

Unveiling Racism: A Systematic Review of Racism Survey Measures in Education

Wendy Castillo¹, Juan Cruz², Rachel Renbarger³,
Christen Priddie⁴, Brein Mosely²,
and Sasha Mejia-Bradford⁵

Abstract

Education policy research aimed at eliminating racism necessitates methodological innovation that fosters both racial equity-centered approaches and robust empirical analysis of the systemic and structural nature of racism. Most quantitative research in educational psychology does not account for racism in its modeling, thereby obscuring the racialized environments in which students learn (DeCuir-Gunby & Schutz, 2014; Strunk & Andrzejewski, 2023). Drawing from Bronfenbrenner's ecological systems theory and the ethos of QuantCrit, this study evaluates various scales to highlight gaps and challenges in measuring racism within school settings. This review aims to uplift measures for researchers to use in their work and advocate for more targeted resources and reparations. Reviewing 66 studies and 50 unique scales, key findings include the overreliance on scales designed to measure racial discrimination at the intrapersonal level, while systemic and structural factors are often overlooked. Most measures were self-report surveys for older students developed more than 15 years ago. The results demonstrate the need to move beyond measures of intrapersonal racism to more up-to-date, multidimensional tools. The study's findings also underscore the necessity of constant reevaluation of the tools so they keep pace with the shifting cultural realities of how racism operates, and the specific ways it

¹Montclair State University, NJ, USA

²Harvard University, Cambridge, MA, USA

³Independent Scholar, Durham, NC, USA

⁴Indiana University, Bloomington, USA

⁵University of Pennsylvania, Philadelphia, USA

Corresponding Author:

Wendy Castillo, Montclair State University, Montclair, NJ 07043-1624, USA.

Email: castillow@montclair.edu

impacts racially minoritized communities. As racism continues to evolve, so too must the methods by which it is measured.

Keywords

diversity, validity/reliability, race, ethnicity, equity, psychometrics, survey research, measurements, racism, race, ethnicity, discrimination, systemic, structural, institutional, interpersonal, intrapersonal, survey

Introduction

Racism, not race, is a root cause of inequities in educational and life outcomes. These inequities stem from an egregious history of colonization, disenfranchisement, dehumanization, forced removal, genocide, and enslavement (Bailey et al., 2017) and have been well-documented (c.f. Harber et al., 2012; Kozol, 1991; Nicholson-Crotty et al., 2016; Okonofua & Eberhart, 2015; U.S. Department of the Treasury, 2023). Furthermore, the presence of racism heavily shapes the K–12 educational experiences of all students, negatively impacting racially minoritized students (Bridgeforth, 2025; Kohli et al., 2017; Oto et al., 2023). Nonetheless, the majority of quantitative research in educational psychology fails to consider the pervasive racism students in K–12 education experience (DeCuir-Gunby & Schutz, 2014; Strunk & Andrzejewski, 2023). When racism is not explicitly considered, researchers, and by extension, the policy-makers and practitioners who rely on their findings, often associate racial disparities with individual or cultural deficits rather than structural conditions. This misinterpretation leads to statistical modeling by researchers that appears neutral when in reality it reproduces racial inequities by reinforcing existing unjust structures and systems. Thus, the omission of racism results in flawed methodology and risks causing further harm by legitimizing and perpetuating existing structural inequities, which further demonstrates a need for measures of racism, not only race.

Education policy research aimed at eliminating racism in the educational environment necessitates methodological innovation that fosters both racial equity-centered approaches and robust empirical analysis of the systemic and structural nature of racism. Mitigating and eliminating racism will be a multifaceted effort. One approach is to move away from the status quo of using racial and ethnic identity as a control for racism and instead directly attempt to estimate the impact of racism (Castillo & Strunk, 2024). Reviews of quantitative measures of racism exist in other sectors, such as health (Atkins, 2014; Furtado et al., 2023; Groos et al., 2018; Hardeman et al., 2022) and psychology (Schooley et al., 2019), but none to date have been completed in education. Though education researchers can and should use these existing measures (when appropriate) to understand racism, many of the measures do not ask questions explicitly related to experiences in schools. This void leaves the field lacking comprehensive measures to accurately assess the racism students encounter in educational environments.

Because operationalization determines *what* is measured and *what remains invisible* (regardless of whether it is a qualitative or quantitative measure), any measure of

racism in schools requires conceptual clarity on how racism manifests at the individual level and within the schooling systems. Different measurement choices will illuminate or obscure dynamics at play at different levels (e.g., classroom norms, school norms, district norms, student–student, teacher–student, etc.). This study therefore examines not only what measures exist, but also how their operationalizations reflect (or fail to reflect) contemporary understandings of racism’s multilevel nature.

This review directly addresses these gaps with the goal of identifying and advancing tools for researchers to use in their work and advocating for more targeted resources and reparations. Specifically, research questions include: 1) What are features of the survey measures, both descriptive and psychometric, used to measure racism in education? 2) What was the purpose of using racism-related surveys, and who was the intended population? 3) How do these measures operationalize racism and take into account (or fail to account for) the multiple contexts of education?

Significance

The murders of George Floyd, Breonna Taylor, and Ahmaud Arbery sparked a resurgence of the Black Lives Matter (BLM) movement during the global pandemic (Kolluri et al., 2023). Their unjust deaths brought race and racism to the forefront of American consciousness, along with increased acknowledgment that racism and white¹ supremacy were still alive and well. In the summer of 2020, polling found that 76% of Americans stated racism was a problem compared to 53% in 2015 (Monmouth University, 2020). Yet, America’s pandemic-era racial reckoning was short-lived. A year later, 90% of the \$50 billion pledged from corporations to racial equity organizations and causes had yet to actually materialize (Jung, 2021), and public support for the BLM movement dropped 20 percentage points, returning to prepandemic levels (Chudy & Jefferson, 2021). This rapid shift underscores the urgency of developing better tools to measure racism: failure to do so enables denial of the existence of racism and hinders advocacy for resources aimed at eliminating racism. Furthermore, if quantitative researchers are not encouraged to measure racism, the field risks maintaining the status quo by studying “race” without ever examining the impact of racism on racial disparities in education (Harper, 2012).

Common adages say, “what gets measured, gets managed” and “what gets measured, gets done.” If we measure racism, we can attempt to: 1) manage and address it effectively with data-informed interventions and policies; 2) track our progress as a society in combating it; and 3) raise awareness in those who may not believe they are directly affected (positively or negatively) by it. An example of the impact of measurement on contextual and relational dimensions of schooling, such as belonging and self-efficacy, can be seen in the expansion of social–emotional learning (SEL) assessment over the last decade (CASEL, n.d.). Though practitioners and communities knew from their cultural intuition (Bernal, 2016) that social–emotional learning (SEL) mattered for students’ well-being and academic success, it was not until researchers developed reliable and valid measures of SEL-related constructs that they were able to systematically demonstrate the effect of SEL on student outcomes (See Cipriano et al.,

2023; Durlak et al., 2011, 2022) and cost savings (Belfield et al., 2015). Today at least 19 large districts, including Austin, Atlanta, Boston, Baltimore, Chicago, Cleveland, Denver, Minneapolis, Nashville, Oakland, Sacramento, and Tulsa, now track SEL indicators (Committee for Children, 2019). The American Teacher and School Leader Panel Survey found that in the 2023–24 school year, 83% of districts used an SEL program or curriculum, up from 66% two years prior (Skoog-Hoffman et al., 2024). And in 2019, Congress passed a bill allocating 123 million in federal funding for SEL (Stringer, 2019).

At the same time, recent scholarship has raised important concerns about the heterogeneity of SEL programs across racial and ethnic groups and the extent to which many measures sufficiently attend to race (Jones et al., 2026). These debates exemplify that measurement does not merely document phenomena. Measurement can shape how constructs are defined, whose experiences are centered, and how findings are used. As Ho (2022) argues in specifying “who uses which scores for what purpose,” educational measurement always carries social consequences that depend on its use and interpretation. In this sense, measurement can function both as a bridge, translating practitioner insight into policy investment, and as a boundary, reflecting the assumptions embedded within design. This dual role motivates our examination of how racism is conceptualized, operationalized, and reported in contemporary educational measurement.

A similar case can be made for the measurement of racism: When constructs are defined, operationalized, and measured, they become visible and actionable for policymakers and practitioners. Nonetheless, given the current political climate and recent executive orders restricting funding for diversity, equity, and inclusion research, government support to measure racism in schools will be an uphill challenge. Yet, some states have refused to back down. Arizona, California, Connecticut, Delaware, Hawaii, Maine, Maryland, Minnesota, Nevada, New Jersey, New York, Oregon, Rhode Island, and Vermont have all signed a letter by their attorneys general expressing their support for Diversity, Equity, and Inclusion (DEI) employment initiatives (Massachusetts Attorney General’s Office, 2025). In our current context, partnerships with supportive states, community-based organizations, and private philanthropy will be essential as they can advance racially oriented measurement and data collection efforts that are insulated from federal political constraints. In this paper, we document the prevalence of these efforts from the individual to larger societal/governmental levels within the bioecological systems theory (Bronfenbrenner, 1977) to provide opportunities wherever efforts can take root.

In advocating for the *quantitative* measurement of racism, it is important to first acknowledge and reckon with the eugenic roots of statistics and psychometrics methods. Frequently used statistical concepts, such as the *F*-test, regression, and correlation, were developed by known eugenicists Ronald Fisher, Francis Galton, and Karl Pearson, respectively, to give scientific backing to racist concepts of the genetic superiority of white people and the intellectual inferiority of Black people (Evans, 2020; Weiss, 2010; Zuberi, 2001). Pioneers in the field of psychometrics were also known eugenicists. Edward Thorndike and Robert Yerkes, creators of early intelligence and

mass testing, asserted that innate differences in intelligence were directly related to racial hierarchy. Yerkes developed the Army Alpha aptitude test that was used specifically to exclude Black men from officer rankings in the military (Castillo & Strunk, 2024). That same framework was used to design the SAT (formerly known as the Scholastic Aptitude Test), which is still used today as a college entrance exam (Lemann, 1999). “Racism is, so to speak, baked into the bread” (Castillo & Strunk, 2024, p. 3) because performing well on many aptitude tests (including the SAT) and other standardized assessments relies on understanding vocabulary and culturally specific associations that reflect predominantly white, middle-class norms (Au, 2020; Gillborn, 2016; Modaffari & Jimenez, 2021). It logically follows that racially minoritized students score poorly on assessments that do not reflect their background and culture, reinforcing a deficit narrative about their abilities that detracts stakeholders from addressing the cumulative impacts of systemic and structural racism (Harper & Davis, 2012; Howard et al., 2016; Noguera, 2003; Randall, 2021). These tests’ racist legacies continue to permeate modern educational practices and assessments, to the detriment of progress and to sustain the status quo (white supremacy).

The roots of quantitative methods are not just historical. Although eugenics fell out of favor post–World War II, eugenicist statisticians continued publishing racist ideas in scholarly research throughout the 60s and 70s (c.f. Cattell, 1965; Jencks, 1972; Terman, 1961). As recently as the ’90s, Charles Murray and Herrnstein published *The Bell Curve* (1994), attributing racial and gender differences to genetics, not systemic discrimination. These methods are still deployed today to justify and explain societal stratification, rationalize under-funding of schools for racially minoritized students, and place the blame on communities of Color for educational and other life outcome disparities (Allen et al., 2008; P. L. Carter, 2005; Strunk & Andrzejewski, 2023). For instance, the frequently repeated (but false) claim that racially minoritized students are less motivated or lack work ethic compared to white students is a socially created rather than a biological explanation for educational disparities. This argument persists in some educational research, despite the fact that it has been thoroughly debunked by rigorous, racially conscious scholars (e.g., Castillo, 2020; Harper & Davis, 2012), further emphasizing the need for changes in quantitative methods.

Many frameworks have attempted to “flip the script” and use data for justice, including data feminism (D’Ignazio & Klein, 2020), QuantCrit (Gillborn et al., 2018), and critical quantitative inquiry (Stage, 2007). This study is primarily driven by three tenets of QuantCrit. This study first and foremost applies the first tenet of QuantCrit by centering on the realities of racism and advocating for its measurement (Gillborn et al., 2018). Aligned with another tenet of QuantCrit, “Data is not neutral,” we acknowledge that data and assessments created by mostly white researchers have been used to legitimize racial inequities (National Center for Education Statistics, n.d.). The last tenet of QuantCrit, “a social justice and equity orientation,” is also applied, as this study’s purpose is to both uplift tools and call for the creation of new tools to more accurately measure racism with the intention of using data to advocate for resources and reparations for oppressed communities. Although we apply QuantCrit, our study relied on traditional systematic review methods (for article and scale selection) to

guide the process. In the implications section, we outline what a QuantCrit-informed systematic review would entail and discuss possibilities for strengthening future critical quantitative methodologies.

A promising practice of QuantCrit is to provide a positionality statement wherein all authors explore their positions in society and potential biases (Boveda et al., 2023; Castillo & Strunk, 2024; Fong & Irizaary, 2025; Randall, 2021). For brevity, and given space constraints, we have provided a team positionality statement:

We are a racially diverse team composed of two Latinas, two Black women, one Latino, and one white woman. We hold many intersectional identities including but not limited to being a member of the LGBTQ community, 1st or 2nd generation immigrants in this country, parents, members of the working/lower class, English Learners, disabled, and first-generation college students. Everyone on the team acknowledges that they have been positively (white author) and negatively (authors of Color) impacted by racism in their schooling. Specifically, team members can recall teachers holding lower expectations for students of Color, tracking, and/or attending segregated, dilapidated public schools. All of us have been trained in traditional statistics and/or psychometric methods, and most at elite universities. Every day we work to unlearn these approaches and relearn statistics with an eye for racial equity. Many, but not all of us, continue to be affiliated with prestigious universities. We recognize the privileges and credibility that are automatically assumed because of where we trained and who we trained with. We pledge to use this amplified credibility for justice for Black and Brown communities. We also are upfront and transparent about using our cultural intuition (Pérez Huber et al., 2018), which stems from our backgrounds and experiences as a source of knowledge to contribute to this study.

Theoretical Framework

According to Bronfenbrenner's (1977) ecology of human development, humans interact with experiences within their environments to grow and progress. Said differently, individuals exist within systems with which they interact. These systems are in turn nested within each other. Bronfenbrenner, in his bioecological systems theory, proposes four levels, moving outward: microsystems, mesosystems (spans between microsystems), exosystems, and macrosystems. In later conceptions of the model, Bronfenbrenner stressed that these systems change throughout the lifespan and depend on a person's personal characteristics, such as race and gender (Bronfenbrenner, 1988; Rosa & Tudge, 2013), yet still did not explicitly consider racism.

Bronfenbrenner's (1977) bioecological systems theory has been criticized by racial equity scholars for its neutral stance on race and failing to take into account the interlocking systems of racism that individuals experience (Crawford et al., 2020; Paat, 2013; Skinner-Dorkenoo et al., 2023; Stern et al., 2022), particularly in the ways Black and Brown individuals experience overlapping systems of overt and covert layers of oppression in both racial, gender, and other types of discrimination. Recently, racial equity scholars have *reclaimed* the theory by explicitly reconceptualizing Bronfenbrenner's model to study race and racism; the adapted model provides an apt framework for understanding racism as it occurs at each of these levels and is perceived by an individual differently throughout their lifetime. For example, Stern et al.

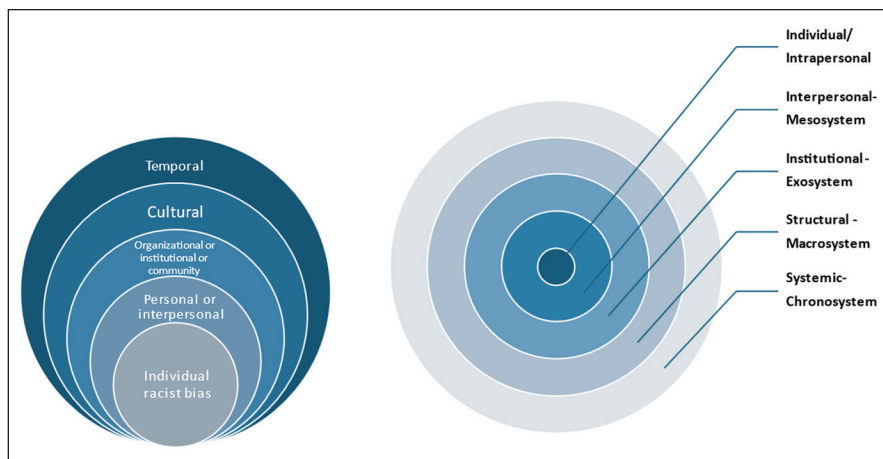


Figure 1. Adaptation of Skinner-Dorkenoo and Bronfenbrenner model.

(2022) applied it to study Black youth, Crawford et al. (2020) to study underrepresented populations in gifted education, and Paat (2013) to study immigrant families. Each example provided the team with historical context for marginalized populations, added examples of racism in each level of the Bronfenbrenner model, and thus supported the definitions used for coding (see coding section).

We follow previous scholarship and also adapt Bronfenbrenner's model to understand the measurement of racism (Crawford et al., 2020; Paat, 2013; Peterson, 2014; Skinner-Dorkenoo et al., 2023; Stern et al., 2022). We closely follow Skinner-Dorkenoo et al.'s (2023) contemporary and relevant approach to using models of nested influence to understand race relations (Bronfenbrenner, 1979; Peterson, 2014). In our study, racism can be behaviors (e.g., discrimination), beliefs (e.g., perceptions, attitudes, cultural norms), or structures/policies (e.g., access to high-speed internet, school-district policy, federal policy, historical legacy). The specific form it takes depends on the microsystem in which it occurs. We have adapted Skinner-Dorkenoo et al.'s model of Bronfenbrenner's (1979) theory as visualized in Figure 1, with Skinner-Dorkenoo's on the left and this study's adaptation on the right.

Defining Levels of Racism

We define the multiple levels of racism following Banaji et al.'s (2021) definition, which builds on Jones's (1971) and Fiske's (1993) prior work. The levels used by Banaji et al. (2021) also map onto Bronfenbrenner's and Skinner-Dorkenoo et al.'s (2023) model. These definitions and adaptations provided examples and guidance across various educational contexts for the coding aspect of the project (see Method).

Like Banaji et al. (2021), we define systemic racism as the most encompassing and expansive form of racism that permeates our society. Systemic racism, analogous to

Bronfenbrenner's chronosystem and Skinner-Dorkenoo's (2023) temporal level, captures the social norms and policies that shape how the education system produces unequal outcomes. Systemic racism focuses on how changes over time (both personal and societal) affect development. It considers the timing, duration, and sequence of these changes.

The next level is structural racism, which is also part of Banaji et al.'s (2021) definition, equivalent to Bronfenbrenner's macrosystem and Skinner-Dorkenoo et al.'s (2023) cultural level. This type of racism occurs when interconnected systems work together over time to produce racial inequities. Structural racism does not relate to a person's specific environment; rather, it revolves around the cultural norms and guidelines that shape smaller settings. These guidelines may be formal, such as the rules that dictate the number of students and resources available in public school classrooms. These guidelines may also be informal or unwritten, such as society's expectations regarding discipline or performance. At this level, racism exists as both beliefs and structures. While systemic and structural are considered similar and at times overlapping concepts, some scholars view them as distinct (Hicken et al., 2018).

Like Banaji et al. (2021) and Skinner-Dorkenoo et al. (2023), we call the next level institutional racism, akin to Bronfenbrenner's exosystem. This type of racism occurs at the community level and varies for subgroups in society based on race. At this level, racism manifests as structural or local policies, practices, or climates. In education, this could be at the school or district level.

Analogous to Bronfenbrenner's mesosystem, we use the term interpersonal racism to refer to racism that is relational and enacted between individuals. Beyond a single person, Bronfenbrenner (1977) called the mesosystem "a system of microsystems" (p. 515), referring to sets of interactions between people and the immediate settings at a particular age. For example, students interact with their families or guardians, while said families or guardians also have relationships with school staff. At this level, racism is a behavior. Skinner-Dorkenoo et al. (2023) refer to this level as the organizational/institutional/community level, which involves larger structures that directly impact students. Banaji et al. (2021) refer to this type of racism as everyday interaction patterns, describing habits and norms. Such larger structures may include government agencies, the media, social services, and extended social networks. Specifically, in everyday schooling, it is the internet services or transportation infrastructure in the student's environment that would impact the student's ability to interact with the smaller systems of school or work.

We define the most micro-level, the individual level, as intrapersonal racism, which is a person's internal perceptions of self and/or groups based on race or ethnicity. At this level, racism is a belief. It is equivalent to Banaji et al.'s (2021) definition of the individual and mental structures level that encompasses learning, memory, attitudes, beliefs, and values. Interpersonal racism maps onto Skinner-Dorkenoo et al.'s (2023) individual racist bias and Bronfenbrenner's microsystem level, which involves an individual's immediate surroundings, such as the home, work, or educational setting.

Literature on Prior Measurement of Racism

The direct measurement of racism is critical to understanding its far-reaching impacts on individuals and society. Historically, researchers have often relied on proxy variables, such as race, to infer the presence of racism, interpreting differences in outcomes between racial groups as potentially driven by individual, structural, or systemic racism (Castillo & Gillborn, 2022). However, this approach has limitations. As a quantitative variable in education, race is commonly self-reported based on the five common census categories (white, Black, Asian, Hawaiian Native/Pacific Islander, Native American/Alaskan Native) (D. J. Baker et al., 2024). Relying solely on race to stand in for racism oversimplifies the nuanced and dynamic ways racism operates. Racism is not experienced uniformly by all members of a racial group; rather, it is a fluid and complex system that manifests differently depending on the context, moment, and individual (Bonilla-Silva, 2021). This complexity highlights the necessity for researchers to move toward more direct measures of racism that can capture its variability and provide a clearer understanding of how it shapes life chances and outcomes.

Survey-based measures remain one of the most prevalent methods for directly assessing individual experiences of racism, particularly among people of Color. These measures aim to capture both overt and subtle forms of discrimination and, therefore, are the focus of our review. For instance, the Adolescent Discrimination Distress Scale includes items such as “You were discouraged from joining an advanced level class,” focusing on more explicit forms of racial discrimination in schooling (Fisher et al., 2000). Similarly, the Perceived Racism Scale includes questions about experiences like exclusion from study groups or discomfort in the classroom based on race (McNeilly et al., 1996). These instruments have been foundational in capturing the everyday realities of racial discrimination, though they are not without limitations. The challenge lies in the fact that racism is constantly evolving, and these measures must be regularly reassessed to ensure they remain relevant and accurate in capturing contemporary forms of racism (Domínguez & Embrick, 2020).

The measurement of microaggressions, a more subtle form of individual-level racism, has also gained traction in recent years. Microaggressions refer to slight, often unconscious comments or behaviors that convey negative racial attitudes (Sue et al., 2007). Instruments like the Racial and Ethnic Microaggressions Scale aim to capture these daily slights, with items such as “Someone assumed that I would have a lower education because of my race” and “Someone avoided eye contact with me because of my race” (Nadal, 2011). However, measuring microaggressions presents distinct challenges. Microaggressions are often subtle and ambiguous, making it difficult for individuals to definitively label their experiences as racially motivated, a phenomenon referred to as racial gaslighting (Sue et al., 2007). This ambiguity can result in underreporting or misinterpretation of experiences, complicating the survey process. Thus, while scales measuring microaggressions provide valuable insights into everyday racism, they also require careful consideration of the subtlety of these interactions

(Domínguez & Embrick, 2020). We therefore include measures that look beyond individuals and extend to institutions and structures to understand the strengths and limitations of measurement of racism at all levels.

We acknowledge the limitations associated with surveys based on self-reports, particularly to measure racist beliefs and attitudes (West et al., 2016). White respondents often deny harboring racist beliefs and/or, due to social desirability bias. Survey takers (students, teachers, parents) may respond to an item in a manner that they believe will please or avoid hurting others, making self-report measures unreliable (Bonilla-Silva, 2021; Paulhus, 2002). Self-report scales are also vulnerable to reference bias, which is the idea that students respond in reference to their classmates, school, peer group, family, and/or community (West et al., 2016). For example, one among many potential sources of reference bias is differences in the characteristics of the schools that have high concentrations of racially minoritized students compared to those that have low concentrations (Reardon et al., 2013; West et al., 2016). Majority white students are more likely to attend high-resourced schools (Reardon et al., 2013) and, as a result, students from varying racial groups will experience different points of reference for gauging what it means to “have teachers who look like you” or “feel like you belong.”

Overall, while survey-based measures are not without challenges, they also provide an essential framework for directly measuring racism and have advantages over other types of quantitative measurement. For large-scale data collection, researchers and practitioners can quickly, easily, and cheaply administer self-report scales to thousands of students (West et al., 2016). Additionally, self-report scales can be tailored to ask questions about specific experiences, while an observation tool can capture only what the researcher observed at one point in time (West et al., 2016). The complexity of racism, especially as it morphs over time and operates differently at individual, interpersonal, institutional, systemic, and structural levels, necessitates ongoing evaluation of existing tools and the development of new, more nuanced measures. Researchers must carefully and continually refine their instruments to ensure they align with contemporary understandings of racism, recognizing that the ways racism is experienced and expressed are subject to ongoing change (Sue et al., 2007; D. R. Williams et al., 1997).

Methods

Systematic reviews can place a concept like the measurement of racism in education “into the broader framework of research areas in a scholarly discipline, [build] upon the body of knowledge in that area and [identify] gaps in the literature” (Zawacki-Richter, 2020, p. v). To answer our research questions, we followed systematic review protocols and practices within educational research (e.g., Polanin et al., 2019). We also incorporated emerging systematic review guidelines on centering DEI in synthesis work, such as creating research questions that challenge the status quo and center specific groups, forming a team that reflects the populations of focus in the studies, encouraging full participation of team members in refining the training to reduce power dynamics and create clarity, and extracting data regarding specific subgroups in the studies (i.e., racial and ethnic proportions) (Polanin et al., 2024).

Table 1. List of Search Terms.

race OR racial OR racism OR discriminat* OR prejudic*
interpersonal OR intrapersonal OR systemic OR institutional OR Internalized OR structural OR belief*
educat* OR K-12 OR higher education OR college OR postsecondary OR student* OR teacher* OR instruct* OR professor* OR educator* OR primary OR elementary OR secondary OR “middle school” OR youth OR young OR child*
survey* OR outcome* OR instrument* OR scale* OR tool* OR validation OR reliability OR measur* OR questionnaire OR assess*

Search

We searched for peer-reviewed articles published 2003–2023 because, as previously stated, it is important to use the most up-to-date measures as society is constantly shifting and cultural realities are shaping how racism operates and the specific ways it impacts marginalized communities. We searched within the academic databases of ERIC, Education Research Complete, PsycInfo, and PsycArticles. We searched article abstracts for key terms of quantitative measures (e.g., “survey”) on race and racism (e.g., “prejudice”) within systems and structures (e.g., “institutional”) specifically in K–12 educational settings (e.g., “middle school”). Our initial results revealed an overrepresentation of high school and elementary school measures. To ensure that middle schools were adequately represented, we explicitly added this term to our search. See Table 1 for initial search terms. To reduce the number of irrelevant articles, we also chose NOT to include settings outside of the United States (e.g., “Canada”) and in irrelevant contexts (e.g., “dental”) that appeared in initial searches.

Inclusion Criteria and Screening. To be included in the review, the study had to include a quantitative measure whether it was from a quantitative or mixed-methods study. Measures for the purposes of this study included questionnaires, surveys, subscales, or individual closed-ended items that were completed by students, parents, teachers, or other school staff. We excluded experimental tools (e.g., picture judgments), proxy variables created to measure racism in other ways (e.g., housing discrimination), and open-ended survey items because these comparisons would require different analytic approaches and would not allow for a meaningful, “apples-to-apples” analysis. Pre-K and early childhood articles and articles that did not include data collection in the United States were excluded. Higher education articles were only included if the focus was on pre-service or in-service educators or educational staff (e.g., school counselors) who we assumed would soon work in K–12 settings.

Articles were excluded if they measured racism generally (including discrimination based on race/ethnicity) and did not explicitly connect racism to K–12 educational settings (e.g., diversity not explicitly about race, discrimination occurring in non-school settings, culturally responsive pedagogy scales that do not mention race). Because racism unfolds and operates distinctly based on location and history, we also

excluded studies conducted outside the United States. To be included, studies needed to contain at least one item addressing institutional or structural racism, schooling experiences of racism, perceptions of racism in school, or racist attitudes and beliefs about schooling. When reviewing scales, items that referenced racism but did not specify educational racism were excluded from our item-level analysis, although the full scales were still retained. Examples of excluded items included questions to parents about their beliefs regarding racism in society, items capturing trauma or discrimination unrelated to educational experiences, and health-focused measures for school-age children that assessed discrimination broadly rather than within school contexts. In addition, two commonly used measures, the Multigroup Ethnic Identity Measure–Revised (Phinney & Ong, 2007) and the Multidimensional Inventory of Black Identity (Sellers et al., 2006), were intentionally excluded because their items address racial identity or racism generally rather than racism in schooling.

We included other forms of educational discrimination, namely characteristics that racialized ethnic individuals experience, such as discrimination based on language, immigration status, and religion. While not all races and ethnicities experience racism due to their language, religion, and/or immigration status, many researchers have noted high rates of discrimination based on these characteristics. Thus, we sought to be as inclusive as possible in understanding how racism might be measured (Cano et al., 2021; Cobb et al., 2019; Gong et al., 2016; Sugarman et al., 2018).

We exported the initial search to Rayyan, an online collaborative systematic review site, for removing duplicates and screening. Rayyan has been increasingly popular in systematic reviews because of the search innovation features that allow screeners to “triage” articles more quickly, such as search bars and highlights of key inclusion/exclusion criteria (Harrison et al., 2020; Olofsson et al., 2017). Three of the six authors screened the studies for inclusion. During both title/abstract and full-text screening, to improve screening reliability, team members screened 10 random articles at a time and conferred on the inclusion/exclusion of articles following each batch until agreement reached 90%. After reaching 90%, the screeners independently screened and marked “maybe” or “unsure” articles for discussion in screening meetings. All “unsure” articles were discussed by the screening team until full consensus was achieved.

Data Extraction. With the final set of articles, we extracted bibliographic information, sample information (e.g., respondent type [student, parent, teacher, other school staff, other], racial groups), and measure information (e.g., name, year created, number of items, number of scales). If authors did not provide the year that the scale, subscale, or item was created, we inferred it was the year the data was collected; if no information was included about the year the data was collected, we used the year of publication as an approximation. We noted the extent to which authors of the articles included the items (none, some, all). If some or none of the items were available in the article and if the authors cited another article for the scale they used, we located the original article and searched for the items. Also, if authors cited other scales that fit our criteria and were not already in our review, we located those articles to screen the full text and determine inclusion. Articles used for extracting item details but not included as part

of the full analysis included Augoustinos and Rosewarne (2001), Bogardus (1933), Brigham (1993), Devine (1989), J. A. Eccles (1997), Fisher et al. (2000), Kulis et al. (2009), Landrine and Klonoff (1996), Neville et al. (2000), Nadal (2011), Vinokurov et al. (2002), Pohan and Aguilar (1994, 2001), Phinney et al. (1998), Seaton et al. (2008), Utsey et al. (2013) and Torres et al. (2010). Lastly, we selected a random 10% of the articles to check references (i.e., backward citation tracking) and add additional articles that met our inclusion criteria. See Figure 2 for the full inclusion process.

Analysis

Features and Description of Measures

Using extracted data, the team used quantitative and qualitative analytical methods to answer the research questions. We calculated descriptive statistics to understand the most frequent measures and their features, such as who completed the measure (parent, teacher, student, or other school staff), demographics of students, setting, and year it was developed versus year it was used. We also calculated descriptive statistics to understand the proportion of scales that aligned within each area of the bioecological systems theory to understand where focus was concentrated in the field. Many scales were used multiple times (across multiple articles), which complicates the descriptive statistics regarding the scale characteristics. To focus on the *scale* characteristics rather than individual *study* characteristics, we calculated the characteristics below based on scales. For example, when three studies used the same scale, we counted the unique characteristics of participant, age, etc., only once. Additionally, coding allowed “select all” for scale characteristics.

In the first coding cycle, we qualitatively coded each of the measures using attribute coding (Saldaña, 2018). Attribute codes included why authors used the scale and the types of racism present in each of the scales and whether they are part of individual beliefs and perceptions (intrapersonal), microsystem (interpersonal), exosystem including course curricula (institutional), macrosystem such as larger cultural stereotypes (systemic), or chronosystem such as racism through the intersection of systems in alignment with bioecological theory adapted for discussions of racism (e.g., Peterson, 2014; Race Forward, 2023; Stern et al., 2022). We did not include anti-racism as it reflects another construct, a proactive charge to undo educational racism at any of the levels (e.g., B. Alvarez, 2024; Smithsonian, n.d.). This was beyond the scope of this project but should be reviewed in future research and will be a follow-up study of this current project.

Some forms of racism can map onto multiple levels of the theoretical model, as systems often overlap. Therefore, scales could receive more than one coding label (e.g., “interpersonal and institutional”). If authors specified they attempted to measure a type of racism, we used the study authors’ classification. If it differed from our coding, we had multiple labels for the scale. For example, Oberoi and Trickett (2018) attempted to measure “Muslim School Discrimination.” Some of the questions presented scenarios about school experiences, such as “A school administrator treated

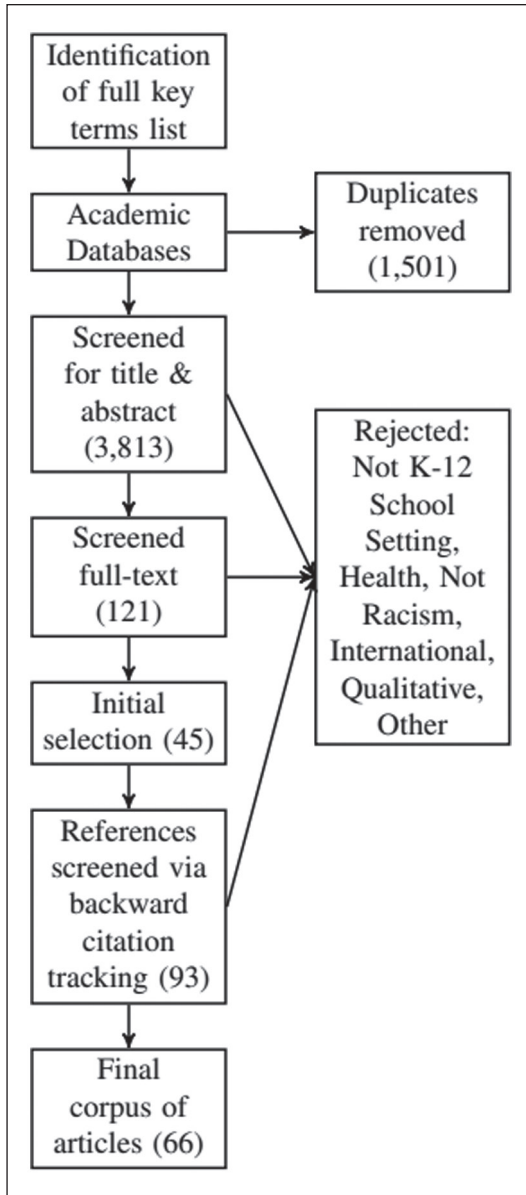


Figure 2. Inclusion and exclusion process.

you unfairly because you are Muslim” and “You got in trouble in school because you did not understand how the school rules work,” which we labeled as attempting to measure institutional racism. However, others were presented, such as “An American

student treated you in some way because you are Muslim,” which we labeled as interpersonal racism because it occurred during student-to-student interactions. There were also other statements about teacher-to-student discrimination, like “A teacher treated you unfairly because you are Muslim,” which could be coded as both interpersonal and potentially institutional racism. Some teachers, particularly in smaller schools, had more autonomy to dictate students’ disciplinary actions, which would take place in the microsystem. Still, most disciplinary actions happened at the school or district leadership level, which would place them in the exosystem (i.e., institution). Because these items cut across multiple ecological levels, this scale, and others with similar overlap, received all applicable coding labels.

If a scale had a question about a student’s perceptions/awareness, we labeled these scales as individual/intrapersonal rather than systemic because they were happening within the person’s thoughts and did not directly measure systemic racism. For example, Seider et al. (2018) used the Awareness of Structural Racism scale to ask students about statements such as “Racism in the educational system limits the success of Blacks, Latinos, and other racial minorities.” We concluded that this question did not measure structural racism but measured a student’s perception and awareness of it. We applied this concept of measuring perceived racism or attitudes/beliefs/endorsement of racism as intrapersonal.

We found only two surveys that attempted to measure structural racism per our definition of racism through the intersection of multiple systems and institutions. Bañales et al.’s (2019) Critical Reflection Achievement Gap Attributions scale attempted to measure structural racism with the following prompt: “There are many theories about why Black students do less well than Whites in school. Indicate how much you think these differences are caused by each of the following: Individual attributions, structural attributions.” Although the scale references structural explanations, we *also* coded it as intrapersonal because these items ultimately assess a respondent’s beliefs and attributions about the causes of the gap rather than directly measuring structural racism itself. The other instance we found of structural racism was in Landrine and Klonoff’s (1996) Schedule of Racist Events, where they asked “How many times have you been treated unfairly by institutions (schools, universities, law firms, the police, the courts, the Department of Social Services, the Unemployment Office, and others) because you are Black?” Although this item references institutional and structural domains, the scale also includes interpersonal items, such as “How many times have you been treated unfairly by teachers and professors because you are Black?” Because unfair treatment by teachers can reflect both interpersonal racism (individual educator bias) and institutional racism (school or district policies that shape disciplinary or evaluative practices), we coded this scale as interpersonal, institutional, and structural.

See Table 2 for example codes across each level. To ensure adequate reliability between coders, the coding team collectively reviewed the extraction form and coded five articles together for practice before refining the extraction form. The coding team then independently coded five articles at a time, engaged in coding discussions, and came to an agreement on extraction and codes. Coding meetings occurred as needed to discuss coding issues, and all disagreements were resolved via 100% consensus.

Table 2. Example Racism Codes in Each Level of Bioecological Systems Theory.

Ecological Level	Racism Code	Team Definition	Examples
Chrono system	Systemic	Societal-level, broad, cultural structures that shape racial and ethnic communities	• Unequal school funding • Lack of access to clean water at schools or heat island impacts • Racial achievement/opportunity gaps • Racially segregated schools
Macro system	Structural	Multiple interconnected parts (systems), all reinforcing inequality together	• Redlining leading to unequal access to high-quality schooling • Environmental racism • Parents' access to quality education, housing, or employment limits their ability to support their child's schooling • Limited transportation and healthcare access leads to chronic student absenteeism • Over-policing in neighborhoods leads to stress and criminalization of youth in schools
Exo system	Institutional	Community level; racism that occurs within subgroups in societies	• District-level policies on hair and clothing that disproportionately affect racially minoritized students • Tracking by schools/districts (e.g. discriminatory access/placement to advanced coursework) • Within-school segregation • Approved school or district curricula that avoid teaching diverse histories and cultures • Disproportionate referrals to special education or gifted and talented programming • Underrepresentation of Black and Brown teachers and school administration • Unfair disciplinary actions/policies
Micro system	Interpersonal	Relational; racism enacted between individuals	• Discrimination by peers or school staff • Low expectations for racially minoritized students • Discipline referral bias and in-class individual discipline bias • Teachers not providing accurate information on race/religion/history/etc. • Racial/ethnic bias attitudes and stereotypes of racially minoritized students
Individual	Intrapersonal	Individual; internal perceptions of self and groups based on race or ethnicity	

Psychometric Examination

We employ the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 2014) to assess psychometric properties. Primarily, we focus on validity, which the *Standards* define as “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (AERA, APA, & NCME, 2014). The *Standards* define sources for validity evidence as (1) test content, (2) response processes, (3) internal structure, (4) relation to other variables, and (5) consequences of testing. We will emphasize numbers (3) and (4), as these are often the most used sources of evidence when designing an instrument. The evidence from test content will not be systematically evaluated, as it is contingent on the purpose the authors assign to each instrument. The exception in which content will be reviewed is when authors assess dimensionality. Authors should provide qualitative evidence or draw from theory if their instrument is unidimensional or multidimensional. Nevertheless, we present qualitative codes for the type(s) of racism each scale is assessing (see Appendix Table A.1). Similarly, many scales do not provide evidence for (2) response processes and (5) consequences of testing. Furthermore, evidence for the consequences of testing generally requires causal relationships, which are frequently out of scope for many initial validation studies.

One form of internal consistency we reviewed was reliability, which refers to the consistency, stability, and repeatability of scores across different contexts (Cronbach, 1951). When examining reliability, we expected to see at least two estimates and not complete reliance on one Cronbach’s alpha, which has been documented as problematic (Cortina, 1993; Yang & Green, 2011). These estimates could be item-total correlations, Omega, Kuder-Richardson Formula 20 (KR-20), test-retest correlations, intraclass correlations, generalizability coefficients, or another critical/new estimate. Authors that do not indicate the magnitude they expected to achieve will be evaluated against the conventional approach of categorizing values greater than .70 as acceptable (Dunn et al., 2014; Nunnally, 1978).

For validity, we used both construct validity and convergent validity. Scales that are described to measure a multidimensional attribute should provide dimensionality reduction estimates from either a confirmatory factor analysis (CFA) or an exploratory factor analysis (EFA). For CFA approaches, the fit was assessed through the root mean square error of approximation (RMSEA), the standardized root means square residual (SRMR), the comparative fit index (CFI), and the Tucker-Lewis index (TLI). An RMSEA $< .06$ indicates an excellent fit, and an SRMR $< .08$ indicates a good fit to the data (Hu & Bentler, 1999). However, an RMSEA $< .08$ will be considered an adequate fit to the data (Browne & Cudeck, 1993). The CFI and the TLI should be over .90 for an adequate fit and over .95 for an excellent fit (Bentler & Bonett, 1980; Schumacker & Lomax, 2010). An acceptable fitting model is defined when three of these four statistics are at least adequate (Hu & Bentler, 1999). Otherwise, an EFA is expected to address content dimensionality. If authors opted to provide dimensionality estimates for

their instrument through a full-information Item-Response Theory factor-analytical approach, proper explication of the parameters should have been addressed. Authors need to explain the usage of a one- vs a two-parameter model whenever appropriate for dichotomous items, especially considering parsimony (F. B. Baker & Kim, 2004). For polytomous items, we looked for the employment of the Generalized Partial Credit Model, the Testlet Model, or the Graded Response Model, depending on the number of categories for each item and the aims described by the authors (F. B. Baker & Kim, 2004). Additionally, if authors aim to create population-invariant parameters, they should employ an IRT framework. This is because item-total correlations, CFA, and EFA are population-dependent (Novick, 1966). On the other hand, IRT posits that the properties of each item remain constant regardless of the sample, the time of measurement, or the presence of other items on the instrument (Brennan et al., 2006).

To evaluate the appropriateness of scales, we have designed and coded four categories incorporating the previously mentioned information. These categories are “R”=Recommendations Met, “M”=Meets Most Recommendations, “N”=Needs More Evidence to Meet Recommendations, and “E”=Evidence Absent. For a detailed description of these categories, see Table 3.

Results

Of the 3,813 imported references, 1,501 were duplicates, and 66 articles aligned with the inclusion criteria. Authors of these articles mostly used different scales; a total of 50 unique scales, subscales, or items were extracted from the articles, and five of the 66 articles used more than one measure. The following scales were the most commonly used (each by at least three articles) listed from highest to lowest frequency: Critical Consciousness Scale (CCS) (Diemer et al., 2017), Index of Race-Related Stress (Seaton, 2003), Everyday Discrimination Scale (Seaton et al., 2008), Schedule of Racist Events (Landrine & Klonoff, 1996), School Climate for Diversity Scale—Secondary (Byrd, 2017), and Color-Blind Racial Attitudes Scale (CoBRAS) (Neville et al., 2000). Table 4 lists all the scales/subscales, the type of coded racism (e.g., intra-personal, interpersonal, institutional, systemic or structural), author(s) who developed it, author(s) who applied it, and who it was designed for (student, parent, teacher, or other school staff). For more detailed information with relevant education-related items, see Appendix Table A.1 and A.2.

Descriptive Features of Measures on Racism

Of the 50 measures/scales/subscales/items that explicitly asked questions about racism in schools, only 11 explicitly operationalized racism and even fewer ($n=2$) met both psychometric recommendations (R) per our coding criteria for reliability and validity estimates. The scales used in this study were developed, on average, in 2008. However, the year of scale development ranged from 1972 to 2023. The mean year of scale usage was 2015, while usage ranged from 2001 to 2023. Most of the scales were used in student settings, with 58% of respondents being students. Teachers were the second largest

Table 3. Rubric for Assessing Measures.

	Recommendations Met (R)	Meets Most Recommendations (M)	Needs More Evidence to Meet Recommendations (N)	Evidence Absent (E)
Evidence for internal structure: reliability	Authors provide at least an estimate of reliability and are acceptably strong (e.g., Cronbach's alpha $>.70$ per subscale) or otherwise explicate the estimate they were hoping to achieve.	Authors provide at least one estimate to reliability, with this measure alpha $>.70$ per subscale) or otherwise explicate the estimate they were hoping to achieve.	Reliability estimates are unacceptable (e.g., Cronbach's alpha $<.70$ per subscale) and do not provide an explanation regarding the estimate authors were hoping to achieve.	Reliability estimates are absent.
Evidence for internal structure: dimensionality	If multidimensional: Authors explain with theory or qualitative evidence why the instrument should be multidimensional. The EFA is appropriate (uses smoothed tetrachoric correlations for polytomous items or uses an appropriate IRT model). The CFA fit statistics are appropriate: at least three out of the RMSEA, SRMR, CFI, TLI are adequate (Hu & Bentler, 1999). Qualitative evidence to test CFA is present. If unidimensional: Authors draw from theory or qualitative evidence to explain that the scale should be unidimensional. Authors provide IRT estimates if they wish to generalize scores to other populations and draw from theory to explain that the scale should be unidimensional.	If multidimensional: Authors do not explain with theory or qualitative evidence why the instrument should be multidimensional. Yet, authors perform an EFA or CFA. Authors perform either an EFA or CFA. The EFA is appropriate (uses smoothed tetrachoric correlations for polytomous items or uses an appropriate IRT model). The CFA fit statistics are appropriate: at least three out of the RMSEA, SRMR, CFI, TLI are adequate (Hu & Bentler, 1999). Qualitative evidence to test CFA is absent. If unidimensional: Authors draw from theory or qualitative evidence to explain that the scale should be unidimensional but do not provide IRT estimates if they wish to generalize scores to other populations.	If multidimensional: Authors do not explain with theory or qualitative evidence why the instrument should be multidimensional. Yet, authors perform an EFA or CFA. Authors perform either an EFA or CFA, with unacceptable results. The EFA is not appropriately employed (uses heuristic correlations for polytomous items or uses an inappropriate IRT model based on the number of parameters being estimated). The CFA fit statistics are unacceptable: less than three out of the RMSEA, SRMR, CFI, TLI are adequate (Hu & Bentler, 1999). Qualitative evidence to test factors is absent. If unidimensional: Authors do not draw from theory or qualitative evidence to explain that the scale should be unidimensional and do not provide IRT estimates if they wish to generalize scores to other populations.	Evidence for internal structure regarding dimensionality and drawing from theory or qualitative data is absent.
Evidence for relations to other variables	Authors provide estimates of convergent/discriminant validity evidence or test-criterion relationships evidence, with significant correlations ($p < .05$). Authors explain the appropriateness of the external variables selected to compute these estimates.	Authors provide estimates of convergent/discriminant validity evidence or test-criterion relationships evidence, with significant correlations ($p < .05$). Authors do not explain the appropriateness of the external variables selected to compute these estimates.	Authors provide estimates of convergent/discriminant validity evidence or test-criterion relationships evidence, with insignificant correlations ($p > .05$).	Evidence for relations to other variables is absent.

Table 4. Descriptive Statistics for Scale Characteristics.

Average year developed	2008
Average year used	2015
% Answered by students	58%
% Answered by parents	16%
% Answered by teacher	34%
% Answered by other school staff	2%
% Measured systemic racism	6%
% Measured structural racism	4%
% Measured institutional racism	44%
% Measured interpersonal racism	50%
% Measured individual/intrapersonal racism	64%
	<i>n</i> = 50

Table 5. Descriptive Statistics for Scale Characteristics of Student Scales.

% Demographic measured was mostly Asian	3.3%
% Demographic measured was mostly Black	53.3%
% Demographic measured was mostly Indigenous	0%
% Demographic measured was mostly Latine	33.3%
% Demographic measured was a mix	33.3%
% Demographic measured was mostly people of Color	10%
% Demographic measured was mostly white	33.3%
% Measured in urban areas	72.7%
% Measured in suburban areas	50%
% Measured in rural areas	20%
% Area not stated	23.3%
% Elementary age	16.7%
% Middle school age	56.7%
% High school age	66.7%
	<i>n</i> = 30

group to be surveyed at 34%. Both parents and other school staff were surveyed at a lower rate. Additionally, 64% of the scales assessed racism at the individual/intrapersonal level. Both interpersonal and institutional racism were measured at comparable rates, 50% and 44%, respectively. Finally, structural and systemic racism were the least captured, at just 4% and 6%, respectively. See Tables 4 and 5 for more descriptive details.

Of the 50 scales, 30 were used in student settings. Table 5 shows that just over half of the student sample consists primarily of Black students, while white, Latine, and mixed settings, with no single group forming a majority, are represented at similar rates. Indigenous and Asian populations were surveyed the least frequently, if at all. In addition, most of the sample consisted of students from urban areas, while students from rural areas were least represented. We know that different racialized groups, and

individuals for that matter, interpret racism differently (E. R. Carter & Murphy, 2015). Thus, the overrepresentation of Black students and the limited inclusion of Asian, Indigenous, and rural students raise concerns about whether existing measures capture within-group heterogeneity and whether perceptions of racism are being interpreted through measures normed on one group's lived experience.

Table 5 also shows that middle-school-aged and high-school-aged students are surveyed more frequently and at similar rates compared to elementary school students. Most scales that sampled elementary students included older students (e.g., 10 years old, upper elementary). Only two scales, Black/White Evaluative Trait Scale (Hughes et al., 2007) and the Social Distance Scale (A. D. Williams et al., 2023), asked younger elementary students about their attitudes toward individuals of varying races and ethnicities. Perhaps researchers intentionally select older students because they assume older students have the developmental awareness to identify racism in school. While teacher demographics are not specified in Table 5, most teachers that were surveyed identified as white, matching the national demographics of the teaching labor force (USAFacts Team, 2025). This pattern is not observed among parents, as their sample is more racially diverse.

Psychometric Evaluation

Overall, the review revealed substantial gaps in the reporting and quality of psychometric evidence across the included assessments. Generally, reliability estimates were either missing or entirely reliant on Cronbach's alpha, a statistic criticized for its misuse and misunderstanding (Cortina, 1993). More notable is that a fifth of all assessments did not report a single reliability estimate, and 14% of the time, the reliability estimates from the scores needed more evidence to align with our recommendations outlined in Table 3. While evidence for relations to other variables (correlational validity) was most infrequently classified as Evidence Absent (E), 72% of the assessments under consideration did not provide a single estimate. Similarly, we find that 70% of scales did not address dimensionality (i.e., factor analysis or mentioning the scale must be unidimensional). More so, 20% of authors estimating internal consistency via dimensionality reduction did not meet most of our outlined recommendations. Generally, 39 out of the 50 scales reviewed provided a Cronbach's alpha and three provided intraclass correlations. Exploratory factor analysis ($N=6$) and confirmatory factor analysis ($N=11$) were employed the most to assess dimensionality. External correlations ($N=11$) and structural equation models ($N=2$) were most often used to assess evidence for relations to other variables. The aforementioned descriptive results are in Table 6; for a detailed coding of each measure, see Table A.2 in the Appendix.

When findings are disaggregated by type of racism (individual/intrapersonal, institutional, interpersonal, structural, and systemic), reporting patterns are broadly consistent across the types (see Table 7). Across all conceptualizations, the most frequently observed classification was Meets Most Recommendations (M), whereas Recommendations Met (R) was less common across categories. For measures assessing institutional and interpersonal racism, most were coded as Meets Most

Table 6. Overview of Psychometric Evaluation of Measures.

	Evidence for internal structure: reliability	Evidence for internal structure: dimensionality	Evidence for relations to other variables
Quality Coded Classification	Percent		
Recommendations met (R)	4%	4%	22%
Meets most recommendations (M)	62%	20%	4%
Needs more evidence to meet recommendations (N)	14%	6%	2%
Evidence absent (E)	20%	70%	72%
N = 50			

Table 7. Overview of Psychometric Evaluation of Measures by Type of Racism.

Type of racism	E	N	M	R	Total*
Individual/Intrapersonal	19%	10%	65%	6%	31
Institutional	20%	10%	70%	0%	20
Interpersonal	16%	12%	72%	0%	25
Structural	0%	0%	100%	0%	2
Systemic	0%	0%	100%	0%	3

Note. E = Evidence Absent, M = Meets Most Recommendations, N = Needs More Evidence to Meet Recommendations, R = Recommendations Met.

*The percentages are calculated based on the total number of times each type of racism was observed in the sample, rather than on the total number of scales.

Recommendations, with smaller proportions falling into other categories. Among measures assessing structural ($N=2$) and systemic ($N=3$) racism, all were coded as Meets Most Recommendations. Overall, the distribution of classifications suggests that patterns of psychometric reporting are relatively similar across types of racism rather than concentrated within any single conceptualization.

Why Authors Measured Racism

A majority of the purposes for scale use were tied to interpersonal, intrapersonal, or individual levels of racism and discrimination, with some studies also connecting these personal experiences to broader systemic, structural, or institutional forms of racism. Across the reviewed articles, four primary reasons emerged for why authors used these scales: (a) to examine socio-emotional and psychosocial processes, (b) to assess ideologies and beliefs, (c) to explore environmental or contextual factors, and (d) to support measurement development. First, several researchers studied well-being, stress, self-compassion, psychological adjustment, and empathy to specifically

understand how those emotional and psychosocial experiences were related to or caused by racism, discrimination, and segregation. Second, others used the scales to examine group-based ideologies and beliefs, including language use, attitudes toward diversity, awareness of racism, racial stereotypes, and social distancing. A couple of researchers explored how scales, such as Teaching About Race/Racism (Brooks & Martin, 2023) and Middle School Social Distance Scale (Brandwein & Donoghue, 2011), helped examine race and racism through programmatic and workshop efforts and how teacher ideologies may influence students' learning on these issues.

Third, a smaller subset of the scales was used to measure environmental factors that relate to racism and discrimination. Most of these scales, such as the School Discrimination (Bennett et al., 2020) and the Teacher Discrimination Scale (Kyere et al., 2023), focused on school climate around race and diversity, school-based discrimination, and exclusionary practices within school settings. One scale, Policy Endorsements: School Desegregation and Affirmative Action (Hughes & Bigler, 2011), was more intentional in measuring how systemic-level policies like affirmative action relate to interpersonal and institutional beliefs about policy impact. Similarly, the Race, Racial Attitudes, and School Segregation Survey (Billingham & Hunt, 2016) and the Perceived Discrimination Scale (Faircloth & Hamm, 2005) linked systemic or institutional inequities with interpersonal and intrapersonal experiences. Lastly, researchers' intentional use of six scales was to either develop more accurate assessments of race and racism concepts or to validate the scale (Aldana et al., 2019; Brand et al., 2003; Diemer et al., 2017; Jacobson et al., 2022; Legette et al., 2021; Pieterse et al., 2016). Although specific intentions for scale use varied, researchers explored all levels of racism through their scholarship using selected scales.

How Authors Operationalize Racism

We also found that the majority of tools focused on one type of discrimination, racial/ethnic, rather than examining discrimination through an intersectional lens that included other forms of oppression (intersectionality of sexism, homophobia, etc.). Eleven discussions using these scales provided more concrete operationalizations of racism specifically. The Index of Race-Related Stress (Seaton, 2003), Awareness of Structural Racism (Seider et al., 2018), and Critical Reflection Achievement Gap Attributions (Bañales et al., 2019) scales were used. Authors using the aforementioned scales discussed multiple levels of racism, including individual, interpersonal, institutional, and structural racism. Other scales were used to examine more systemic or institutional dimensions of racism. For example, the Racism and Inequity Beliefs Questionnaire (Legette et al., 2023) draws on critical race theory to explore beliefs about school-based racial inequities, while the Teaching About Race/Racism Scale (Brooks & Martin, 2023) captures how educators address history and white supremacy in their instruction. Authors discussed historical representations of racism and race-related concepts for the Cognitive and Reflective Responses to Lessons Scale (Hughes et al., 2007) but did not state or explicitly define racism. Relevant discussions about race-conscious policies and color-blind racism for the Policy Endorsements: School

Desegregation and Affirmative Action Scale (Hughes & Bigler, 2011), Teachers Race Talk Survey (A. Alvarez & Milner IV, 2018), and CoBRAS (W. Wang et al., 2023) Scales; racial segregation for the Race, Racial Attitudes, and School Segregation Survey (Billingham & Hunt, 2016); and structural dominance with the Teacher Candidates' Perceptions About Adolescents Based on the Adolescents' Physical Characteristics scale (DeCastro-Ambrosetti & Cho, 2011).

Seventeen scales either focused on concepts of discrimination (e.g., racial/ethnic, gender, and/or religion) or contextualized race through discriminatory practices and experiences, which tended to be more focused on the institutional, interpersonal, and intrapersonal. Authors using the Everyday Discrimination Scale (Saleem et al., 2022), Perceived Discrimination based on ethnic group membership (Faircloth & Hamm, 2005), School-Based Racial Discrimination (J. S. Eccles et al., 2006), Perceived Institutional Discrimination (Hope & Jagers, 2014), Perceived Discrimination Scale (Karcher et al., 2022), and the Perceived Ethnic Discrimination (Nieri et al., 2022) Scales focused discussions on how race-based discrimination is related to negative experiences. Other scales, such as the Students' Awareness and Endorsement of Cultural Stereotypes About Race and Math (Cvencek et al., 2015), Experiences of Discrimination Based on Gender or Race (Copur-Gencturk et al., 2023), and Bias-based Bullying (Vigna et al., 2018), had discussions around biases and stereotypes to frame the development and use of the scale. Lastly, some researchers using the School Discrimination (Bennett et al., 2020), Teacher Discrimination Scale (Kyere et al., 2023), and Acculturative Hassles adapted (Oberoi & Trickett, 2018) focused more on structural, school-based discrimination to discuss scales. The other 15 scales either explained their scale through anti-racism efforts, critical reflection practices, general belief systems and ideologies, and the importance of diversity and multiculturalism, or did not have context provided for the exploration of race and racism.

Promising Examples. Based on our review, we wanted to highlight promising examples for the field. We chose measures based on the following criteria: created within the last 10 years (to account for the ever-evolving nature of racism research) and have at least one reliability AND validity estimate from the scores that Met Most Recommendations (M). Some scales were designed to allow multiple stakeholders to take them, including students, teachers, parents, and/or other school staff. The *teacher* scales that met (M) criteria (see Appendix B) were: Attributions of Mathematical Excellence Construct (Jacobson et al., 2022), Racism and Inequity Beliefs (RIB) Questionnaire (Legette et al., 2023), and School/Teacher Racial Encounter Stress Scale (STRESS) (Bentley-Edwards et al., 2020). The *student* scales that met those criteria were: School Climate for Diversity Scale—Secondary (Byrd, 2017) and School Racial Threat adapted (Bentley-Edwards et al., 2020). Some scales were designed to allow multiple stakeholders to take them, including students, teachers, parents, and/or other school staff. The scales that had multiple audiences and met the criteria were: Critical Consciousness Scale (CCS) (Diemer et al., 2017) and Critical Reflection Achievement Gap Attributions (Bañales et al., 2019).

Of the teacher scales, RIB (Legette et al., 2023) was the only scale to explicitly define racism and also measure systemic factors. RIB drew on critical race theory

(CRT) to incorporate definitions of racism (cultural, systemic, individual), cultural deficit, and color-evasiveness. We attributed the type of racism the scale attempted to measure as systemic and intrapersonal/individual racism. Omega was used as a reliability measure (.7 and above for all subscales) and CFA was conducted for validation. The scale sampled a predominantly white teacher population with both rural and urban representation in a Southeastern state. The scale consists of items designed to capture causes of racial inequity that aligned with individualistic, post-racial, deficit, and systemic beliefs, including “In general, racial/ethnic minorities do not work as hard in school as white children,” “In general, teachers treat students differently based on race,” and “In general, racial/ethnic discrimination is a factor in Black kids’ suspensions.” For a larger list of the items, see Appendix Table A.1.

The Critical Reflection Achievement Gap Attributions (Bañales et al., 2019) is another promising example that was administered to students, parents, and school staff that explicitly defined racism as structural versus individual: “Structural attributions point to institutional racism, discrimination, and other systemic impediments to success as causes of social inequities. . . . In contrast, individual attributions are causal explanations that credit people’s work ethic, attitudes, and merit as the causes of achievement gaps.” We coded the type of racism they attempted to measure as intrapersonal/individual and systemic. Authors included Cronbach’s alpha estimates (.8 and above for all subscales) for reliability, and CFA was conducted for validation. The scale was used in a study for urban, mostly Black high school students in a Southeastern region of the United States. To measure structural attributions, the scales indicated “White students usually go to schools with more resources than Black students,” and to measure individual attributions, the scales indicated “Black students don’t work as hard as white students regarding the Black/white academic achievement gap.” Additionally, these items were asked with the following prompt: “There are many theories about why Black students do less well than whites in school. Indicate how much you think these differences are caused by each of the following: individual attributions, structural attributions.” These promising examples can provide a foundation for measuring various facets of racism in educational settings because their items offer specific ways to ask about different aspects of systemic and individual racism and how they manifest within school environments.

Discussion

Although we know the way individuals experience racism changes as our society changes over time, the field of measurement has not kept up, as the average year of development for these measures was 16 years ago (2008). This constant evolution underscores the need for researchers to remain vigilant in ensuring that their instruments accurately reflect current racial dynamics. A majority of the scales centered on personal and individual-based racial discriminatory practices and experiences, but a much smaller subset focused on how those concepts connect to systemic and institutional inequities or were able to tie their scale use to an explicit definition of how they operationalized racism. While beliefs are just one part of structural change, they are

essential in shifting attitudes, though this alone does not equate to structural change (it might over time). This finding indicates a conceptual and methodological gap between how racism operates according to our adapted ecological systems theory, and how it is currently measured. This misalignment reinforces the QuantCrit premise that data are not neutral; what researchers choose to measure reflects prevailing assumptions about where racism lives and how it should be understood.

Our review found a focus on capturing whether respondents experienced or witnessed overt acts of discrimination (e.g., interpersonal and intrapersonal emotions and feelings), but instruments were limiting in three ways: 1) surveys did not ask individuals about potential racist actions they committed; 2) intersectionality was not addressed, and 3) the cumulative effects of microaggressions were not assessed. For the interpersonal level of racism, only one survey asked the respondent to identify if they had demonstrated racist conduct: the Middle School Social Distance Scale (Brandwein & Donoghue, 2011). While most scales made statements about others engaging in racist acts like “Teachers grade you harder because of your race or gender” (Butler-Barnes et al., 2022; Kyere et al., 2023) or responding to statements like “Students of different races/ethnicities work together in class” (Byrd, 2017), the Middle School Social Distance Scale (Brandwein & Donoghue, 2011) asks specifically about the willingness of middle schoolers to accept members of various ethnic groups as friends. Participants indicate the closest level of relationship they would be willing to have with each target group by selecting a number from 1 to 7: 1=best friends, 2=eat lunch with, 3=sit beside in class, 4=say hi only, 5=member of homeroom only, 6=member of school only, and 7=exclude them. Nonetheless, it is about a hypothetical situation versus an after-the-fact interaction, highlighting the urgent need to develop surveys about self-reflection and real interactions.

Similar to discrimination being analyzed mostly through an “experiencing” versus “committing” lens, surveys only analyzed racism through one dimension, ignoring gender, LGBTQIA, disability, religion, immigration, and other forms of marginalized intersectional identities. There were no scales that asked about students’ experiences with disabilities as people of Color. Only one scale asked about being bullied because they were part of the LGBTQIA community, but it was not within the context of being a student of Color (Vigna et al., 2018). Both the Teacher Discrimination Scale (J. A. Eccles, 1997) and the Experiences of Discrimination Based on Gender or Race adapted by Copur-Gencturk et al. (2023) asked about gender; however, this was separate, not in addition to race/ethnicity. The Acculturative Hassles adapted (Vinokurov et al., 2002) did ask specific questions for Muslim students and serves as an example of what a scale tailored toward Islamic students can entail, asking questions such as “An American student treated you in some way because you are Muslim” and “You saw another Muslim student treated badly or discriminated against.” Despite the fact that many students of Color navigate these identities simultaneously, all surveys only captured one dimension of racial discrimination.

Microaggressions are a complex experience that can be defined as being stereotyped, invalidated, or excluded in academic and social settings (Sue et al., 2007). Measuring microaggressions can serve as a bridge between interpersonal experiences

and the institutional practices that reproduce inequities over time, since they can also include environmental factors that uphold racist practices and policies. Unfortunately, our study only found one survey that reliably measured microaggressions (Nadal, 2011) but did not have adequate correlational or construct validity evidence from the scores. This may reflect the reality that many of the reviewed measures were developed around 2008 before microaggressions became widely recognized within educational and public discourse. Yet because these subtle forms of racism are often dismissed as benign, their measurement is critical to understanding how racism persists in “race-neutral” educational environments. If we want to change systems, we need to measure more than student peer interactions, the intersectional aspect of discrimination, and track the cumulative effects and progress over time.

These scales analyzed in this review unveil important patterns on how racism is operationalized (or not) in quantitative K–12 education research. Most measures equate racism with discriminatory events they witnessed from their peers or school staff and ignore asking about their own engagement in such acts, thereby privileging interpersonal manifestations they witness while leaving institutional and structural mechanisms as well as their own acts of racism largely unmeasured. As a result, the field lacks tools that capture racism embedded in policies (e.g., discipline infraction codes, gifted identification procedures, special education identification procedures), resource distribution, and/or historical and political forces shaping school contexts. They also lack survey instruments that would allow respondents to self-reflect and take accountability for their actions. When measures define racism primarily as personal experience and omit asking individuals about their own actions, they risk (in)advertently hiding the larger systems that are a result of those experiences. More explicit and expansive operationalizations of racism, such as those measuring policy beliefs, structural attributions, or institutional practices, demonstrate that it is possible to design tools that capture racism beyond the interpersonal, but these remain the exception rather than the norm.

While we support the use of high-quality psychometric tools to ensure they are applied appropriately within communities, we also recognize that excessively high standards can reinforce systemic and structural inequities, presenting a complex balance between traditional rigor and equity. For example, we would not want a racism measure validated in a Latino population in Chicago used to measure racism for Black students in the South. At the same time, we acknowledge that access to the knowledge, skills, funding, and large sample sizes needed to validate scales is often limited to prestigious institutions, whose scholars are usually well-funded, predominantly white individuals. Our review carefully analyzed reliability and validity estimates in accordance with the Standards for Educational and Psychological Testing (2014). However, we aimed to critically assess potential biases that could perpetuate dominant norms in psychometrics. To avoid reinforcing the “status quo” (i.e., white norms), we reported the psychometric evidence available for each measure (see Appendix Table A.2). Using any scale should be accompanied by both quantitative evidence and clearly articulated qualitative definitions. We recognize that traditional metrics, such as Cronbach’s alpha, potentially provide only a partial view of a measure’s

appropriateness and may reflect normative assumptions that do not generalize across diverse populations. Considering these limitations and the specific constructs each measure was designed to assess is critical when interpreting psychometric evidence and selecting measures for research. Ultimately, prioritizing qualitative evidence in item design aligns with recommended measurement practices (Ho, 2024) and supports more equitable and culturally responsive approaches to assessing racism.

Implications and Future Research

We see this review as a start to raising awareness and increasing research around the extent and forms of racism within education. We call for researchers to discontinue the practice of using race/ethnicity as a default control for racism and instead operationalize and define race and racism. Secondly, systematic research should consider adopting QuantCrit as a guiding framework. Applying QuantCrit in this review allowed us to center racism, challenge traditional psychometric properties, and acknowledge the systemic and complicated nature of racism. Infusing a QuantCrit framework more thoroughly may involve incorporating all tenets at every stage of the review process. For example, rather than relying on traditional screening practices that privilege peer-reviewed articles, reviewers might adopt an intentional approach that includes community-facing reports and briefs. The future of QuantCrit-informed reviews will also require developing new approaches and methodologies for conducting literature reviews, potentially including community-based participatory elements, as aligned with the fourth QuantCrit tenet, *Voice and insight: Data cannot speak for itself*.

Thirdly, we seek to advance more robust and rigorous measures of systemic racism in education, incorporating mixed-methods, longitudinal data, and contextually grounded tools that capture not just attitudes but also structural, systemic, institutional, interpersonal, and intrapersonal aspects of racism. This review demonstrates that respondents' racial identities shape how they perceive and experience racism, underscoring the need for more systemic measures and for tools that capture the complexity of racism across different racial groups. We also caveat that this is a limitation of all measurements, as no one measure can 100% accurately represent an entire group of people; however, more nuance and specificity for particular groups is warranted for future research. We uncovered the necessity to create more measures for rural, Asian, and Indigenous students. Only six scales surveyed a rural student population, only one used a majority Asian student sample, and none used a majority Indigenous sample. We acknowledge that a limitation of this study is its focus on K–12 studies, despite evidence that racism begins early and persists throughout the lifespan. We understand the need to study this, especially given not only the additional considerations in surveying or assessing young children, but also the reality of adult implicit bias toward young children in disciplinary, observational, assessment data among others (Blackson et al., 2022; Gilliam et al., 2016; Giordano et al., 2021; Meek & Gilliam, 2016). Nonetheless, future research should examine how to measure racism to understand the

nuances and needs of early age groups, potentially through instruments that assess early educators' or parents' critical consciousness (e.g., Vesely et al., 2023).

We also acknowledge that survey instruments measuring racist attitudes must adapt continuously to reflect evolving forms of racial bias and exclusion. Given the evolving nature of racism, we recommend a combination of both in-the-moment surveys and retrospective surveys after students have graduated. In-the-moment surveys would provide immediate insights into students' experiences of racism within the current educational environment, while post-graduation surveys allow for a reflective view of how racism and bias may have impacted them over time. Additionally, future measurements of racism should include historical, geographical, political, and theoretical aspects to capture its multifaceted, systemic, and structural properties.

It is crucial to create multidimensional measures that go beyond individual scales and self-report surveys by using innovative quantitative and qualitative components to capture a broader spectrum of experiences. According to Ho (2024), "Measurement must be qualitative, then quantitative, then qualitative again." Building on this perspective, we emphasize the importance of using mixed methods in survey development to capture the complex and evolving nature of racism more effectively, as well as how these phenomena vary across different populations over time. For instance, employing an exploratory sequential design (Creswell, 2022) enables researchers first to gather qualitative data to identify key factors shaping perceptions of racism, which can then inform the development of targeted quantitative survey instruments. Respondents, especially youth or individuals with internalized racism, might lack the terminology or awareness to recognize structural forces, leading to systematic underreporting in survey-based measures of systemic racism. Therefore, we advise researchers to incorporate qualitative data within an explanatory mixed-methods approach to contextualize and interpret the survey results (Creswell, 2022). This enables us to understand how institutional policies and systemic forces influence the patterns seen in the survey data. This approach ensures that the resulting measures are not only relevant but also grounded in the lived experiences of the population being studied. Given the multidimensional and shifting nature of racism, survey developers should also integrate qualitative methods, such as interviews, ethnographic studies, and observational research, to capture both overt and subtle forms of racial discrimination. By including these diverse data sources, researchers can create more comprehensive and nuanced surveys that provide deeper insights into how racism is experienced and perceived.

To address the drawback of survey measures, researchers have developed promising examples. While not the primary focus of this paper, we note the presence of anti-racist measures and tools that hold potential for advancing this work further, particularly to encourage future research and highlight innovative work. We want to warn that despite these innovative approaches, these tools remain controversial, with some scholars questioning their reliability and validity (Schimmack, 2021b). The Implicit Association Test (IAT) attempts to assess unconscious biases by measuring reaction times when participants (usually young pre-school-aged) are asked to associate positive or negative attributes with racial groups (Greenwald et al., 1998). Recent research has documented the need to understand both implicit and explicit bias in

educators using either the Child Race IAT or Adult Race IAT, as self-report measures differ vastly from implicit measures (Blackson et al., 2022). The authors recommend expanding the work into early childhood and early school settings, as early discriminatory experiences can have long-lasting effects on children (Blackson et al., 2022). Nevertheless, recent critiques suggest that the IAT may exhibit poor internal consistency and construct validity evidence (Schimmack, 2021a, 2021b). We advise using this tool carefully and following the recommended practices outlined by Greenwald et al. (2022).

Another innovative approach to measuring school segregation, which can be (not always) examples of institutional and systemic racism, is Fiel's (2013) study. Fiel (2013) used a decomposition approach to assess the relative contributions of distributive processes and compositional changes to the observed resegregation of schools between 1993 and 2010. Other proxies that researchers have used to measure school segregation include the Index of Dissimilarity (D), which is frequently used for measuring residential segregation (Frankel & Volij, 2011; Frankenberg, 2013). D is a two-group measure that assesses the degree to which groups are evenly distributed across geographical areas, and education scholars have used D to measure within-district segregation in racially isolated districts (Frankenberg, 2013). Lastly, Polos et al. (2022) measured school-level structural racism by analyzing a combination of student demographics, sense of connectedness, perceived life opportunities, school discipline records, and attendance data. While these alternatives exist, it is crucial to sift through research goals before using any of these measures. As recommended, use qualitative definitions and see whether they fit the aims of each scale, aligning with measurement best practices (Ho, 2024).

Although we can attempt to quantitatively measure racism, this is not the same as actively working toward reducing and eliminating the interconnected systems that perpetuate racism and other forms of oppression. This work will not be done overnight and requires deconstructing and rebuilding across all sectors, including public and private educational institutions. The process of quantifying racism in education will be iterative and aspirational in nature (O. Johnson Jr., personal communication, June 2024). Quantitative methods will eventually reach their limits, and we may never be able to unpack the complexity and layers of systemic and structural racism. However, we remain committed to trying to measure it as best we can, with the understanding that better data can lead to more resources for racially minoritized students and the development of anti-racist programs and policies. We aim to inspire ongoing research and action toward racial justice, understanding that this work requires the constant reevaluation of our measures and methods to match the shifting cultural realities of how racism operates, and the specific ways it impacts marginalized communities. As racism continues to evolve, so too must the methods by which it is measured, and given today's political environment, this work is even more important.

Acknowledgments

The team would like to thank Katherine Aragon for helping with the second round of coding and references, and Neil Charlery for helping with the second round of coding.

ORCID iDs

Wendy Castillo  <https://orcid.org/0009-0004-5278-2400>

Juan Cruz  <https://orcid.org/0009-0006-5007-8035>

Rachel Renbarger  <https://orcid.org/0000-0001-8754-5510>

Christen Priddie  <https://orcid.org/0000-0001-9009-8328>

Brein Mosely  <https://orcid.org/0009-0002-9474-2410>

Sasha Mejia-Bradford  <https://orcid.org/0000-0003-4848-9645>

Ethical Considerations

All data were gathered through secondary public sources, and no IRB was required.

Author Contributions

Drs. Castillo and Renbarger conceptualized the article. Dr. Castillo was the lead on ensuring we adhered to the a priori coding as well as leading meetings and discussions to resolve any conflicting coding. Drs. Castillo and Renbarger wrote the majority of the article and helped with all rounds of coding and revisions. Mr. Cruz supported the second and final round of coding and the writing of the psychometric sections. He also helped significantly revise the manuscript. Dr. Priddie helped with the second round of coding and writing of two sections as well as revisions. Ms. Mosely helped with the second round of coding and the descriptive analysis as well as revisions. Mrs. Mejia-Bradford helped with the initial searches and the first round of coding

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

Note

1. We intentionally use a lowercase white to distance ourselves from white supremacists who capitalize the term (Daniszewski, 2020).

References

* Indicates the articles included as part of the systematic review

- *Aldana, A., Bañales, J., & Richards-Schuster, K. (2019). Youth anti-racist engagement: Conceptualization, development, and validation of an Anti-Racism Action Scale. *Adolescent Research Review*, 4(4), 369–381. <https://doi.org/10.1007/s40894-019-00113-1>
- Allen, W. R., Suh, S. A., Gonzalez, G., & Yang, J. (2008). Qui bono? Explaining—or defending—winners and losers in the competition for educational achievement. In T. Zuberi & E.

- Bonilla-Silva (Eds.), *White logic, white methods: Racism and methodology* (pp. 217–238). Rowman and Littlefield.
- *Alvarez, A., & Milner, H. R., IV. (2018). Exploring teachers' beliefs and feelings about race and police violence. *Teaching Education*, 29(4), 383–394. <https://doi.org/10.1080/10476210.2018.1512091>
- Alvarez, B. (2024). What does it mean to be an anti-racist? *NEAToday*. <https://www.nea.org/nea-today/all-news-articles/what-does-it-mean-be-anti-racist>
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association. <https://www.testingstandards.net/>
- *Anyiwo, N., Anderson, R. E., Marchand, A. D., Diemer, M. A., & Garrett, J. M. (2023). They raised me to resist: Examining the sociopolitical pathways between parental racial socialization and Black youth's racial justice action. *Journal of Community & Applied Social Psychology*, 33(2), 270–286. <https://doi.org/10.1002/casp.2652>
- *Anyiwo, N., Watkins, D. C., & Rowley, S. J. (2022). “They Can’t Take Away the Light”: Hip-hop culture and Black youth’s racial resistance. *Youth & Society*, 54(4), 611–634. <https://doi.org/10.1177/0044118X211001096>
- Atkins, R. (2014). Instruments measuring perceived racism/racial discrimination: Review and critique of factor analytic techniques. *International Journal of Social Determinants of Health and Health Services* 44(4), 711–734. <https://doi.org/10.2190/HS.44.4.c>
- Au, W. (2020). Testing for whiteness? How high-stakes, standardized tests promote racism, undercut diversity, and undermine multicultural education. In S. May & C. Sleeter (Eds.), *Visioning multicultural education* (pp. 99–113). Routledge. <https://doi.org/10.4324/9781003095644-9>
- Augoustinos, M., & Rosewarne, D. L. (2001). Stereotype knowledge and prejudice in children. *British Journal of Developmental Psychology*, 19, 143–156. <https://doi.org/10.1348/026151001165912>
- Bailey, Z. D., Krieger, N., Agénor, M., Graves, J., Linos, N., & Bassett, M. T. (2017). Structural racism and health inequities in the USA: Evidence and interventions. *Lancet (London, England)*, 389(10077), 1453–1463. [https://doi.org/10.1016/S0140-6736\(17\)30569-X](https://doi.org/10.1016/S0140-6736(17)30569-X)
- Baker, D. J., Ford, K. S., Viano, S., & Johnston-Guerrero, M. P. (2024). Racial category usage in education research: Examining the publications from AERA journals. *AERA Open*, 10, 23328584241293681. <https://doi.org/10.1177/23328584241293681>
- Baker, F. B., & Kim, S. H. (Eds.). (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). CRC Press.
- Banaji, M. R., Fiske, S. T., & Massey, D. S. (2021). Systemic racism: Individuals and interactions, institutions and society. *Cognitive Research: Principles and Implications*, 6(1), 82. <https://doi.org/10.1186/s41235-021-00349-3>
- *Bañales, J., Marchand, A. D., Skinner, O. D., Anyiwo, N., Rowley, S. J., & Kurtz-Costes, B. (2020). Black adolescents’ critical reflection development: Parents’ racial socialization and attributions about race achievement gaps. *Journal of Research on Adolescence*, 30: 403–417. <https://doi.org/10.1111/jora.12485>
- *Bennett, M., Roche, K. M., Huebner, D. M., & Lambert, S. F. (2020). School discrimination and changes in Latinx adolescents’ internalizing and externalizing symptoms. *Journal of Youth and Adolescence*, 49(10), 2020–2033. <https://doi.org/10.1007/s10964-020-01256-4>
- Belfield, C., Bowden, A. B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). The economic value of social and emotional learning. *Journal of Benefit-Cost Analysis*, 6(3), 508–544. doi:10.1017/bca.2015.55

- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588–606. <https://doi.org/10.1037/0033-2909.88.3.588>
- *Bentley-Edwards, K. L., Stevenson, H. C., Jr., Thomas, D. E., Adams-Bass, V. N., & Coleman-King, C. (2020). Teaching scared: Pre-service teacher appraisals of racial stress, socialization and classroom management self-efficacy. *Social Psychology of Education*, 23(5), 1233–1257. <https://doi.org/10.1007/s11218-020-09578-8>
- Bernal, D. (2016). Cultural intuition: Then, now, and into the future. *Center for Critical Race Studies Research Briefs*, 1, 1-4. https://d1wqtxts1xzle7.cloudfront.net/51206194/Cultural_Intuition.Research_Brief_2016_copy-libre.pdf?1483637246=&response-content-disposition=inline%3B+filename%3DCultural_Intuition_Then_Now_and_Into_the.pdf&Expires=1786040751&Signature=e1JvBVdgXaTvkz40PvR9gJKITE12hiHRdk2sNvI~YrjKbOYdG8MqgPW1v~ligdylSXippF6D81Z61VvZ8smFomtQjU-RWfnHVJOCB9kfXrUDSAFXrU0~MTIBQbbhWqRtMx2UkFMhNe-7uWRmwIeTTvGIBUroI8EnW-toQnO0XM6sKv9JaFiUFiXlsyKeK04ONUHKsUXjCWZOJ-QEgAtlCpOmKpwJ1gDw-pQbiOnO1PuOVLXQ6fGr5oRaMdCG6dZtHQsEGc7ITahAYhQWI9m8heOx9Zbl71M-tlCgVZYibRmS7AafbmenI1X1FjGDRRC5fppmu6RBgZR1FfbqLrCTwZ3w__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
- *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>
- Blackson, E. A., Gerdes, M., Segan, E., Anokam, C., & Johnson, T. J. (2022). Racial bias toward children in the early childhood education setting. *Journal of Early Childhood Research*, 20(3), 277–292. <https://doi.org/10.1177/1476718x221087051>
- Bogardus, E. S. (1933). A social distance scale. *Sociology and Social Research*, 17, 265–271.
- Bonilla-Silva, E. (2021). *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States* (6th ed.). Rowman & Littlefield.
- Boveda, M., Ford, K. S., Frankenberg, E., & López, F. (2023). Editorial Vision 2022–2025. *Review of Educational Research*, 93(5), 635–640. <https://doi.org/10.3102/00346543231170179>
- *Brand, S., Felner, R., Shim, M., Seitsinger, A., & Dumas, T. (2003). Middle school improvement and reform: Development and validation of a school-level assessment of climate, cultural pluralism, and school safety. *Journal of Educational Psychology*, 95(3), 570–588. <https://doi.org/10.1037/0022-0663.95.3.570>
- *Brandwein, D., & Donoghue, C. (2011). A multicultural grassroots effort to reduce ethnic and racial social distance among middle school students. *Multicultural Education*, 18(2), 27–32.
- Brennan, R. L., National Council on Measurement in Education, & American Council on Education. (2006). *Educational measurement* (4th ed.). Praeger.
- Bridgeforth, J. C. (2025). Leading through the fire: Responding to racism in K-12 schools. *Journal of Research on Leadership Education*, 20(4), 399–427. <https://doi.org/10.1177/19427751251341299>
- Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology*, 23, 1933–1967. <https://doi.org/10.1111/j.1559-1816.1993.tb01074>
- *Brody, G. H., Chen, Y.-f., Murry, V. M., Ge, X., Simons, R. L., Gibbons, F. X., Gerrard, M., & Cutrona, C. E. (2006). Perceived discrimination and the adjustment of African American youths: A five-year longitudinal analysis with contextual moderation effects. *Child Development*, 77(5), 1170–1189. <https://doi.org/10.1111/j.1467-8624.2006.00927.x>
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32(7), 513. <https://doi.org/10.1037/0003-066X.32.7.513>

- Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, 34(10), 844–850. <https://doi.org/10.1037/0003-066X.34.10.844>
- Bronfenbrenner, U. (1988). Interacting systems in human development. *Research paradigms: Present and future*. In N. Bolger, A. Caspi, G. Downey, & M. Moorehouse (Eds.), *Persons in contexts: Developmental processes* (pp. 25–49). Cambridge University Press.
- *Brooks, C., & Martin, D. (2023). History teachers thinking about race and whiteness: Possibilities and challenges. *Social Studies Research and Practice*, 18(2), 97–109. <https://doi.org/10.1108/SSRP-01-2023-0007>
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.
- *Butler, A. C. (2022). Sexual and physical assault before age 18 and young adults' perception of unfair treatment. *Journal of Interpersonal Violence*, 37(11–12), NP10296–NP10328. <https://doi.org/10.1177/0886260520985481>
- *Butler-Barnes, S. T., Leath, S., Inniss-Thompson, M. N., Allen, P. C., D'Almeida, M. E. D. A., & Boyd, D. T. (2022). Racial and gender discrimination by teachers: Risks for Black girls' depressive symptomatology and suicidal ideation. *Cultural Diversity & Ethnic Minority Psychology*, 28(4), 469–482. <https://doi.org/10.1037/cdp0000538>
- *Byrd, C. M. (2017). The complexity of school racial climate: Reliability and validity of a new measure for secondary students. *British Journal of Educational Psychology*, 87(4), 700–721. <https://doi.org/10.1111/bjep.12179>
- *Byrd, C. M., & Legette, K. B. (2022). School ethnic-racial socialization and adolescent ethnic-racial identity. *Cultural Diversity & Ethnic Minority Psychology*, 28(2), 205–216. <https://doi.org/10.1037/cdp0000449>
- Cano, M., Portillo, A. G. P., Figuereo, V., Rahman, A., Reyes-Martínez, J., Rosales, R., Ángel Cano, M., Salas-Wright, C. P., & Takeuchi, D. T. (2021). Experiences of ethnic discrimination among US Hispanics: Intersections of language, heritage, and discrimination setting. *International Journal of Intercultural Relations*, 84, 233–250. <https://doi.org/10.1016/j.ijintrel.2021.08.006>
- Carter, P. L. (2005). *Keepin' it real: School success beyond black and white*. Oxford University Press.
- Carter, E. R., & Murphy, M. C. (2015). *Group-based differences in perceptions of racism: What counts, to whom, and why?* *Social and Personality Psychology Compass*, 9(6), 269–280. <https://doi.org/10.1111/spc3.12181>
- CASEL. (n.d.). *Our history*. Author. <https://casel.org/about-us/our-history/>
- Castillo, W. (2020). Do elementary students reading motivation levels differ by racial/ethnic and/or immigrant background? *Journal of Latinos and Education*, 22(2), 669–680. <https://doi.org/10.1080/15348431.2020.1805615>
- Castillo, W., & Gillborn, E. (2022). *How to “QuantCrit:” Practices and questions for education data researchers and users* (EdWorkingPaper: 22–546). Annenberg Institute at Brown University. <https://doi.org/10.26300/v5kh-dd65>
- Castillo, W., & Strunk, K. K. (2024). *How to QuantCrit: Applying critical race theory to quantitative data in education* (1st ed.). Routledge. <https://doi.org/10.4324/9781003429968>
- Cattell, R. B. (1965). *The scientific analysis of personality*. Penguin Books.
- *Cho, H., Chou, A., & Wiens, P. D. (2023). Teacher education students' perceptions about bilingualism and emergent bilingual students. *Action in Teacher Education*, 45(2), 107–123. <https://doi.org/10.1080/01626620.2023.2174203>
- Chudy, J., & Jefferson, H. (2021). *Support for Black Lives Matter surged last year. Did it last?* *The New York Times*, May 22. <https://www.nytimes.com/2021/05/22/opinion/blm-movement-protests-support.html>

- Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development, 94*(5), 1181–1204. <https://doi.org/https://doi.org/10.1111/cdev.13968>
- Cobb, C. L., Meca, A., Branscombe, N. R., Schwartz, S. J., Xie, D., Zea, M. C., Fernandez, C. A., & Sanders, G. L. (2019). Perceived discrimination and well-being among unauthorized Hispanic immigrants: The moderating role of ethnic/racial group identity centrality. *Cultural Diversity and Ethnic Minority Psychology, 25*(2), 280. <https://doi.org/10.1037/cdp0000227>
- Committee for Children. (2019, January 15). *Record high demand for social-emotional learning in US schools*. <https://www.cfchildren.org/blog/2019/01/record-high-demand-for-social-emotional-learning-in-us-schools/>
- *Copur-Gencturk, Y., Thacker, I., & Cimpian, J. R. (2023). Teachers' race and gender biases and the moderating effects of their beliefs and dispositions. *International Journal of STEM Education, 10*(1), 31. <https://doi.org/10.1186/s40594-023-00420-z>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Crawford, B. F., Snyder, K. E., & Adelson, J. L. (2020). Exploring obstacles faced by gifted minority students through Bronfenbrenner's bioecological systems theory. *High Ability Studies, 31*(1), 43–74. <https://doi.org/10.1080/13598139.2019.1568231>
- Creswell, J. W. (2022). *A concise introduction to mixed methods research* (2nd ed.). SAGE Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. <https://doi.org/10.1007/BF02310555>
- *Cvencek, D., Nasir, N. S., O'Connor, K., Wischnia, S., & Meltzoff, A. N. (2015). The development of math-race stereotypes: "They say Chinese people are the best at math." *Journal for Research on Adolescence, 25*, 630–637. <https://doi.org/10.1111/jora.12151>
- Daniszewski, J. (2020, July 20). *Why we will lowercase white*. Associated Press. <https://blog.ap.org/announcements/why-we-will-lowercase-white>
- *DeCastro-Ambrosetti, D., & Cho, G. (2011). A look at "lookism": A critical analysis of teachers' expectations based on students' appearance. *Multicultural Education, 18*(2), 51–54.
- DeCuir-Gunby, J. T., & Schutz, P. A. (2014). Researching race within educational psychology contexts. *Educational Psychologist, 49*(4), 244–260. <https://doi.org/10.1080/00461520.2014.957828>
- *Dedeoglu, H., & Lamme, L. L. (2011). Selected demographics, attitudes, and beliefs about diversity of preservice teachers. *Education and Urban Society, 43*(4), 468–485. <https://doi.org/10.1177/0013124510380718>
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18. <https://doi.org/10.1037/0022-3514.56.1.5>
- *Diemer, M. A., & Rapa, L. J. (2016). Unraveling the complexity of critical consciousness, political efficacy, and political action among marginalized adolescents. *Child Development, 87*(1), 221–238. <https://doi.org/10.1111/cdev.12446>
- *Diemer, M. A., Rapa, L. J., Park, C. J., & Perry, J. C. (2017). Development and validation of the critical consciousness scale. *Youth and Society, 49*, 461–483. <https://doi.org/10.1177/0044118X14538289>

- Domínguez, S., & Embrick, D. G. (2020). Racial microaggressions: Bridging psychology and sociology and future research considerations. *Sociology Compass*, 14(8), Article number 12803. <https://doi.org/10.1111/soc4.12803>
- Dunn, T. J., Baguley, T., & Bruns, J. (2014). From alpha to omega: A practical guide to the estimation of Cronbach's alpha and omega coefficients. *European Journal of Psychological Assessment*, 30(3), 220–227.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>
- Durlak, J. A., Mahoney, J. L., & Boyle, A. E. (2022). What we know, and what we need to find out about universal, school-based social and emotional learning programs for children and adolescents: A review of meta-analyses and directions for future research. *Psychological Bulletin*, 148(11–12), 765–782. <https://doi.org/10.1037/bul0000383>
- D'Ignazio, C., & F. Klein, L. (2020). Seven intersectional feminist principles for equitable and actionable COVID-19 data. *Big Data & Society*, 7(2), 1–6. <https://doi.org/10.1177/2053951720942544>
- Eccles, J. A. (1997). *MADICS study of adolescent development in multiple contexts, 1991–2012* [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/TVXUIN>
- *Eccles, J. S., Wong, C. A., & Peck, S. C. (2006). Ethnicity as a social context for the development of African-American adolescents. *Journal of School Psychology*, 44(5), 407–426. <https://doi.org/10.1016/j.jsp.2006.04.001>
- Evans, R. J. (2020, July 30). *R. A. Fisher and the science of hatred*. *New Statesman*. <https://www.newstatesman.com/long-reads/2020/07/ra-fisher-and-science-hatred>
- *Faircloth, B. S., & Hamm, J. V. (2005). Sense of belonging among high school students representing 4 ethnic groups. *Journal of Youth and Adolescence*, 34(4), 293–309. <https://doi.org/10.1007/s10964-005-5752-7>
- Fiel, J. E. (2013). Decomposing school resegregation: Social closure, racial imbalance, and racial isolation. *American Sociological Review*, 78(5), 828–848. <https://doi.org/10.1177/0003122413496252>
- *Fike, K. J., Mattis, J. S., Nickodem, K., & Guillaume, C. (2023). Black adolescent altruism: Exploring the role of racial discrimination and empathy. *Children and Youth Services Review*, 150, Article number 106990. <https://doi.org/10.1016/j.childyouth.2023.106990>
- *Fine, M., Torre, M. E., Frost, D. M., & Cabana, A. L. (2018). Queer solidarities: New activism erupting at the intersection of structural precarity and radical misrecognition. *Journal of Social and Political Psychology*, 6(2), 608–630. <https://doi.org/10.5964/jspp.v6i2.905>
- Fisher, C. B., Wallace, S. A., & Fenton, R. E. (2000). Discrimination distress during adolescence. *Journal of Youth and Adolescence*, 29(6), 679–695. <https://doi.org/10.1023/A:1026455906512>
- Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, 48, 621–628. <https://doi.org/10.1037/0003-066x.48.6.621>
- *Flores, B. B., & Smith, H. L. (2008). Teachers' characteristics and attitudinal beliefs about linguistic and cultural diversity. *Bilingual Research Journal*, 31(1–2), 323–358. <https://doi.org/10.1080/15235880802640789>
- Fong, C. J., & Irizarry, Y. (2025). Too quant to crit? Advancing QuantCrit methodologies in educational psychology. *Contemporary Educational Psychology*, 80, Article number 102341. <https://doi.org/10.1016/j.cedpsych.2025.102341>
- Frankel, D. M., & Volij, O. (2011). Measuring school segregation. *Journal of Economic Theory*, 146(1), 1–38. <https://doi.org/10.1016/j.jet.2010.10.008>

- Frankenberg, E. (2013). The role of residential segregation in contemporary school segregation. *Education and Urban Society*, 45(5), 548–570. <https://doi.org/10.1177/0013124513486288>
- Furtado, K., Rao, N., Payton, M., Brown, K., Balu, R., & Dubay, L. (2023). *Measuring structural racism*. Urban Institute. Retrieved February 5, 2024, from <https://www.urban.org/research/publication/measuring-structural-racism>
- Gillborn, D. (2016). Softly, softly: Genetics, intelligence and the hidden racism of the new geneism. *Journal of Education Policy*, 31(4), 365–388. <https://doi.org/10.1080/02680939.2016.1139189>
- Gillborn, D., Warmington, P., & Demack, S. (2018). QuantCrit: Education, policy, ‘Big Data’ and principles for a critical race theory of statistics. *Race Ethnicity and Education*, 21(2), 158–179. <https://doi.org/10.1080/13613324.2017.1377417>
- 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.
- *Golden, A. R., & Byrd, C. M. (2022). Examining critical reflection as a mediator between school racial climate experiences and anti-racist action. *Journal of Research on Adolescence*, 32(3), 1109–1119. <https://doi.org/10.1111/jora.12778>
- Gong, F., Xu, J., & Takeuchi, D. T. (2016). Racial and ethnic differences in perceptions of everyday discrimination. *Sociology of Race and Ethnicity*, 3(4), 506–521. <https://doi.org/10.1177/2332649216681587>
- Giordano, K., Interra, V. L., Stillo, G. C., Mims, A. T., & Block-Lerner, J. (2021). Associations Between Child and Administrator Race and Suspension and Expulsion Rates in Community Childcare Programs. *Early Childhood Education Journal*, 49, 125–133. <https://doi.org/10.1007/s10643-020-01049-1>
- Greenwald, A. G., Brendl, M., Cai, H., Cvencek, D., & Nosek, B. A. (2022). Best research practices for using the Implicit Association Test. *Behavior Research Methods*, 54, 1161–1180. <https://doi.org/10.3758/s13428-021-01624-3>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Groos, M., Wallace, M., Hardeman, R., & Theall, K. P. (2018). Measuring inequity: A systematic review of methods used to quantify structural racism. *Journal of Health Disparities Research and Practice*, 11(2), 190–205. <https://digitalscholarship.unlv.edu/jhdrp/vol11/iss2/13>
- Harber, K. D., Gorman, J. L., Gengaro, F. P., Butisingh, S., Tsang, W., & Ouellette, R. (2012). Students’ race and teachers’ social support affect the positive feedback bias in public schools. *Journal of Educational Psychology*, 104(4), 1149–1161. <https://doi.org/10.1037/a0028110>
- Hardeman, R. R., Homan, P. A., Chantarat, T., Davis, B. A., & Brown, T. H. (2022). Improving the measurement of structural racism to achieve antiracist health policy: Study examines measurement of structural racism to achieve antiracist health policy. *Health Affairs*, 41(2), 179–186. <https://doi.org/10.1377/hlthaff.2021.01489>
- Harper, S. R. (2012). Race without racism: How higher education researchers minimize racist institutional norms. *Review of Higher Education*, 36(1), 9–29. <https://doi.org/10.1353/rhe.2012.0047>
- Harper, S. R., & Davis, C. H., III. (2012). They (don’t) care about education: A counternarrative on Black male students’ responses to inequitable schooling. *Educational Foundations*, 26, 103–120.

- Harrison, H., Griffin, S. J., Kuhn, I., & Usher-Smith, J. A. (2020). Software tools to support title and abstract screening for systematic reviews in healthcare: An evaluation. *BMC Medical Research Methodology*, 20, 1–12. <https://doi.org/10.1186/s12874-020-0897-3>
- *Heberle, A. E., Hoch, N., Wagner, A. C., Caccia Cruz, A. G., Zhou, L., & Khan, A. K. (2024). Examining feasibility, acceptability, and participant experiences of an anti-racist parenting intervention for White U.S. parents. *The American Journal of Orthopsychiatry*, 94(1), 33–47. <https://doi.org/10.1037/ort0000703>
- Hicken, M. T., Kravitz-Wirtz, N., Durkee, M., & Jackson, J. S. (2018). Racial inequalities in health: Framing future research. *Social Science & Medicine*, 199, 11–18. <https://doi.org/10.1016/j.socscimed.2017.12.027>
- Ho, A. D. (2022). Specifying the three Ws in educational measurement: Who uses which scores for what purpose? *Journal of Educational Measurement*, 59, 418–422. <https://doi.org/10.1111/jedm.12355>
- Ho, A. D. (2024). Measurement must be qualitative, then quantitative, then qualitative again. *Educational Measurement: Issues and Practice*, 43(4), 137–145. <https://doi.org/10.1111/emip.12662>
- *Hope, E. C., & Jagers, R. J. (2014). The role of sociopolitical attitudes and civic education in the civic engagement of Black youth. *Journal of Research on Adolescence*, 24(3), 460–470. <https://doi.org/10.1111/jora.12117>
- Howard, T. C., Douglas, T. R. M., & Warren, C. A. (2016). “What works”: Recommendations on improving academic experiences and outcomes for Black males. *Teachers College Record*, 118(6), 1–10. <https://doi.org/10.1177/016146811611800602>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>
- *Hughes, J. M., & Bigler, R. S. (2011). Predictors of African American and European American adolescents’ endorsement of race-conscious social policies. *Developmental Psychology*, 47(2), 479–492. <https://doi.org/10.1037/a0021309>
- *Hughes, J. M., Bigler, R. S., & Levy, S. R. (2007). Consequences of learning about historical racism among European American and African American children. *Child Development*, 78, 1689–1705. <https://doi.org/10.1111/j.1467-8624.2007.01096.x>
- *Jacobson, E., Cross Francis, D., Willey, C., & Wilkins-Yel, K. (2022). Race, gender, and teacher equity beliefs: Construct validation of the Attributions of Mathematical Excellence Scale. *AERA Open*, 8, 1–17. <https://doi.org/10.1177/23328584221130999>
- Jencks, C. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. Basic Books.
- Jones, J. M. (1971). The political dimensions of black liberation. *The Black Scholar*, 3(1), 67–75. <https://doi.org/10.1080/00064246.1971.11431187>
- Jones, T. M., Kim, B. K. E., Fleming, C. B., Deng, J., Duane, A., Gavin, A. R., & Shapiro, V. B. (2026). To whom do these results apply? Assessing evidence for the generalizability of social and emotional learning programs among specific racial and ethnic groups. *Review of Educational Research*, 96(2), 543–583. <https://doi.org/10.3102/00346543241310184>
- Jung, L. (2021). Big business pledged nearly \$50 billion for racial justice after George Floyd’s death. Where did the money go? *The Washington Post*, August 23. <https://www.washingtonpost.com/business/interactive/2021/george-floyd-corporate-america-racial-justice/>
- *Kang, S., & Harvey, E. A. (2020). Racial differences between Black parents’ and White teachers’ perceptions of Attention-Deficit/Hyperactivity Disorder Behavior. *Journal of Abnormal Child Psychology*, 48(5), 661–672. <https://doi.org/10.1007/s10802-019-00600-y>

- *Karcher, N. R., Klaunig, M. J., Elsayed, N. M., Taylor, R. L., Jay, S. Y., & Schiffman, J. (2022). Understanding associations between race/ethnicity, experiences of discrimination, and psychotic-like experiences in middle childhood. *Journal of the American Academy of Child and Adolescent Psychiatry*, 61(10), 1262–1272. <https://doi.org/10.1016/j.jaac.2022.03.025>
- Kohli, R., Pizarro, M., & Nevárez, A. (2017). The “new racism” of K–12 schools: Centering critical research on racism. *Review of Research in Education*, 41(1), 182–202. <https://doi.org/10.3102/0091732X1668694>
- Kolluri, S., Hypolite, L., Patterson, A., & Young, K. (2023). *The Racial reckoning and the role of schooling: Exploring the potential of integrated classrooms and liberatory pedagogies*. UCLA Civil Rights Project/Proyecto Derechos Civiles. <https://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/racial-reckoning-and-the-role-of-schooling-exploring-the-potential-of-integrated-classrooms-and-liberatory-pedagogies/Schools-and-Racial-Reckoning-082923.pdf>
- Kozol, J. (1991). *Savage inequalities*. Crown Pub.
- Kulis, S., Marsiglia, F. F., & Nieri, T. (2009). Perceived ethnic discrimination versus acculturation stress: Influences on substance use among Latino youth in the Southwest. *Journal of Health and Social Behavior*, 50(4), 443–459. <https://doi.org/10.1177/002214650905000405>
- *Kyer, E., Hong, S., & Gentle-Genitty, C. S. (2023). Mediation effect of teacher-based discrimination on academic performance: An intersectional analysis of race, gender, and income/class. *Education Sciences*, 13(4), 387. <https://doi.org/10.3390/educsci13040387>
- Landrine, H., & Klonoff, E. A. (1996). The schedule of racist events: A measure of racial discrimination and a study of its negative physical and mental health consequences. *Journal of Black Psychology*, 22(2), 144–168. <https://doi.org/10.1177/00957984960222002>
- *Lee, R. M., Grotevant, H. D., Hellerstedt, W. L., & Gunnar, M. R. (2006). Cultural socialization in families with internationally adopted children. *Journal of Family Psychology*, 20(4), 571. <https://doi.org/10.1037/0893-3200.20.4.571>
- *Legette, K. B., Halberstadt, A. G., & Majors, A. T. (2021). Teachers’ understanding of racial inequity predicts their perceptions of students’ behaviors. *Contemporary Educational Psychology*, 67, 1–11. <https://doi.org/10.1016/j.cedpsych.2021.102014>
- *Legette, K. B., Supple, A., & Halberstadt, A. G. (2023). The Racism and Inequity Beliefs Questionnaire: Construction and initial validation. *The Urban Review*, 55(5), 605–623. <https://doi.org/10.1007/s11256-023-00663-8>
- Lemann, N. (1999, February 14). *The big test* [Review of the book *The Big Test: The Secret History of the American Meritocracy*, by N. Lemann]. *The New York Times Book Review*, 104(6), 6. <https://doi.org/10.22439/asca.v32i2.2772>
- *Lesane-Brown, C. L., Brown, T. N., Caldwell, C. H., & Sellers, R. M. (2005). The comprehensive race socialization inventory. *Journal of Black Studies*, 36(2), 163–190. <https://doi.org/10.1177/0021934704273457>
- *Liang, X., & Zhang, G. (2009). Indicators to evaluate pre-service teachers’ cultural competence. *Evaluation & Research in Education*, 22(1), 17–31. <https://doi.org/10.1080/09500790903082354>
- Massachusetts Attorney General’s Office. (2025, February 13). *Multi-State guidance concerning diversity, equity, inclusion, and accessibility employment initiatives*. <https://www.mass.gov/doc/multi-state-guidance-concerning-diversity-equity-inclusion-and-accessibility-employment-initiatives/download>
- McNeilly, M. D., Anderson, N. B., Armstead, C. A., Clark, R., Corbett, M., Robinson, E. L., Pieper, C. F., & Lepisto, E. M. (1996). *Perceived Racism Scale* [Database record]. APA PsycTests. <https://doi.org/10.1037/t70724-000>

- Meek, S. E., & Gilliam, W. S. (2016). Expulsion and suspension in early education as matters of social justice and health equity: NAM Perspectives [Discussion paper]. *National Academy of Medicine*. <https://doi.org/10.31478/201610e>
- *Metz, M. (2018). Exploring the complexity of high school students' beliefs about language variation. *Linguistics and Education*, 45, 10–19. <https://doi.org/10.1016/j.linged.2018.02.003>
- *Metz, M. (2019). Accommodating linguistic prejudice? Examining English teachers' language ideologies. *English Teaching: Practice & Critique*, 18(1), 18–35. <https://doi.org/10.1108/ETPC-09-2018-0081>
- *Milner, H. R. (2017). Race, talk, opportunity gaps, and curriculum shifts in (teacher) education. *Literacy Research: Theory, Method, and Practice*, 66(1), 73–94. <https://doi.org/10.1177/2381336917718804>
- Modaffari, J., & Jimenez, L. (2021, May 5). The education data that matter most to parents and school stakeholders. *Center for American Progress*. Retrieved May 13, 2024, from <https://www.americanprogress.org/article/education-data-matter-parents-school-stakeholders/>
- Monmouth University. (2020, June 2). *Protestors' anger justified even if actions may not be*. Monmouth University Polling Institute. https://www.monmouth.edu/polling-institute/reports/monmouthpoll_us_060220/
- Murray, C., & Herrnstein, R. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Nadal, K. L. (2011). The Racial and Ethnic Microaggressions Scale (REMS): Construction, reliability, and validity. *Journal of Counseling Psychology*, 58(4), 470. <https://doi.org/10.1037/a0025193>
- National Center for Education Statistics. (n.d.). *History and innovation of assessment design*. Retrieved May 5, 2024, from <https://nces.ed.gov/nationsreportcard/about/timeline.aspx>
- Neville, H. A., Lilly, R. L., Duran, G., Lee, R. M., & Browne, L. (2000). Construction and initial validation of the Color-Blind Racial Attitudes Scale (CoBRAS). *Journal of Counseling Psychology*, 47(1), 59–70. <https://doi.org/10.1037/0022-0167.47.1.59>
- Nicholson-Crotty, S., Grissom, J. A., Nicholson-Crotty, J., & Redding, C. (2016). Disentangling the causal mechanisms of representative bureaucracy: Evidence from assignment of students to gifted programs. *Journal of Public Administration Research and Theory*, 26(4), 745–757. <https://doi.org/10.1093/jopart/muw024>
- *Nieri, T., Ayón, C., Yoo, M., & Webb, M. (2022). Perceived ethnic discrimination, ethnic-racial socialization, and substance use among ethnic minority adolescents. *Journal of Ethnicity in Substance Abuse*, 21(1), 70–89. <https://doi.org/10.1080/15332640.2019.1707141>
- Noguera, P. A. (2003). The trouble with Black boys: The role and influence of environmental and cultural factors on the academic performance of African American males. *Urban Education*, 38(4), 431–459. <https://doi.org/10.1177/0042085903038004005>
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3, 1–18. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- *Oberoi, A. K., & Trickett, E. J. (2018). Religion in the hallways: Academic performance and psychological distress among immigrant origin Muslim adolescents in high schools. *American Journal of Community Psychology*, 61(3–4), 344–357. <https://doi.org/10.1002/ajcp.12238>
- 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>
- Olofsson, H., Brolund, A., Hellberg, C., Silverstein, R., Stenström, K., Österberg, M., & Dagerhamn, J. (2017). Can abstract screening workload be reduced using text mining?

- User experiences of the tool Rayyan. *Research Synthesis Methods*, 8(3), 275–280. <https://doi.org/10.1002/jrsm.1237>
- Oto, R., Rombalski, A., & Grinage, J. (2023). The role of racial literacy in US K-12 education research: A review of the literature. *Race Ethnicity and Education*, 26(1), 94–111. <https://doi.org/10.1080/13613324.2022.2047635>
- Paat, Y. F. (2013). Working with immigrant children and their families: An application of Bronfenbrenner's ecological systems theory. *Journal of Human Behavior in the Social Environment*, 23(8), 954–966. <https://doi.org/10.1080/10911359.2013.800007>
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Lawrence Erlbaum Associates Publishers.
- *Pena-Shaff, J. B., Bessette-Symons, B., Tate, M., & Fingerhut, J. (2019). Racial and ethnic differences in high school students' perceptions of school climate and disciplinary practices. *Race Ethnicity and Education*, 22(2), 269–284. <https://doi.org/10.1080/13613324.2018.1468747>
- Pérez Huber, L., Vélez, V. N., & Solórzano, D. (2018). More than 'papelitos': a QuantCrit counter-story to critique Latina/o degree value and occupational prestige. *Race Ethnicity and Education*, 21(2), 208–230. <https://doi.org/10.1080/13613324.2017.1377416>
- Peterson, R. D. (2014). Examining the intersection of race, gender, and environment using Bronfenbrenner's bioecological systems theory of human development. *Theoretical Frameworks in Qualitative Research*, 80(1), 34–58.
- Phinney, J. S., Madden, T., & Santos, L. J. (1998). Psychological variables as predictors of perceived ethnic discrimination among minority and immigrant adolescents. *Journal of Applied Social Psychology*, 28(11), 937–953. <https://doi.org/10.1111/j.1559-1816.1998.tb01661.x>
- Phinney, J. S., & Ong, A. D. (2007). Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology*, 54(3), 271–281. <https://doi.org/10.1037/0022-0167.54.3.271>
- *Pieterse, A. L., Utsey, S. O., & Miller, M. J. (2016). Development and initial validation of the anti-racism behavioral inventory (ARBI). *Counselling Psychology Quarterly*, 29(4), 356–381. <https://doi.org/10.1080/09515070.2015.1101534>
- Pohan, C. A., & Aguilar, T. E. (1994). *Initial development and preliminary testing of two beliefs about diversity measures*. University of Nebraska, Center for Curriculum and Instruction, Teachers College.
- Pohan, C. A., & Aguilar, T. E. (2001). Measuring educators' beliefs about diversity in personal and professional contexts. *American Educational Research Journal*, 38(1), 159–182. <https://doi.org/10.3102/00028312038001159>
- Polanin, J. R., Edwards, I., & Peko-Spicer, S. (2024). *Considerations for centering diversity, equity, and inclusion in research syntheses*. Methods of Synthesis and Integration Center. https://www.air.org/sites/default/files/2024-04/DEI_Considerations.pdf
- Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotper, J. K. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Research Synthesis Methods*, 10(3), 330–342. https://krimdok.uni-tuebingen.de/Record/1912467348?utm_source=chatgpt.com
- Polos, J. A., Koning, S. M., Hargrove, T. W., Kershaw, K. N., & McDade, T. W. (2022). Structural racism in school contexts and adolescent depression: Development of new indices for the National Longitudinal Study of Adolescent to Adult Health and beyond. *SSM-Population Health*, 19, 101237. <https://doi.org/10.1016/j.ssmph.2022.101237>

- Race Forward. (2023). *What is racial equity?* <https://www.raceforward.org/resources/what-racial-equity-1>
- Randall, J. (2021). "Color-neutral" is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues and Practice*, 40, 82–90. <https://doi.org/10.1111/emip.12429>
- Reardon, S. F., Valentino, R. A., Kalogrides, D., Shores, K. A., & Greenberg, E. H. (2013). *Patterns and trends in racial academic achievement gaps among states, 1999–2011*. Stanford Center for Education Policy Analysis. <https://cepa.stanford.edu/content/patterns-and-trends-racial-academic-achievement-gaps-among-states-1999-2011>
- *Riskowski, J. L. (2010) Students views of diversity: Viewing transformation during the middle school years. *Multicultural Education*, 17, 2–12.
- *Romero, A. J., & Roberts, R. E. (2003a). Stress within a bicultural context for adolescents of Mexican descent. *Cultural Diversity and Ethnic Minority Psychology*, 9(2), 171–184. <https://psycnet.apa.org/doi/10.1037/1099-9809.9.2.171>
- *Romero, A. J., & Roberts, R. E. (2003b). The impact of multiple dimensions of ethnic identity on discrimination and adolescents' self-esteem. *Journal of Applied Social Psychology*, 33(11), 2288–2305. <https://doi.org/10.1111/j.1559-1816.2003.tb01885.x>
- Rosa, E. M., & Tudge, J. (2013). Urie Bronfenbrenner's theory of human development: Its evolution from ecology to bioecology. *Journal of Family Theory & Review*, 5(4), 243–258. <https://doi.org/10.1111/jftr.12022>Digital Object Identifier (DOI)
- Saldaña, J. (2018). *The coding manual for qualitative researchers* (3rd ed.). SAGE publications.
- *Saleem, F., Lambert, S., & Rose, T. (2022). Ethnic–racial socialization as a moderator of associations between discrimination and psychosocial well-being among African American and Caribbean Black adolescents. *Cultural Diversity & Ethnic Minority Psychology*, 28(2), 145–157. <https://doi.org/10.1037/cdp0000521>
- Schimmack, U. (2021a). The Implicit Association Test: A method in search of a construct. *Perspectives on Psychological Science*, 16(2), 396–414. <https://doi.org/10.1177/1745691619863798>
- Schimmack, U. (2021b). Invalid claims about the validity of Implicit Association Tests by prisoners of the implicit social-cognition paradigm. *Perspectives on Psychological Science*, 16(2), 435–442. <https://doi.org/10.1177/1745691621991860>
- Schooley, R. C., Lee, D. L., & Spanierman, L. B. (2019). Measuring Whiteness: A systematic review of instruments and call to action. *The Counseling Psychologist*, 47(4), 530–565. <https://doi.org/10.1177/0011000019883261>
- Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). Routledge/Taylor & Francis Group.
- *Seaton, E. K. (2003). An examination of the factor structure of the Index of Race-Related Stress among a sample of African American adolescents. *Journal of Black Psychology*, 29(3), 292–307. <https://doi.org/10.1177/0095798403254211>
- *Seaton, E. K. (2006). Examination of a measure of racial discrimination among African American adolescents. *Journal of Applied Social Psychology*, 36(6), 1414–1429. <https://doi.org/10.1111/j.0021-9029.2006.00066.x>
- *Seaton, E. K. (2010). The influence of cognitive development and perceived racial discrimination on the psychological well-being of African American youth. *Journal of Youth and Adolescence*, 39(6), 694–703. <https://doi.org/10.1007/s10964-009-9438-4>
- Seaton, E. K., Caldwell, C. H., Sellers, R. M., & Jackson, J. S. (2008). The prevalence of perceived discrimination among African American and Caribbean Black youth. *Developmental Psychology*, 44(5), 1288–1297. <https://doi.org/10.1037/a0012747>

- *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>
- Sellers, R. M., Copeland-Linder, N., Martin, P. P., & Lewis, R. L. (2006). Racial identity matters: The relationship between racial discrimination and psychological functioning in African American adolescents. *Journal of Research on Adolescence*, 16, 187–216. <https://doi.org/10.1111/j.1532-7795.2006.00128.x>
- *Seider, S., Kelly, L., Clark, S., Jennett, P., El-Amin, A., & Graves, D. (2018). Fostering the sociopolitical development of African American and Latinx adolescents to analyze and challenge racial and economic inequality. *Youth and Society*, 52(5), 756–794. <https://doi.org/10.1177/0044118X18767783>
- *Sharma, S., Mustanski, B., Dick, D., Bolland, J., & Kertes, D. A. (2019). Protective factors buffer life stress and behavioral health outcomes among high-risk youth. *Journal of Abnormal Child Psychology*, 47(8), 1289–1301. <https://doi.org/10.1007/s10802-019-00515-8>
- *Simons, R. L., Chen, Y.-f., Stewart, E. A., & Brody, G. H. (2003). Incidents of discrimination and risk for delinquency: A longitudinal test of strain theory with an African American sample. *Justice Quarterly*, 20(4), 827–854. <https://doi.org/10.1080/07418820300095711>
- Skinner-Dorkenoo, A. L., George, M., Wages, J. E., Sánchez, S., & Perry, S. P. (2023). A systemic approach to the psychology of racial bias within individuals and society. *Nature Reviews Psychology*, 2, 392–406. <https://doi.org/10.1038/s44159-023-00190-z>
- Skoog-Hoffman, A., Miller, A. A., Plate, R. C., Meyers, D. C., Tucker, A. S., Meyers, G., Diliberti, M. K., Schwartz, H. L., Kuhfeld, M., Jagers, R. J., Steele, L., & Schlund, J. (2024). *Social and emotional learning in U.S. schools: Findings from CASEL's nationwide policy scan and the American Teacher Panel and American School Leader Panel surveys*. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1822-2.html
- Smithsonian. (n.d.). *Talking about race: Being anti-racist*. Retrieved March 4, 2024, from <https://nmaahc.si.edu/learn/talking-about-race/topics/being-antiracist>
- Stage, F. K. (2007). Answering critical questions using quantitative data. *New Directions for Institutional Research*, 133, 5–16. <https://doi.org/10.1002/ir.200>
- Stern, J. A., Barbarin, O., & Cassidy, J. (2022). Working toward anti-racist perspectives in attachment theory, research, and practice. *Attachment & Human Development*, 24(3), 392–422. <https://doi.org/10.1080/14616734.2021.1976933>
- *Stevenson, H. C., Aisenbrey, E. A., Mejia-Bradford, S. C., & Rovine, M. J. (2024). RECASTing racial coping stress in school: Self-efficacy as a buffer for adolescent agency. *Cultural Diversity & Ethnic Minority Psychology*, 30(3), 587–598. <https://doi.org/10.1037/cdp0000647>
- Stringer, K. (2019). SEL in the House: Democrats approve millions in landmark federal funding for social-emotional learning in bill that now faces test in Senate. *The 74 Million*, June 20. <https://www.the74million.org/sel-in-the-house-democrats-approve-millions-in-landmark-federal-funding-for-social-emotional-learning-in-bill-that-now-faces-test-in-senate/>
- Strunk, K. K., & Andrzejewski, C. E. (2023). Racisms of commission and omission in educational psychology: A historical analysis and systematic review. *Educational Psychologist*, 58(2), 57–69. <https://doi.org/10.1080/00461520.2022.2152031>
- Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, 62(4), 271–286. <https://doi.org/10.1037/0003-066X.62.4.271>

- Sugarman, D. B., Nation, M., Yuan, N. P., Kuperminc, G. P., Hassoun Ayoub, L., & Hamby, S. (2018). Hate and violence: Addressing discrimination based on race, ethnicity, religion, sexual orientation, and gender identity. *Psychology of Violence, 8*(6), 649–656. <https://doi.org/10.1037/vio0000222>
- Terman, L. (1961). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale*. Houghton Mifflin.
- Torres, L., Driscoll, M. W., & Burrow, A. L. (2010). Racial microaggressions and psychological functioning among highly achieving African-Americans: A mixed-methods approach. *Journal of Social and Clinical Psychology, 29*(10), 1074–1099. <https://doi.org/10.1521/jscp.2010.29.10.1074>
- Utsey, S. O., Belvet, B., Hubbard, R. R., Fischer, N. L., Opore-Henaku, A., & Gladney, L. L. (2013). Development and validation of the prolonged activation and anticipatory race-related stress scale. *Journal of Black Psychology, 39*(6), 532–559. <https://doi.org/10.1177/0095798412461808>
- USAFacts Team. (2025, June 2). Who are the nation's teachers? Key insights in five charts. *USAFacts*. <https://usafacts.org/articles/who-are-the-nations-teachers/>
- U.S. Department of the Treasury. (2023, June 9). *Post 5: Racial differences in educational experiences and attainment*. U.S. Department of the Treasury. <https://home.treasury.gov/news/featured-stories/post-5-racial-differences-in-educational-experiences-and-attainment>
- *Vargas, S. M., Calderon, V., Beam, C. R., Cespedes-Knadle, Y., & Huey, S. J., Jr. (2021). Worse for girls?: Gender differences in discrimination as a predictor of suicidality among Latinx youth. *Journal of Adolescence, 88*, 162–171. <https://doi.org/10.1016/j.adolescence.2021.02.007>
- Vesely, C. K., Sansbury, A. B., Call-Cummings, M., Dodman, S., Chen, X., N'Diaye, N., Bethea, C., DeMulder, E., Frank, T., Letiecq, B., Shaklee, B., & View, J. (2023). Putting antiracism into action in teacher education: Developing and implementing an antiracist pedagogy course audit. *Journal of Early Childhood Teacher Education, 44*(4), 943–961. <https://doi.org/10.1080/10901027.2023.2192433>
- *Vigna, A. J., Poehlmann-Tynan, J., & Koenig, B. W. (2018). Does self-compassion covary with minority stress? Examining group differences at the intersection of marginalized identities. *Self and Identity, 17*(6), 687–709. <https://doi.org/10.1080/15298868.2018.1457566>
- Vinokurov, A., Trickett, E. J., & Birman, D. (2002). Acculturative hassles and immigrant adolescents: A life-domain assessment for Soviet Jewish refugees. *The Journal of Social Psychology, 142*(4), 425–445. <https://doi.org/10.1080/00224540209603910>
- *Wang, M.-T., & Huguley, J. P. (2012). Parental racial socialization as a moderator of the effects of racial discrimination on educational success among African American adolescents. *Child Development, 83*(5), 1716–1731. <https://doi.org/10.1111/j.1467-8624.2012.01808.x>
- *Wang, W., Spinrad, T. L., Laible, D. J., Janssen, J., Xiao, S. X., Xu, J., Berger, R. H., Eisenberg, N., Carlo, G., Gal-Szabo, D. E., Fraser, A., Lopez, J., & Xu, X. (2023). Parents' color-blind racial ideology and implicit racial attitudes predict children's race-based sympathy. *Journal of Family Psychology, 37*(4), 475–485. <https://doi.org/10.1037/fam0001047>
- Weiss, S. F. (2010). After the fall: Political whitewashing, professional posturing, and personal refashioning in the postwar career of Otmar Freiherr von Verschuer. *Isis, 101*(4), 722–758. <https://doi.org/10.1086/657474>
- West, C. P., Dyrbye, L. N., Erwin, P. J., & Shanafelt, T. D. (2016). Interventions to prevent and reduce physician burnout: A systematic review and meta-analysis. *Lancet, 388*(10057), 2272–2281. [https://doi.org/10.1016/S0140-6736\(16\)31279-X](https://doi.org/10.1016/S0140-6736(16)31279-X)

- *Williams, A. D., Bigler, R. S., & Ramirez, M. C. (2023). Latinx children's race- and ethnicity-related identities, beliefs, and preferences. *Group Processes & Intergroup Relations*, 26(1), 120–139. <https://doi.org/10.1177/13684302211050553>
- Williams, D. R., Yan, Yu, Jackson, J. S., & Anderson, N. B. (1997). Racial differences in physical and mental health: Socio-economic status, stress and discrimination. *Journal of Health Psychology*, 2(3), 335–351. <https://doi.org/10.1177/135910539700200305>
- Yang, Y., & Green, S. B. (2011). Coaching and performance: Is Cronbach's alpha sufficient? *Educational and Psychological Measurement*, 71(4), 661–676.
- Zawacki-Richter, O. (2020). Introduction: Systematic reviews in educational research. In M. Newman & D. Gough (Eds.), *Systematic reviews in educational research* (pp. v–xiv). Springer.
- Zuberi, T. (2001). *Thicker than blood: How racial statistics lie*. University of Minnesota Press.

Author Biographies

Wendy Castillo is an assistant professor in educational foundations at the College of Education and Engaged Learning (CEEL) at Montclair State University, 1 Normal Ave, Montclair, NJ 07043; castillow@montclair.edu. Her research interests lie at the intersection of race and quantitative data and methods, including QuantCrit and other racial equity frameworks, to advance and improve the lives of students of color.

Juan Cruz is a PhD student in education policy at Harvard University, 13 Appian Way, Cambridge, MA 02138; email: juan_cruz@g.harvard.edu. His research focuses on access to and attainment in higher education, heterogeneity treatment effects, and psychometric methods for measuring educational experiences and outcomes.

Rachel Renbarger is a research scientist at FHI 360, 359 Blackwell Street, Suite 200, Durham, NC 27701; email: rachelrenbarger@gmail.com. Her research focuses on improving educational systems in K–12 and higher education for all populations and improving research practices.

Christen Priddie is an assistant research scientist at Indiana University Bloomington, 107 S. Indiana Ave., Bloomington, Indiana, 47405; cpriddie@iu.edu. Her research focuses on using critical quantitative methods to analyze Black people's experiences in higher education.

Brein Mosely is a PhD candidate at Harvard University, 13 Appian Way, Cambridge, MA 02138; email: breinmosely@g.harvard.edu. Her research focuses on the intersection of race, mental health, and power.

Sasha Mejia-Bradford a PhD candidate in interdisciplinary studies in human development at Penn GSE, 3700 Walnut St, Philadelphia, PA 19104; scmejia@upenn.edu. Her current research investigates the role of cultural values in mitigating the effects of racial-ethnic discrimination (RED) stress among Latino adolescents, utilizing wearable technology to examine their physiological responses to racism.