ModelCounselor <- rma.mv(yi, vi, mods = ~ (Policy + Hiring + District + Parent +
Principal + Student + Teacher),
        random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelCounselor, digits=3)
ModelParent<- rma.mv(yi, vi, mods = ~ Policy + Hiring + District + Counselor +
Principal + Student + Teacher,
       random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelParent, digits=3)
ModelPrincipal<- rma.mv(yi, vi, mods = ~Policy + Hiring + District + Counselor +
Parent + Student + Teacher,
        random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelPrincipal, digits=3)
```

```
ModelStudent <- rma.mv(yi, vi, mods = ~ Policy + Hiring + District + Counselor +
Parent + Principal + Teacher,
        random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelStudent, digits=3)
ModelTeacher <- rma.mv(yi, vi, mods = ~ Policy + Hiring + District + Counselor +
Parent + Principal + Student,
        random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelTeacher, digits=3)


# Determine the potential moderating effect of institutional vs individual.
aggregate(dat2$White, list(dat2$PerpTarget), FUN=sum)
PerpTarget
dat2$Inst <- car::recode(dat2$PerpTarget,"
      'Parent to school'=0;
      'Policy to student'=1;
      'Principal to student'=1;
      'Psychologist to student'=1;
      'School to parent' =1;
      'District to families'=1;
      'School to Teacher' =1;
      'Student to peer' =0;
      'Teacher to student' =1")
dat2$Individ <- car::recode(dat2$PerpTarget,"
      'Parent to school'=1;
      'Policy to student'=0;
      'Principal to student'=0;
      'Psychologist to student'=0;
      'School to parent' =0;
      'District to families'=0;
      'School to Teacher' =0;
      'Student to peer' =1;
      'Teacher to student' =0")
IndividualR<- rma.mv(yi, vi, mods = ~ Inst, random = list(~ 1 | id, ~1 | MLID),
tdist=TRUE, data=dat2)
summary(IndividualR, digits=3)
InstitutionalR<- rma.mv(yi, vi, mods = ~ Individ, random = list(~ 1 | id, ~1 | MLID),
tdist=TRUE, data=dat2)
summary(InstitutionalR, digits=3)

# Determine the potential moderating effect of White sample.
dat2$WhiteSamp <- car::recode(dat2$WhiteOnly,"
      '1'=1;
      '0'=0")
```

```
dat2$MixedSamp <- car::recode(dat2$WhiteOnly,"
      '1'=0;
      '0'=1")
WhiteOnly<- rma.mv(yi, vi, mods = ~ MixedSamp, random = list(~ 1 | id, ~1 | MLID),
tdist=TRUE, data=dat2)
summary(WhiteOnly, digits=3)

MixedSamp<- rma.mv(yi, vi, mods = ~ WhiteSamp, random = list(~ 1 | id, ~1 | MLID),
tdist=TRUE, data=dat2)
summary(MixedSamp, digits=3)


########Target race
# Determine the potential moderating effect of target race.
aggregate(dat2$White, list(dat2$RaceComp), FUN=sum)
dat2$Black <- car::recode(dat2$RaceComp, "1 = 1; 2=0;3=0;4=0; 5=0; 6=0")
dat2$Latine <- car::recode(dat2$RaceComp, "1 = 0; 2=1;3=0;4=0; 5=0; 6=0")
dat2$Asian <- car::recode(dat2$RaceComp, "1 = 0; 2=0;3=0;4=1; 5=0; 6=0")
dat2$Native <- car::recode(dat2$RaceComp, "1 = 0; 2=0;3=1;4=0; 5=0; 6=0")
dat2$Minority <- car::recode(dat2$RaceComp, "1 = 0; 2=0;3=0;4=0; 5=1; 6=0")
dat2$Biracial <- car::recode(dat2$RaceComp, "1 = 0; 2=0;3=0;4=0; 5=0; 6=1")
ModelBlack <- rma.mv(yi, vi, mods = ~ Latine + Asian + Native + Minority + Biracial,
         random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelBlack, digits=3)
ModelLatine<- rma.mv(yi, vi, mods = ~ Black + Asian + Native + Minority + Biracial,
         random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelLatine, digits=3)
ModelAsian<- rma.mv(yi, vi, mods = ~ Black + Latine + Native + Minority + Biracial,
         random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelAsian, digits=3)
ModelMinority<- rma.mv(yi, vi, mods = ~ Black + Latine + Native + Asian + Biracial,
        random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelMinority, digits=3)


########Region
# Determine the potential moderating effect of region.
dat2$All <- car::recode(dat2$C.Region, "'Multiple/All' = 1;
'South'=0;'Midwest'=0;'Northeast'=0; 'West'=0")
dat2$South <- car::recode(dat2$C.Region, "'Multiple/All' = 0;
'South'=1;'Midwest'=0;'Northeast'=0; 'West'=0")
dat2$Midwest <- car::recode(dat2$C.Region, "'Multiple/All' = 0;
'South'=0;'Midwest'=1;'Northeast'=0; 'West'=0")
```

```
dat2$Northeast <- car::recode(dat2$C.Region, "'Multiple/All' = 0;
'South'=0;'Midwest'=0;'Northeast'=1; 'West'=0")
dat2$West <- car::recode(dat2$C.Region, "'Multiple/All' = 0;
'South'=0;'Midwest'=0;'Northeast'=0; 'West'=1")
ModelSouth <- rma.mv(yi, vi, mods = ~ All + Midwest + Northeast + West,
        random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelSouth, digits=3)
ModelMidwest <- rma.mv(yi, vi, mods = ~ All + South + Northeast + West,
        random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelMidwest, digits=3)
ModelNortheast <- rma.mv(yi, vi, mods = ~ All + South + Midwest + West,
         random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelNortheast, digits=3)
ModelWest <- rma.mv(yi, vi, mods = ~ All + South + Midwest + Northeast,
         random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelWest, digits=3)
ModelAll <- rma.mv(yi, vi, mods = ~ West + South + Midwest + Northeast,
        random = list(~ 1 | id, ~ 1 | MLID), tdist=TRUE, data=dat2)
summary(ModelAll, digits=3)


########Participant design
# Determine the potential moderating effect of participant design.
dat2$Within <- car::recode(dat2$Assignment, "'Within-subjects' = 1; 'Random'=0")
dat2$Between <- car::recode(dat2$Assignment, "'Within-subjects' = 0; 'Random'=1")
Within <- rma.mv(yi, vi, mods = ~ Between, random = list(~ 1 | id, ~
                        1 | MLID), tdist=TRUE, data=dat2)
summary(Within, digits=3)
Between <- rma.mv(yi, vi, mods = ~ Within, random = list(~ 1 | id, ~
                        1 | MLID), tdist=TRUE, data=dat2)
summary(Between, digits=3)


########Publication status
# Determine the potential moderating effect of publication status.
dat2$pstatpub <- car::recode(dat2$PubStatus, "'Peer-reviewed publication' = 1;
        'Brief' = 0; 'Dissertation'=0; 'Thesis'=0")
dat2$pstatnotpub <- car::recode(dat2$PubStatus, "'Peer-reviewed publication' = 0;
        'Brief' = 1; 'Dissertation'=1; 'Thesis'=1")
aggregate(dat2$White, list(dat2$PubStatus), FUN=sum)
notpublished <- rma.mv(yi, vi, mods = ~ pstatpub, random = list(~ 1 | id, ~
                        1 | MLID), tdist=TRUE, data=dat2)
summary(notpublished, digits=3)
published <- rma.mv(yi, vi, mods = ~ pstatnotpub, random = list(~ 1 | id, ~
```

```
                        1 | MLID), tdist=TRUE, data=dat2)
summary(published, digits=3)

#########################################################################
####################Final Model
#########################################################################
# Estimate the final overall effect by fitting an intercept-only model.
overallfinal <- rma.mv(yi, vi, mods = ~ Individ + All + Midwest + Northeast + West,
       random = list(~ 1 | id, ~ 1 | MLID),
       tdist= TRUE, data=dat2)
summary(overallfinal, digits=3)
forest(overallfinal,slab = paste(dat2$BIPOC, dat2$Author, as.character(dat2$Year), sep =
", "),
   xlim=c(-4,5), digits=c(2,1), addpred=TRUE, header="Author and Year", order="obs")

#Once you have tested everything. Build the final model to examine the residual within
study and
# within-study variance.
modelnovarR2 <- rma.mv(yi, vi, mods = ~ Individ + All + Midwest + Northeast + West,
        random = list(~ 1 | id, ~ 1 | MLID),
        sigma2=c(0,NA), tdist=TRUE, data=dat2)
anova(overall,modelnovarR2)

# between-study variance.
modelnovar3 <- rma.mv(yi, vi, mods = ~ Individ + All + Midwest + Northeast + West,
        random = list(~ 1 | id, ~ 1 | MLID),
        sigma2=c(0,NA), tdist=TRUE, data=dat2)
anova(modelnovar2,modelnovar3)

modelnovar3 <- rma.mv(yi, vi, mods = ~ Individ + All + Midwest + Northeast + West,
        random = list(~ 1 | id, ~ 1 | MLID),
        sigma2=c(0,NA), tdist=TRUE, data=dat2)

#########################################################################
####################Publication Bias
#########################################################################
Eggers <- rma.mv(yi, vi, mods = ~ vi,
        random = list(~ 1 | id, ~ 1 | MLID),
        sigma2=c(0,NA), tdist=TRUE, data=dat2)
Eggers

#Funnel plot
funnelplot0 <- rma.mv(yi, vi,
        random = list(~ 1 | id, ~ 1 | MLID),
```

```
        sigma2=c(0,NA), tdist=TRUE, data=dat2, slab=paste(Author, Year, sep=", "))
funnelplot2 <- rma.mv(yi, vi, mods = ~ Individ,
        random = list(~ 1 | id, ~ 1 | MLID),
        sigma2=c(0,NA), tdist=TRUE, data=dat2, slab=paste(Author, Year, sep=", "))
funnelplot1 <- rma.mv(yi, vi, mods = ~ All + Midwest + Northeast + West,
        random = list(~ 1 | id, ~ 1 | MLID),
        sigma2=c(0,NA), tdist=TRUE, data=dat2, slab=paste(Author, Year, sep=", "))

funnel(funnelplot0, label="out")

funnel(funnelplot0, level=c(90, 95, 99), shade=c("White", "gray55", "gray75"),
   refline=0, legend=TRUE, cex.lab = 1, cex.axis = 1,
   digits=3L, at=(c(-2, -1, 0, 1, 2)))
funnel(funnelplot1, level=c(90, 95, 99), shade=c("White", "gray55", "gray75"),
   refline=0, legend=TRUE, cex.lab = 1, cex.axis = 1,
   digits=3L, at=(c(-2, -1, 0, 1, 2)))
funnel(funnelplot2, level=c(90, 95, 99), shade=c("White", "gray55", "gray75"),
   refline=0, legend=TRUE, cex.lab = 1, cex.axis = 1,
   digits=3L, at=(c(-2, -1, 0, 1, 2)))

#########################################################################
####################Analysis without aggregate function
#########################################################################
# All analyses were repeated, but the data were not aggregated and the fourth level was
added to the random command. Differences are highlighted below.

#Three level model
dat <- read_excel("Desktop/SubgroupSensitivity.xlsx")
agg<- agg(id = id, es = D ,var = V, n.1 = Nx1, n.2 =Nx2, cor =0.5, method = "BHHR",
data = dat)
dat1 <- merge(agg, dat[, c(1,2:106)], by='id')
dat0<- dat1[!duplicated(dat1$id),]
dat2<-escalc(yi=es, vi=var, data=dat0)
overall <- rma.mv(yi, vi, random = list(~ 1 | id, ~ 1 | MLID),
        tdist= TRUE, data=dat2)
summary(overall, digits=3)

#Four level model
dat <- read_excel("Desktop/SubgroupSensitivity.xlsx")
datT<- escalc(yi=D, vi=V, data=dat)
overall0 <- rma.mv(yi, vi, random = list(~1 | MLESID, ~ 1 | id, ~ 1 | MLID),
        tdist= TRUE, data=datT)
summary(overall0, digits=3)
```

APPENDIX E

SENSITIVITY EXPLORATION OF THE MULTILEVEL MODEL

Moderator	*k*	*d*	[95% CI]		*t*	$Q_{residual}$	*F*
Experimental Method							
Audit	7	0.11	[0.09 ,	0.14]	8.86***	578.17***	5.02**
Task	17	0.09	[0.07 ,	0.12]	7.36***		
Vignette	45	0.16	[0.13 ,	0.19]	10.80***		
Natural	2	—	——	——	—		
Perpetrator Type							
Counselor ^{Pr,Sp}	15	0.5	[-0.05 ,	0.14]	0.94	541.68***	7.37***
District Office	1	—	——	——	—		
Hiring	2	—	——	——	—		
Parent/Community ^{C,Pr,Sp,T}	9	0.13	[0.11 ,	0.16]	10.17***		
Policy	2	—	——	——	—		
Principal $^{Pc, Sp}$	3	0.07	[-0.04 ,	0.14]	1.89+		
Student/pupil ^{C,Pc,Pr,T}	3	0.47	[0.35 ,	0.59]	7.76***		
Teacher ^{Pc,Sp}	36	0.08	[0.05 ,	0.11]	5.69***		
Level of Racism							
Individual Racism	12	.015	[0.12	0.18]	11.60***	582.25***	10.99***
Institutional Racism	59	0.10	[0.08	0.11]	10.63***		
Sample Composition							
White	16	0.31	[0.12 ,	0.50]	3.20**	579.56***	3.60+
Mixed	55	.10	[-0.01 ,	0.21]	1.74+		
Target Race							
Black-White	52	0.11	[0.10 ,	0.13]	13.20***	590.09***	0.63
Latine-White	10	0.14	[0.10 ,	0.18]	6.58***		
Native-White	1	—	——	——	—		
Asian-White	4	0.06	[-0.12 ,	0.23]	0.64		
Minority-White	3	0.12	[0.07 ,	0.18]	4.37***		
Biracial-White	1	—	——	——	—		
Region							
Midwest	6	0.12	[-0.07 ,	0.30]	1.28	478.44***	7.48***
All/Multiple $^{S, N}$	32	0.10	[0.08 ,	0.12]	9.84***		
Northeast ^{A,W}	8	0.22	[0.17 ,	0.26]	9.16***		
South ^{A,W}	8	0.20	[0.14 ,	0.26]	6.65***		
West $^{S, N}$	6	0.06	[-0.00 ,	0.12]	1.94+		
Study Design							
Between-Subjects	47	0.11	[0.09 ,	0.14]	8.97***	593.07***	0.18
Within-Subjects	24	0.12	[0.10 ,	0.14]	12.50***		
Publication Status							
Published	47	0.11	[0.10 ,	0.13]	14.12***	585.54***	7.70**
Unpublished	24	0.19	[0.14,	0.25]	6.70***		