EXPLORING THE METHODOLOGICAL APPROACHES TO STUDYING RACIAL IMPLICIT BIAS OF TEACHERS IN PK–12 EDUCATION

Isabella Camfield

This article seeks to highlight and evaluate the different methodological approaches to researching teachers' racial implicit bias in an educational environment. It is evident that much of the research on this topic has focused primarily on examining the prevalence of implicit bias, understanding the various aspects of the schooling system that it impacts (including both grading and behavior), and exploring different ways to counteract it. Research on the topic has included both the qualitative methods of ethnography and phenomenological research and the primarily quantitative approaches of survey design, correlational design, and experimental design. After an overview of the methodologies of existing literature on teachers' racial implicit bias, this review will then explore some of the ethical considerations of these empirical studies. Next, the review will discuss the methodological strengths of instrument choice, pilot surveys, and accounting for other variables, and the weaknesses of unclear sample size reasoning, lack of generalizability, self-reported data, and opt-in participation. Lastly, suggestions for future research will be offered that emerge from gaps in existing literature, including intersections between implicit bias and artificial intelligence, implicit bias and anti-critical race theory policy, and implicit bias and growth mindset.

Keywords: PK–12 education, racial bias, implicit bias, teacher bias

Racial implicit bias is an important and growing topic of study because the internalized beliefs of teachers can unknowingly impact both the academic performance and social wellbeing of students of color. As defined by the American Psychological Association (2018b), implicit prejudice—which I liken to implicit bias—is "a negative attitude, of which one is not consciously aware, against a specific social group" (para. 1). Thus, individuals can have implicit bias regarding gender, socioeconomic status, ethnicity, race, and more. However, this review will focus specifically on teacher racial implicit bias because of the potential impact of the increasingly divisive rhetoric surrounding race throughout the country (Agathocleous et al., 2024; Amnesty International, 2017).

Behavioral researcher De Houwer (2019) explained that because implicit bias is a behavioral response, compared to just a hidden force inside people, it can be lessened over time with the right intervention. De Houwer further explained that there are two ways to study implicit bias:

> Behavioral research on implicit bias can be seen as directed at documenting the environmental conditions under which implicit group-based behavior occurs (i.e., the

> moderators of implicit bias), [and] cognitive research on implicit bias can be seen as focused on documenting the mental processes that mediate implicit group-based behavior (i.e., the mental mediators of implicit bias). (p. 838)

With this in mind, stakeholders need to have a better understanding of what can be done to help educators become more aware of the issue; we must examine both the environmental factors and teachers' individual mental processes at play, and we must try to diminish the harmful impact it can cause. However, in order to do that, we need to collect accurate and appropriate information on the topic. In other words, *how* teachers' racial bias is studied has a significant impact on the recommendations made to mitigate its effects. This literature review explores the contours of those studies.

Method

This review of the literature gathered 30 empirical studies on teachers' racial implicit bias to explore common themes, identify common errors, and then determine areas for further study. Specifically, I searched for articles that explored how teachers' racial implicit bias towards their students can be studied and controlled, narrowing the field to include only articles published within the past 10 years to increase relevance to today's educational landscape. I intentionally focused on teacher bias in PK–12 education because earlier influences on students of color can impact life outcomes (e.g., whether they choose to pursue higher education). However, the goal was also to obtain the highest level and widest range of information possible within those parameters.

Utilizing an institutional library search function, which accesses 531 databases (e.g., Taylor and Francis Online, ProQuest, SpringerLink, and JSTOR), I began by searching "implicit bias and education." This initially identified 4,566 results. To narrow it down, I added *racial* implicit bias and education and then also searched "teacher racial implicit bias." I then limited it further to only include peer-reviewed articles and dissertations. Because it appears to be a newer topic of exploration, I deemed dissertation research to be of value. (Ultimately, eight of the 30 studies gathered were dissertations.) Many of the remaining studies identified in the search results explored medical students' implicit biases, so further sorting to find articles devoted to K–12 education and teachers was necessary. Finally, I sought articles that performed empirical studies and were not just theoretical so that the methodological approaches could be examined.

Each study was added to a table that tracked the different data sources that the study engaged, the collection methods employed, the possible ethical concerns for the study, and strengths and weaknesses of the study's application of the method; color-codes track themes in each category (see Appendix). While it is certain that not every study on the topic was included in examination, I determined that saturation was reached in identifying methodological themes in the literature.

Findings

Correlational Design Approaches

The first method used in studying teacher implicit bias is correlational design where researchers try to find a relationship between two or more variables. According to Creswell and Guetterman

(2019), this design is beneficial for relating "two or more variables to see if they are associated with each other… [and/or] to predict an outcome" (p. 343).

Bias and Student Performance

One example of this design is from Chin et al. (2020) who examined data from multiple tests of implicit bias: The Civil Rights data collection archive, the Stanford Education data archive, and Project Implicit's data on scores on the Implicit Association Test (IAT) to explore the relationship between teacher implicit bias and student outcomes. Thus, the variables were teacher implicit bias and student performance, attempting to support the assertion that as the bias increases, performance decreases.

Similarly to Chin et al. (2020) in their strategy of drawing from existing data, both Mortenson (2018) and Mekky (2023) used correlational design to examine the relationship between teacher bias and student outcomes but in quite different contexts. Mortenson examined preexisting data from the Education Longitudinal Study of 2002 and state test scores to examine the correlation between teacher implicit bias and student scores, trying to identify a factor for the racial achievement gap in testing. In contrast, Mekky gathered new data from teachers using the IAT and a survey with demographic data to determine teachers' perceptions and how implicit bias could impact the performance of their students.

Bias and Teacher Internal Bandwidth and Mindsets

Utilizing another perspective, Costa et al. (2023) examined the relationship between teacher implicit bias and teacher burnout, attempting to support the assertion that when a teacher is more burned out, they will show more implicit bias. This aligns with research on cognitive load, which is defined by the American Psychological Association (2018a) as "the relative demand imposed by a particular task, in terms of mental resources required" (para. 1). The logic is that when a task is overloading cognitive abilities as can be experienced in burnout, people are more likely to fall into implicit biases because they cannot do the cognitive work to catch themselves, and this study showed that to be the case with teachers, using the IAT and Relational Responding Task to measure implicit bias and the Maslach Burnout Inventory to measure implicit bias.

In contrast, Kumar et al. (2022) and Stephens et al. (2022) both used correlational design to examine different aspects of teacher's mindsets regarding implicit bias. Both studies were focused on preservice teachers to explore what their biases are and whether they could change. Kumar et al.'s study examined the relationship between implicit bias and teacher's belief in culturally responsive pedagogy. If a preservice teacher tested higher on the IAT, they were less likely to espouse the beliefs of cultural responsibility and respect. Extending this idea, Stephens et al. examined how the relationship between bias and intelligence mindset regarding minority students may change over time after they have spent more time with the students (in this case, Māori students). To measure changes, the researchers in both studies asked teachers to complete both a longitudinal and cross-sectional survey. The researchers did find that after preservice teachers have actually spent more time engaging with their students, their implicit bias scores went down.

Bias and Student Discipline

In examining the effects of teacher bias, there is not only an effect on teacher-student relationships to explore but also correlational implications between bias and discipline. Shirrell et al. (2024) examined the correlation between teacher's implicit bias and the way that students were disciplined. They used preexisting data from New York Public Schools' personnel files and administrative student outcome files to determine whether there was a difference in the disciplinary outcome depending on the race of the teacher and student.

Lastly, Starck et al. (2020) examined the data from the Project Implicit database to determine whether there is a correlation in bias for teachers compared to other groups. They found that teachers had either the same or less implicit bias than the general population.

Experimental Design Approaches

Experimental design for implicit bias has focused on overall interventions to reduce implicit bias and also more specific strategies for reducing its impact (e.g., on grading and behavior). Compared to correlational design where the researcher is just trying to establish that there is a relationship between two variables, experimental design helps "establish possible cause and effect between your independent and dependent variables" (Creswell & Guetterman, 2019, p. 295).

Identifying the Most Effective Bias Intervention

Lai et al. (2014) and Norris (2022) conducted experiments that compared different interventions to determine which was most effective at reducing implicit bias. Some of these interventions included education, perspective-taking, exposure, and evaluative conditioning. They had control and experimental groups (each receiving different interventions), and the researchers examined the effects on the dependent variable (their scores on implicit bias measures) between the two groups.

Testing an Intervention

Other studies focused on one specific intervention. Mossing et al. (2025) focused on teacher coaching to decrease implicit bias towards Black students, measuring bias before and after the coaching. Whitford and Emerson (2019) attempted an empathy intervention to reduce bias again with a pretest and posttest using the IAT to measure their success. Larson et al. (2024) attempted to use contemplative practices in a mixed methods experiment that involved interviews with some of the participants afterwards to gather a deeper understanding of their individual experiences. Pepis (2022) compared a group of teachers that received an intervention on implicit bias compared to a group that did not, examining the effects of stereotype threat and colorblindness. Essentially, the group that received the intervention was uncomfortable with the idea that they may hold implicit bias and slipped into the problematic refrain of "I don't see color." Lastly, Gagliardi (2015) utilized impact awareness as an intervention against implicit bias, using a pre-post measurement framework with the Affect Misattribution Procedure that tests the level of pleasantness participants feel when viewing various racialized stimuli. Although Gagliardi's work included an experiment, it also incorporated grounded theory in interviewing the participants on their responses to sample student vignettes.

Identifying the Severity of the Bias Issue

Instead of focusing on implicit attitudes as a whole, both Quinn (2020) and Johnson (2020) used a student-names-based experiment to look specifically at how implicit bias impacts grading and assessment. Quinn was more focused on proving that this bias exists. In their experiment, they gave teachers a student writing sample and found that when the student name was assumed to be that of an African-American student, they were graded more harshly than if the sample was labeled with a name that was perceived to be that of a White student. Johnson, using a similar experimental design, examined the lack of teacher recommendation for Black male students to take advanced classes, proving an implicit bias that affects grades and educational outcomes.

Approaching the topic of teacher bias from a discipline and behavior angle, Ura and d'Abreu (2022) and Jacoby-Senghor et al. (2016) used different forms of data collection. Ura and d'Abreu used vignette scenarios of student behavior and asked teachers how they would intervene. In one condition, the student's name was a "White name" and in the other it was a "Black name" to see if teachers' decision making would be different depending on the race of the student. In contrast, Jacoby-Senghor et al. actually observed teachers in real-time in the classroom to examine the behavior of teachers and students when the instructor had different amounts of implicit bias as predetermined by their results on a subliminal priming task (Dovidio et al., 1997, 2002, as cited in Jacoby-Senghor et al., 2016).

Survey Design Approaches

Survey designs are beneficial when a researcher wants to "identify important beliefs and attitudes of individuals" (Creswell & Guetterman, 2019, p. 385) or to identify larger trends within a group of people. Surveys in this topic seemed to have two major foci: (a) examining teachers' fear of appearing racist and their social desirability, and (b) the effect of implicit bias on teacher-student relationships. To clarify, social desirability is the "tendency for people to present themselves in a generally favorable fashion … particularly within the field of self-report assessment of personality and attitudes" (Holden & Passey, 2009, p. 441). Many of the surveys under examination here also used a more mixed-methods approach, incorporating open-ended questions that were coded for themes along with more quantitative data.

Exploring Teachers' Comfort Level Discussing Bias Issues

Tropp and Rucinski (2022) used surveys to examine the larger trend of how teachers' fear of seeming racist and their implicit biases impact the way they discuss race with their students. These researchers surveyed two different groups of teachers: one that was using curriculum from the Teaching Tolerance website and one from an urban school district. In both cases, they asked participants to answer questions regarding their concerns of appearing racist, their intention and confidence in engaging the topic of race, their social desirability, support, and motives.

Understanding Specifically How Social Desirability Intersects with Implicit Bias

Ayala et al. (2024) and Marcucci (2020) focused their surveys specifically on the question of social desirability and how it connects to implicit bias. Ayala et al. (2024) used a control group whose survey had noncontroversial statements to compare to the group receiving questions to assess

implicit bias. They found that there was a significant bias in teachers' beliefs about immigrant students' abilities but that social desirability did not significantly sway them. Marcucci (2020), on the other hand, used vignettes about student behavior as a tool to measure bias and found that teachers because of social desirability were actually harder on the White students. This, of course, is not the case in actual discipline cases throughout the country. Their work highlights the insidious nature of implicit bias because social desirability prevents honest reflection from practitioners.

Examining Specific Impacts on Teacher-Student Relationships

The impact of implicit bias on teacher-student relationships was explored by Childs (2023) and Blackson et al. (2022). Both of the surveys they used were implemented online, and both included multiple scales to determine racial preferences and implicit bias. Both surveys utilized the IAT as one of their measures. Childs was focused on all PK–12 public school teachers in the United States, included open-ended questions, and included data about the impact of teacher-student racial mismatch to explore its impact on bias and relationships. Blackson et al., in contrast, focused specifically on early childhood educators and whether their bias towards students of color is impacting their ability to form strong relationships. This is, of course, an important group to study since it is a child's first schooling experience and thus their treatment from educators can have a lasting impact. In addition to the IAT, Blackson et al. used a racial preference scale and temperature scale to measure explicit bias along with a questionnaire about perceptions on racial disparities in education.

Ethnographic Design Approaches

Ethnography allows researchers to develop "procedures for describing, analyzing, and interpreting a culture-sharing group's shared patterns of behavior, beliefs, and language that develop over time" (Creswell & Guetterman, 2019, p. 474). Ethnographic studies on implicit bias consist of case studies examining the perceived impact of training to reduce bias and field studies to examine how the issue plays out in classrooms.

Studying Postintervention Results

A case study was conducted by Crutchfield et al. (2022) examining the skin tone bias of preservice teachers using the IAT, interviews, and journals. Participants were asked to reflect on their IAT results to explore the impact of awareness-building. In addition, both Palmer (2023) and Lindvall-Östling (2024) in their separate studies interviewed teachers after they had engaged in a specific training intended to address implicit bias. Their intention was to understand how teachers perceived the intervention went, how effective it was, and how lasting an impact it had on their practice. However, Lindvall-Östling's (2024) specific training was on gendered implicit bias, not examining race. Even so, this study might still be seen as relevant because their actual focus was on how lasting implicit bias training can be and how much training can actually influence practice.

Observing the Prevalence and Effects of Bias

In contrast to the previous section, İnan-Kaya and Rubie-Davies (2022) performed a qualitative study to provide a sense of what is occurring in classrooms in general, not just after a particular

intervention. They examined both verbal and nonverbal behaviors of teachers to identify indicators of bias towards students of color, utilizing field notes to examine the behavior of teachers over multiple class periods. They found trends in teacher behaviors to identify the difference in treatment of students of color and their White peers. Thus, their focus was not so much on how to address the issue but more to highlight the connection between implicit biases and explicit behavior differences, emphasizing the severity of the problem and raising awareness.

Other Qualitative Designs

There were several other qualitative studies that do not fit into any of the specific designs identified above. For instance, Yeung (2022) used a series of open-ended questions to find out how White teachers perceive their Black students. Young wanted to deliver this questionnaire on paper instead of conducting observations or interviews to avoid the potential for social desirability bias as mentioned previously. Thus, they categorize their work as a *phenomenological qualitative study*. Phenomenological research is similar to ethnography but focuses more on individuals than a group. The idea is to focus "on the study of an individual's lived experiences within the world" (Neubauer et al., 2019, p. 90).

Lastly, Chiu et al. (2022) conducted a *DisCrit qualitative framework* to study the implicit biases that occur in special education classrooms, identifying an important intersection between critical race theory and critical disability studies. Their research methods included the IAT and participant reflections to explore the issue. This is particularly important due to the intersectionality between race and special education designation as Black students are overrepresented in special education (Hines et al., 2021).

Discussion and Analysis

As mentioned in the introduction of this literature review, it is imperative to understand the cause and the impact of teachers' racial implicit bias. However, before concluding anything about that topic, we must ensure that we have collected the best data possible from which to draw conclusions. Each of the aforementioned research structures and their associated methods have affordances and potential pitfalls researchers must guard against. Therefore, this section will examine ethical considerations of these studies to evaluate the merits of these various approaches and to determine gaps in the literature.

Ethics Considerations for Research Done on Teachers' Racial Implicit Bias

Institutional Review Board Approval Achieved

Seventeen of the studies under examination explicitly mentioned following the Institutional Review Board (IRB) process before conducting their research. This is integral to determining that there was informed consent gleaned from participants. According to the Collaborative Institutional Training Initiative,

> researchers are required to provide information in a manner understandable to the subjects. The regulations emphasize that consent must begin with a "concise and focused

> presentation" to the subject and provide information that would help the subject determine whether to participate. (Hicks, 2019)

However, nine studies (Ayala et al., 2024; Blackson et al., 2022; Jacoby-Senghor et al., 2016; Lai et al., 2014; Lindvall-Östling, 2024; Marcucci, 2020; Pepis, 2022; Quinn, 2020; Tropp & Rucinski, 2022) did not specify how they went through the IRB process though most still mentioned informed consent. In addition, four other studies (Chin et al., 2020; Mortenson, 2018; Shirrell et al., 2024; Starck et al., 2020) did not need IRB approval because they were working with previously released data that was de-identified.

Use of Incentives

In addition, many of the studies used incentives to get participants to be involved. For some, it was extra credit in a course (e.g., Gagliardi, 2015; Jacoby-Senghor et al., 2016). For others, it was monetary compensation (e.g., Childs, 2023; Kumar et al., 2022). Still others had all students in a course participate in an experiment, but students could choose whether they wanted their data to be included or not in the paper (Chiu et al., 2022; Pepis, 2022). Obviously, in all of these situations, it is important to consider how these decisions could potentially have an impact on results. In particular, it would be helpful to know whether Pepis and Chiu et al. had participants who decided to rescind consent after the results were gathered since they were asking their whole class to be involved. There is definitely a concern for conflict of interest or for students feeling pressured to participate in those circumstances.

Use of Already Released Data

Four of the studies utilized data that had already been collected and was therefore already de-identified and public. This included Mortenson (2018), Johnson (2020), Starck et al. (2020), and Shirrell et al. (2024) and involved data such as enrollment numbers, standardized test scores, and the overall data from the Project Implicit website. This meant that the studies did not need to go through the same type of IRB approval because they were not generating new data with human subjects. Therefore, there is less of a likelihood for ethical concerns with these data regarding the potential risk to individuals (e.g., there is no chance of breaking confidentiality).

Member Checking

The case study by Crutchfield et al. (2022) used member checking to ensure accuracy and ethics. Member checking involves asking participants to review the transcripts of what was recorded in their interviews and make any changes that they would like before the data are published. According to Creswell and Guetterman (2019), member checking means asking "participants about many aspects of the study, such as whether the description is complete and realistic, if the themes are accurate to include, and if the interpretations are fair and representative" (p. 261). Because their work was qualitative and depended primarily on interview data, it was important for this step to be offered to ensure higher quality. Ideally, all of the qualitative empirical studies that involved interviews should have done the same.

Methodological Critiques for Research Done About This Topic

Strength: Instrument Choice

One of the methodological strengths of many of these studies was that they used instruments that have been previously utilized, which helps to make them more valid and reliable. Bastos et al. (2014) explained that creating a new instrument requires much more time and resources, so they "recommend developing new instruments only when there are no other options for measuring the phenomenon in question or when the existing ones have huge and confirmed limitations" (p. 920). Thus, choosing an instrument that has been used frequently by other researchers, which is what many of the studies did when selecting the IAT as one of their measures, ensured an effective method. Further, because it has been used in previous studies on implicit bias, it can be assumed that it is a valid method as well. Validity means that not only is the instrument reliable (it will be consistent in its results) but also that it is the appropriate choice of instrument for what the researchers want to discover (Siegle, 2019).

Strength: Pilot Survey

Another strength was the use of a pilot survey in the research done by Marcucci (2020) and Childs (2023). For Marcucci, they piloted the survey with two undergraduate students to see how well it functioned and then piloted with two teachers in vastly different contexts (urban secondary vs. rural elementary). This allowed them to check the quality of the questions, and they also used it to help determine the ultimate sample size they would need for their official survey. Childs asserted that they used the pilot to check for validity and response burden. According to Creswell and Guetterman (2019), a key piece of developing a survey instrument is to administer "the instrument to a small number of individuals and [make] changes based on their feedback" (p. 394). Thus, it is a sign of higher quality that these researchers took that step.

Strength: Adjusting for Other Variables

The third strength would be how the correlational studies explicitly mentioned how they adjusted for covariates that could affect their findings. For instance, Starck et al. (2020) adjusted for covariates of gender, education level, race, and political affiliation; Shirrell et al. (2024) addressed the potential influences of gender, socioeconomic status (SES), emergent bilingual classification, and special education classification in their results. According to Morris et al. (2022), "covariate adjustment is desirable because, if a covariate predicts outcome, accounting for its effect on outcome will improve power to detect a treatment effect unless none of the covariates in a model are prognostic" (p. 1). In other words, adjusting for the other variables that are not the main variables under study can make the correlational findings more accurate.

Weakness: Inappropriate Sample for Generalizability

One the other hand, one of the biggest issues with much of the research that has been performed on this topic is that it has been conducted with preservice teachers. For example, in Lindvall-Östling (2024) none of the participants were active teachers; they either had experience as substitute teachers or practicum experience. As seen from the literature (Stephens et al., 2022),

implicit bias has the potential to change once educators have had more experience working with diverse groups of students. Therefore, drawing conclusions from preservice teacher data may be inaccurate to the experiences of more seasoned educators, causing an issue for generalizability of the results. According to Kukull and Ganguli (2012), "fair samples must provide valid estimates of the population characteristics being studied" (p. 1887). Thus, if we genuinely want to draw conclusions for the population of educators writ large, much more work needs to be done to get a representative sample.

Weakness: Questionable Sample Sizes

Another issue that was common in the studies under examination was a lack of explanation when it came to sample sizes. Some of the studies had small samples when it was inappropriate to do so such as Norris (2022) and Mossing et al. (2025) both of whom used experimental design but only had five participants in one (Norris, 2022) and two teachers and six students in the other (Mossing et al., 2025). For others, the sample size was larger, but the sample was not very representative of the entire population of teachers either in terms of gender (İnan-Kaya & Rubie-Davies, 2022; Norris, 2022; Whitford & Emerson, 2019) or grade-level (Crutchfield et al., 2022; Mekky, 2023; Ura & d'Abreu, 2022). At times, this was because they were using a class as their sample, but to be more rigorous, they could have included multiple classes or even compared classes from different schools or universities. Implicit bias can be influenced by so many factors, so if generalization is a study's goal, the sample must be intentionally determined to support this purpose.

In addition, for many quantitative designs researchers must be able to explain why they selected their sample size due to their confidence level, and none of the studies gave a justification for their numbers. Althubaiti (2022) explained that "review committees often expect an explicit justification of the sample size … [and] when sample size calculation is not mentioned, reviewers may wonder whether the sample size is adequate" (p. 72). Thus, the articles undermine their quality by omitting this information. If the sample size in proportion to the population is not appropriate, each participant has more weight in influencing the results.

Weakness: Opt-In Participants

Another weakness is how individuals were gathered to participate in many of these studies. Many of these studies used snowball sampling (e.g., Norris, 2022; Palmer, 2023) or volunteers who self-selected to collect their data (e.g., Costa et al., 2023; Crutchfield et al., 2022). This, of course, is less credible than a design in which participants were selected at random from a larger population. According to Creswell and Guetterman (2019), random sampling is most rigorous because

> any individual has an equal probability of being selected from the population. The intent of simple random sampling is to choose individuals to be sampled who will be representative of the population. Any bias in the population will be equally distributed among the people chosen. (p. 141)

Many of the studies did not have the ability to do this because they already had a challenge in getting enough participants in the first place, but it is important to acknowledge that that may ethically muddy the waters and not be a strong representation of the larger population of teachers.

This also leads to a concern about the generalizability of the results. Teachers who are willing to be part of a study on implicit bias are likely already more socially aware and willing to change compared to the general population; thus, results may imply that implicit bias is not as severe as it actually is.

Weakness: Self-Reported Data

Lastly, the data on the IAT, which was one of the most utilized measures to assess implicit bias, was often self-reported. So even though the IAT is considered a valid and reliable instrument, researchers had to trust that participants were being honest in reporting how they performed on the test, which obviously could greatly affect the findings.

Suggestions for Future Research

Because of the ethical and methodological weaknesses named in the previous section, future research should seek to replicate these studies and correct for those problems. Additionally, the studies that emerged through this review did not consider several topic areas ripe for further exploration some of which are highlighted below.

Growth Mindset and Reducing Implicit Bias

Building upon the work of Stephens et al. (2022) on preservice teachers in Australia, it would be useful to do more work on the connection between growth mindset—when people "believe that human traits are malleable and can be changed through effort and flexibly adopting useful strategies, whereas when people endorse fixed mindset beliefs, they believe that human traits are rigid and unchanging" (Kroeper et al., 2022, p. 372)—and a decrease in implicit bias. If, as supported by De Houwer (2019), that implicit bias is a behavioral response able to change over time, having a growth mindset compared to a fixed mindset could influence how easily such a change might occur and how teachers could adapt to a change. It would be beneficial to conduct further studies to explore how to develop a growth mindset in the ability to decrease implicit bias for all educators and to learn how such development happens over time.

Interestingly, Kern et al. (2023) discovered that when implicit bias is described to participants, they are more likely to read more about the bias if it is presented in a growth mindset framework compared to a fixed mindset; however, more work must be done to determine whether that impacts outcomes in practice.

Implicit Bias and Anti-Critical Race Theory Policy

It would also be beneficial to do a mixed methods study using ethnographic methods and surveys to compare the levels of implicit bias for teachers who have received training in culturally responsive pedagogy in their teacher education programs or departments compared to teachers who are not allowed to discuss racial issues at work. In the past few years, we have seen an increase in supposed anti-critical race theory (CRT) policies across the country. As Welton et al. (2023) stated, "the unwritten policy and rhetoric against CRT at-large are still having a chilling effect on implementation practices where educators are generally fearful of addressing issues of race let alone broader topics on equity" (p. 3). With this being the case, one could hypothesize that implicit

bias could be more severe in states that are afraid to discuss issues of race in education because biases will go unchallenged and uninterrupted. This is particularly important in Texas, because of all the states to implement new anti-CRT policy, "Texas is the only state where White students have the lowest enrollment among all racial/ethnic groups in the state's PK–12 public school system" (Welton et al., 2023, p. 14). If teachers are going to be working with a majority of students of color, it is all the more vital to be aware of the potential for implicit bias.

Implicit Bias and Artificial Intelligence

Another related area of study could be implicit bias and artificial intelligence (AI). There are high-level claims being made on both sides of AI arguments. Some scholars are arguing that AI has the potential to mitigate some of the implicit bias that humans hold. Lin et al. (2021) argued that although AI may still have its drawbacks, it can be programmed to ignore some of the biases that humans have. For instance,

> AI systems can be programmed to ignore information that is irrelevant to certain decisions (e.g., a job applicant's gender in hiring a computer programmer). This allows AI to analyze only information that is relevant to the job requirements (e.g., programming skills) to reach an unbiased decision. (Lin et al., 2021, p. 66)

Of course, this would require careful programming so that the system is not just perpetuating bias.

However, incorporating AI is not without risk. Instead of reducing bias, it could perpetuate it more insidiously. According to Roshanaei (2024), "programmers, often without realizing it, can transfer their own biases into AI systems. These biases, whether they pertain to race, gender, or other stereotypes, are integrated into the algorithms, leading to AI decisions that could unfairly affect certain groups" (p. 18961). This is why more research and caution must be utilized before fully committing to any change.

This brings up the question of how AI could be incorporated into the classroom to reduce teacher bias. One potential area could be in grading. Texas state tests are moving towards a hybrid of humans and AI when assessing the essays on the State of Texas Assessments of Academic Readiness. Could a move like this increase equity in the classroom because it can adjust for teacher bias or will there be other issues to contend with? Preliminary study suggests that the issue is complex. For instance, Chai et al. (2024) found that students believed AI to be more "fair" when it came to evaluation than they believed their own professors to be. This was the case because they felt the AI was more transparent. Thus, when teachers were better at explaining why they scored students the way that they did, students felt less of a difference in fairness. Building on this perception, Flodén (2024) conducted a study that compared how AI graded students compared to humans and found that there were, on average, no significant differences in the scores. AI tends to avoid as many extreme scores but was normally within a 10% range from the human score. This, however, could be used as an argument either for or against its use. If the score is not significantly different, it does not seem to necessarily have a dramatic difference in its amount of bias compared to a human. However, much more research must be done on this emerging field, including how AI might be used to reveal or blunt the effects of teacher racial bias.

Limitations

Teachers' racial bias is an area that has been growing in scholarship in recent years and for good reason. Many of the studies on this topic were dissertations, which shows that it is a newer focus in the zeitgeist of educational research. However, this means that the topic does still need more peer-reviewed studies to deepen the credibility and generalizability of the research.

Further, while every attempt was made to cover the different types of research methods available, due to the specificity of the topic not every method was found (which implies areas for further research). In addition, while the attempt was made to separate each of these studies into separate methodological categories, not every study clearly identified themselves into one category, and some overlap was present. For instance, many of the studies using a survey design combined quantitative and qualitative methods. Moreover, there are more specific side roads that could have been more explicitly examined (e.g., the impact of the race of the teacher on implicit bias towards students), but the intention of this article was to give a more high-level overview of the types of research design and methods being utilized in the research on this topic.

Conclusion

Overall, teachers' racial implicit bias is a topic that has been emerging in focus in recent years. With many new developments in technology and in policy, it is all the more important that educational researchers continue to explore the effects of implicit bias and work to address them more effectively. We can see that the most effective studies used reliable and valid measures (such as the IAT), adjusted for confounding variables, and utilized member checking and pilot studies to ensure accuracy. One of the greatest challenges in studying implicit bias is navigating participants' social desirability bias that can influence research findings.

By examining what methodologies previous studies have implemented, critiquing what they have done well in their practices and what can be improved, and exploring what avenues future studies could pursue, this article has attempted to serve as a guide for researchers interested in addressing teachers' implicit racial bias in the educational field.

References

Agathocleous, A., Conrad, K., & Moore, R. (2024). *Trump on DEI and anti-discrimination law*. ACLU. https://www.aclu.org/publications/trump-on-dei-and-anti-discrimination-law

Althubaiti, A. (2022). Sample size determination: A practical guide for health researchers. *Journal of general and family medicine*, *24*(2), 72–78. https://doi.org/10.1002/jgf2.600

American Psychological Association. (2018a). Cognitive load. *APA Dictionary of Psychology*. https://dictionary.apa.org/cognitive-load

American Psychological Association. (2018b). Implicit prejudice. *APA Dictionary of Psychology.* https://dictionary.apa.org/implicit-prejudice

Amnesty International. (2017). *'Politics of demonization' breeding division and fear*. Amnesty International. https://www.amnesty.org/en/latest/press-release/2017/02/amnesty-international-annual-report-201617

Ayala, M. C., Webb, A., Maldonado, L., Canales, A., & Cascallar, E. (2024). Teacher's social desirability bias and migrant students: A study on explicit and implicit prejudices with a

list experiment. *Social Science Research*, *119*, Article 102990. https://doi.org/10.1016/j.ssresearch.2024.102990

Bastos, J. L., Duquia, R. P., González-Chica, D. A., Mesa, J. M., & Bonamigo, R. R. (2014). Field work I: Selecting the instrument for data collection. *Anais Brasileiros de Dermatologia*, *89*(6), 918–923. https://doi.org/10.1590/abd1806-4841.20143884

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. http://doi.org/10.1177/1476718X221087051

Chai, F., Ma, J., Wang, Y., Zhu, J., & Han, T. (2024). Grading by AI makes me feel fairer? How different evaluators affect college students' perception of fairness. *Frontiers in Psychology, 15*, Article 1221177. https://doi.org/10.3389/fpsyg.2024.1221177

Childs, T. M. (2023). *A mixed method study of how teachers' racial bias relates to student-teacher relationships* (Publication No. 30246688) [Doctoral dissertation, University of South Carolina]. ProQuest Dissertations and Theses Global.

Chin, M. J., Quinn, D. M., Dhaliwal, T. K., & Lovison, V. S. (2020). Bias in the air: A nationwide exploration of teachers' implicit racial attitudes, aggregate bias, and student outcomes. *Educational Researcher*, *49*(8), 566–578. https://doi.org/10.3102/0013189X20937240

Chiu, C. L., Sayman, D., Lusk, M. E., Kressler, B., & Cote, D. (2022). "Does this mean I am racist, distrust, or dislike people of color?" A DisCrit qualitative study of implicit bias among preservice and practicing special educators. *Issues in Teacher Education*, *31*(1), 6–34. http://eric.ed.gov/ERICWebPortal/detail?accno=EJ1337918

Costa, S., Pirchio, S., Shevchuk, A., & Glock, S. (2023). Does teachers' ethnic bias stress them out? The role of teachers' implicit attitudes toward and expectations of ethnic minority students in teachers' burnout. *International Journal of Intercultural Relations, 93*, Article 101757. https://doi.org/10.1016/j.ijintrel.2023.101757

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.) Pearson.

Crutchfield, J., Sparks, D., Williams, M., & Findley, E. (2022). In my feelings: Exploring implicit skin tone bias among preservice teachers. *College Teaching, 70*(4), 469–481. http://doi.org/10.1080/87567555.2021.1979456

De Houwer, J. (2019). Implicit bias is behavior: A functional-cognitive perspective on implicit bias. *Perspectives on Psychological Science, 14*(5), 835–840. https://doi.org/10.1177/1745691619855638

Flodén, J. (2024). Grading exams using large language models: A comparison between human and AI grading of exams in higher education using ChatGPT. *British Educational Research Journal*, *51*(1), 201–224. https://doi.org/10.1002/berj.4069

Gagliardi, J. A. (2015). *Reducing implicit racial bias in preservice teachers by facilitating impact awareness* (Publication No. 1587343) [Specialist thesis, Western Carolina University]. ProQuest Dissertations and Theses Global.

Hicks, L. (2019). *Informed consent-SBE*. CITI Program. https://www.citiprogram.org/index.cfm?pageID=14&message=64#view

Hines, D. E., Boveda, M., & Lindo, E. J. (2021). *Racism by another name: Black students, overrepresentation, and the carceral state of special education.* Information Age Publishing.

Holden, R. R., & Passey, J. (2009). Social desirability. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 441–454). The Guilford Press.

İnan-Kaya, G., & Rubie-Davies, C. M. (2022). Teacher classroom interactions and behaviours: Indications of bias. *Learning and Instruction, 78,* Article 101516. https://doi.org/10.1016/j.learninstruc.2021.101516

Jacoby-Senghor, D. S., Sinclair, S., & Shelton, J. N. (2016). A lesson in bias: The relationship between implicit racial bias and performance in pedagogical contexts. *Journal of Experimental Social Psychology, 63,* 50–55. https://doi.org/10.1016/j.jesp.2015.10.010

Johnson, L. E., Jr. (2020). *Perception, beliefs, or implicit bias: Investigating the relationship between teacher recommendation and African American males selection in gifted and advanced placement courses* (Publication No. 28764172) [Doctoral dissertation, Youngstown State University]. ProQuest Dissertations and Theses Global.

Kern, M. C., Rattan, A., & Chugh, D. (2023). A growth mindset frame increases opting in to reading information about bias. *Personality and Social Psychology Bulletin*, *51*(3), 409–422. https://doi.org/10.1177/01461672231186853

Kroeper, K. M., Fried, A. C., & Murphy, M. C. (2022). Towards fostering growth mindset classrooms: Identifying teaching behaviors that signal instructors' fixed and growth mindsets beliefs to students. *Social Psychology of Education*, *25*(2–3), 371–398. https://doi.org/10.1007/s11218-022-09689-4

Kukull, W. A., & Ganguli, M. (2012). Generalizability: The trees, the forest, and the low-hanging fruit. *Neurology, 78*(23), 1886–1891. https://doi.org/10.1212/WNL.0b013e318258f812

Kumar, R., Gray, D. L., & Kaplan Toren, N. (2022). Pre-service teachers' desire to control bias: Implications for the endorsement of culturally affirming classroom practices. *Learning and Instruction, 78,* Article 101512. https://doi.org/10.1016/j.learninstruc.2021.101512

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., & Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765–1785. https://doi.org/10.1037/a0036260

Larson, K. E., Chaturvedi, A., Dunn, M., & Chavers, N. (2024). Using a contemplative practice intervention in teacher education courses to neutralize implicit racial bias: A feasibility study. *Teaching Education*, *35*(2), 162–183. https://doi.org/10.1080/10476210.2023.2252353

Lin, Y.-T., Hung, T.-W., & Huang, L. T.-L. (2021). Engineering equity: How AI can help reduce the harm of implicit bias. *Philosophy & Technology, 34*, 65–90. http://doi.org/10.1007/s13347-020-00406-7

Lindvall-Östling, M. (2024). "If you are aware of something, you can do something about it": Investigating the lasting effects of an implicit bias training activity. *Teaching and Teacher Education, 140,* Article 104484. https://doi.org/10.1016/j.tate.2024.104484

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

Mekky, N. (2023). *Teacher perceptions of implicit bias on student achievement outcomes* (Publication No. 30521506) [Doctoral dissertation, Western Illinois University]. ProQuest Dissertations and Theses Global.

Morris, T. P., Walker, A. S., Williamson, E. J., & White, I. R. (2022). Planning a method for covariate adjustment in individually randomised trials: A practical guide. *Trials, 23,* Article 328. https://doi.org/10.1186/s13063-022-06097-z

Mortenson, B. C. (2018). *The role of teacher implicit bias in the racial achievement gap* (Publication No. 10791341) [Master's thesis, Georgetown University]. ProQuest Dissertations and Theses Global.

Mossing, K. W., Collins, T. A., Newman, D. S., King, K. A., & Pollard, J. F. (2025) Race-conscious teacher coaching: Disrupting implicit bias and improving teachers' relationships with Black students. *Journal of Educational and Psychological Consultation*, *35*(1), 114–141. https://doi.org/10.1080/10474412.2024.2359673

Neubauer, B. E., Witkop, C. T., & Varpio, L. (2019). How phenomenology can help us learn from the experiences of others. *Perspectives on Medical Education, 8*(2), 90–97. https://doi.org/10.1007/s40037-019-0509-2

Norris, K. J. (2022). *An examination of the effects of education, perspective-taking, and exposure on the implicit racial bias of teachers* (Publication No. 29394185) [Doctoral dissertation, The Chicago School of Professional Psychology]. ProQuest Dissertations and Theses Global.

Palmer, D. O. (2023). *Teachers' perceptions of implicit bias training on cultural competence: A qualitative descriptive study* (Publication No. 30490520) [Doctoral dissertation, Grand Canyon University]. ProQuest Dissertations and Theses Global.

Pepis, T. A. (2022). Stereotype threat and colorblindness. *Journal for Multicultural Education, 16*(2), 184–194. https://doi.org/10.1108/JME-09-2020-0095

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

Roshanaei, M. (2024). Towards best practices for mitigating artificial intelligence implicit bias in shaping diversity, inclusion and equity in higher education. *Education and Information Technologies*, *29*(14), 18959–18984. https://doi.org/10.1007/s10639-024-12605-2

Shirrell, M., Bristol, T. J., & Britton, T. A. (2024). The effects of student–teacher ethnoracial matching on exclusionary discipline for Asian American, Black, and Latinx students: Evidence from New York city. *Educational Evaluation and Policy Analysis*, *46*(3), 555–580. https://doi.org/10.3102/01623737231175461

Siegle, D. (2019). *Educational research basics*. Neag School of Education. https://researchbasics.education.uconn.edu/instrument_reliability/

Starck, J. G., Riddle, T., Sinclair, S., & Warikoo, N. (2020). Teachers are people too: Examining the racial bias of teachers compared to other American adults. *Educational Researcher*, *49*(4), 273–284. https://doi.org/10.3102/0013189X20912758

Stephens, J. M., Rubie-Davies, C., & Peterson, E. R. (2022). Do preservice teacher education candidates' implicit biases of ethnic differences and mindset toward academic ability change over time? *Learning and Instruction, 78,* Article 101480. https://doi.org/10.1016/j.learninstruc.2021.101480

Tropp, L. R., & Rucinski, C. L. (2022). How implicit racial bias and concern about appearing racist shape K–12 teachers' race talk with students. *Social Psychology of Education*, *25*(4), 697–717. https://doi.org/10.1007/s11218-022-09715-5

Ura, S. K., & d'Abreu, A. (2022). Racial bias, social-emotional competence, and teachers' evaluation of student behavior. *Children & Schools*, *44*(1), 17–26. https://doi.org/10.1093/cs/cdab028

Welton, A., Diem, S., & Lent, S. (2023). Let's face it, the racial politics are always there: A critical race approach to policy implementation in the wake of anti-CRT rhetoric. *Education Policy Analysis Archives, 31*. https://doi.org/10.14507/epaa.31.7694

Whitford, D. K., & Emerson, A. M. (2019). Empathy intervention to reduce implicit bias in pre-service teachers. *Psychological Reports*, *122*(2), 670–688. https://doi.org/10.1177/0033294118767435

Yeung, R. P. (2022). *Implicit bias, discipline, and academic disparities: A phenomenological study to understand White teachers' perceptions of Black students* (Publication No. 29261184) [Doctoral dissertation, Baylor University]. ProQuest Dissertations and Theses Global.

Appendix

Source	Design	Data Sources	Ethical Consideration	Methodological Strengths/Weaknesses
Ayala et al. (2024)	Survey	-Survey data -Census data	A consulting firm selected the participants, but they do not specify IRB or how informed consent was achieved.	Strength: list experiment to mitigate social desirability bias (control group gets noncontroversial statements, experimental group gets both) Strength: large sample (1st time: 429 teachers, 333 did whole survey 2nd time: 309 teachers, 230 completed) Weakness: not great response rate Weakness: overrepresented middle school teachers, so not as generalizable Strength: incorporated control variables into their model
Blackson et al. (2022)	Survey	-IAT -Racial Preference Scale -Temperature Scale -Questionnaire	Recruited in person Written consent, but no direct mention of IRB Web-based survey	Strength: invited staff at three child care centers to participate (57 consented and 48 completed study) Strength: graph of percentages on the IAT clearly shows the implicit bias Weakness: not a lot of information about the specific child care centers – seems to be convenience sampling since these centers were already participating in a larger SEL intervention.
Childs (2023)	Survey	Survey with 7 sections: -Implicit bias -Explicit bias -Teacher-student relationship -Teacher-student racial mismatch -Open-ended questions -Demographics Future research participation	IRB completed Anonymity assured Online implementation Raffle for $50 gift card (6 participants got one)	Strength: purposive sampling strategy to recruit teachers Strength: table listing advantages and disadvantages of the different study measures used Strength: explained how they prevented bots from completing the survey and affecting results Strength: completed a pilot test of the survey to ensure how well it worked (validity and response burden) Strength: three different recruitment methods (closed networks, like email, then social media, then recruited through districts) to achieve higher number of participants
Chin et al. (2020).	Correlational design	-Project Implicit IAT scores -Stanford Education Data Archive -Civil Rights Data Collection	Using large data sets, so I don't see how individuals could be impacted IRB not needed	Strength: multiple measures Strength: controlled for some mediating variables Strength: clear on limitations of the data Weakness: IAT data is self-reported
Chiu et al. (2022)	Quasi-experimental	-IAT	Attained IRB	Strength: strong number of participants (68)

		Participant reflections on results	Advised that participation was voluntary and would not affect grade, but it was presented as a class assignment	Strength: identified themes in the responses via coding Strength: member-checking to ensure validity of results
Costa et al. (2023)	Correlational design	-IAT -Relational responding test -Maslach Burnout Inventory -Modern Racial Prejudice Scale -Demographic Questionnaire	Informed consent embedded into the beginning of the surveys No IRB	Strength: multivariable analysis Strength: Larger enough sample size (104) Weakness: teachers opted-in
Crutchfield et al. (2022)	Case Study	-IAT -Interviews -Journals	Completed member-checking: participants checked transcript and were able to make sure info couldn't be traced to them IRB approved	Weakness: most of their sample were preservice secondary teachers (not representative of all teachers) Weakness: only eight teachers participating Weakness: participants self-selected Strength: multiple data sources (triangulation) Strength: coding themes in the results to write about
Gagliardi (2015)	Experimental	-Questionnaire -IAT -Affect Misattribution Procedure -Responses to a narrative about implicit bias -Responses to Teaching Tolerance classroom activities -Rating scale responses to vignettes	Offered extra credit in class for participating No mention of IRB, but they did give informed consent	Strength: participants gathered from two different psychology classes (so only quasi-experimental) that allowed for hopefully less discussion between the control and experimental group Strength: reasonable sample size (34) Strength: control group did the Affect Misattribution Procedure and vignette responses without the intervention to serve as a comparison.
İnan-Kaya and Rubie-Davies (2022)	Exploratory Qualitative Field study	-Field notes -Verbal behaviors -Nonverbal behaviors	Notified parents of kids in the classes even though the teachers are really the ones being observed New Zealand version of IRB done	Strength: mix of schools in low and high SES areas Weakness: most of the teachers participating are women Strength: observed for a whole class period four different times Strength: used saturation to determine that the four observation periods were enough Maybe weakness: only one observer
Jacoby-Senghor et al. (2016)	Experimental	-Subliminal priming task to	Participants received course credit (not ideal,	Strength: tested for multiple things at once finding that when an instructor has more implicit bias, they are more

		measure implicit bias -"Attitudes Towards Blacks" scale for explicit bias -Video recording of lessons -Analysis of video using Likert scale -Learner performance assessment questionnaire	because could lead to pressure to participate) No clear IRB	nervous and deliver a worse lesson AND worse testing outcomes for Black students Strength: added layer to make sure the results were not coming from alt. hypothesis, that Black learners worries that they are the target of prejudice affect their test results, by having a group of non-Black students watch a video of the lesson and see how they were impacted. Weakness: 165 participants with no explanation as to why they chose that number
Johnson (2020)	Quasi-experimental	-Enrollment data -District demographics -AP and gifted enrollment data -Attendance data -Achievement data -Discipline data -Zip code information -Interview information	Utilized only preexisting data Attained IRB	Strength: justification for study is strong (Black boys are only 3.5% of students in gifted classes and Black students make up only 9% of AP classes) Weakness: the methodology is not super clear. The author stated that this is quasi-experimental, but it seems more to be examining the different factors causing Black students to not enter AP/gifted programs. The author is not putting participants through conditions.
Kumar et al. (2022)	Correlational	-IAT -5 item scale about explicit bias -4 item scale about need to appear unbiased and need to actually be unbiased -5 item scale about promoting respect -7 item scale on cultural responsibility	Invited to participate online $25 gift card compensation Assured of anonymity No direct IRB mention	Strength: sample size of 258 preservice teachers Strength: clear exploration of all parts of their hypotheses and variables (for instance, that the more explicit bias someone has, the less they believed in cultural responsibility and respect)
Lai et al. (2014)	Experimental	-IAT -Multi-Category IAT -Self-reported racial attitudes	Contest between different researchers to see which intervention would be the most effective at reducing implicit bias No mention of IRB or how participants were compensated	Strength: calculated sample size for a 99% power Strength: large sample (17,021 total participants) Weakness: participants were selected from people who visited the Project Implicit website Strength: random assignment to control or one of the experimental categories

				Strength: random assignment of a pretest or not since pretest could impact how they perform later
Larson et al. (2024)	Experimental Design	-IAT -Reflective journaling	IRB approved Presented to classes what the study would be and got volunteers Anonymous No effect on grade	Weakness: only 22 participants Strength: pre and posttest Strength: intervention was done multiple times: breathing, loving kindness meditation, gratitude writing Weakness: not all participants attended every session Strength: thematic analysis of the qualitative journaling
Lindvall-Östling (2024)	Ethnography: Case study	Semistructured interview (same themes but not word-for-word)	Gave language preference options for interview No direct mention of consent-process	Strength: all participants had engaged in the same pedagogic activity, but the time since they had done it differed Weakness: none of the participants were active teachers, had either substitute experience or practicum experience
Marcucci (2020)	Survey	-Feelings thermometer to measure explicit bias without revealing purpose of survey -Responses to vignettes using Likert scales	Acknowledge the effect of social desirability on results No IRB mention	Strength: piloted the survey with two undergraduate students to see how well it functioned and then two teachers in very different contexts (urban secondary vs rural elementary) Strength: conducted pilot survey to calculate the needed sample size for the official survey Strength: random assignment to group with vignettes about Black/White students Strength: consulted teachers via in-depth interview to determine the possible disciplinary choices to offer in the options on survey Strength: examined data of whole sample and then subsamples of just White teachers and just high school teachers
Mekky (2023)	Correlational Design	-IAT -Demographic data from teachers	IRB approved Ensured privacy Informed consent Incentive option	Weakness: limited to K–5 teachers Strength: clear rationale for research questions (whether there is a difference in implicit bias levels between low and high performing Title-I schools) Weakness: convenience sampling (what teachers responded to the email) and low response rate (sent 2,302 emails and only got 112 completed surveys) Strength: explained the validity of the measure (IAT)

Mortenson (2018)	Correlational Design	-Education Longitudinal Study of 2002 -12th grade standardized math score	Using preexisting data, so no consent forms or IRB needed	Strength: identified a large number of variables that could have an impact that would need to be controlled Strength: large sample size (from preexisting data) Strength: acknowledged limitations Weakness: response rates were not great for the longitudinal study over time
Mossing et al. (2025)	Experimental design	-Teacher self-reporting -Observational data done by graduate students -Student-Teacher relationship scale -Student Perception of Affective Relationship with Teacher Scale -Teacher Coaching Evaluation Scale	IRB was completed. Teachers were compensated with gift cards; what about students? Teachers self-selected (could change results)	Weakness: very small sample (two teachers and six students) Strength: A-B-A-B design, so each classroom served as its own control Weakness: part of it involves teacher self-reporting (their fidelity to the intervention) Strength: did interobserver agreement data to ensure alignment across different people collecting data Weakness: 20-minute observation periods – is that enough? Strength: used multiple measures that have been utilized before (reliability)
Norris (2022)	Experimental design	-IRAP test each day (measuring implicit racial bias) -Acceptance and Action Questionnaire	Recruited using facebook and flyers IRB mentioned	Weakness: only five participants Weakness: all of the participants are women Strength: tried multiple interventions, with multiple posttests after each Strength: examined intervention package as a whole and the effect of each individual intervention
Palmer (2023)	Qualitative Descriptive Study	-Semi-structured interviews -Focus group -Demographic info	Informed consent at beginning Posted the invitation to participate on several teacher Facebook groups Snowball sampling IRB approved	Strength: included elementary and secondary teachers Weakness: out of the population of the Facebook group (75,534) they only selected 20 participants (though they gave a justification for that)
Pepis (2022)	Experimental	-IAT data (3x) -Demographic questions -Semi-structured interview (~1 hour)	All students in a course did the intervention, but could decline to have their data as a part of the study (still sounds like they could be pressured)	Strength: solid sample size (63 for the IAT and then six interview subjects) Strength: random assignment to Condition 1 where they received their IAT scores before an intervention that was intended to reduce bias, or Condition 2, did not receive scores Weakness: all the people who agreed to be interviewed had randomly been in Group 2.

			No IRB mention	
Quinn (2020)	Experimental	-Web-based survey experiment (grade a writing sample with differences on whether the author was Black or white) -IAT (for implicit bias) -Feelings thermometer (for explicit bias)	Not specified on how the consent process worked or if there was any compensation	Strength: large sample size that is justified using stats (.80 power) Weakness: many participants who began the survey stopped when asked to complete the IAT part. Strength: random assignment to one of the two conditions Strength: consider the ecological validity of giving a sample the way an actual teacher would grade, but also acknowledges the limitations, then, of testing for reliability Strength: selected out the bias for K–2 teachers from the total sample since they would be more familiar with grade level expectations for this age
Shirrell et al. (2024)	Correlational	-Deidentified student outcome files -Deidentified administrative personnel files	The data were preexisting and had no clear identifiers IRB not needed	Strength: HUGE sample (3,524,408 student-year observations, with roughly 70,000 unique teachers and 350,000 unique students each year) Weakness: doesn't factor length of time with teacher–elementary kids with the same teacher all day vs. older students Strength: identified controls of gender, SES, emergent bilingual classification and special education classification Strength: gave estimates of linear probability and used logistic regression for a robustness check
Starck et al. (2020)	Correlational	-IAT data (Teacher category vs All other categories) -ANES 2008 Times Study Series: Affective Misattribution Procedure	Using preexisting data from the Project Implicit database ANES study randomly selected addresses and got consent, and compensated participants	Weakness: The explicit bias part is self-reported (the test asked you the level of warmth you feel towards Black people vs. White people) Strength: adjusted for covariates of gender, education level, race, and political affiliation Weakness: how representative would the population of people who have taken the IAT be to generalize to the entire population? Strength: compensated for that issue by their second study, using more nationally representative data (which showed teachers have just as much implicit bias as other professions, whereas the other showed they had less) Strength: very large sample sizes
Stephens et al. (2022)	Correlational	-IAT data (but adapted to be with Maori surnames vs European)	Data collected in computer lab of 1st and 3rd year students	Strength: performed both a cross-sectional and longitudinal study to see difference between earlier years of program and later

		-Intelligence mindset measures on a 6-point scale -How much time they spend with Maori people	At end of the survey, they can select whether they consent for the info to be used or destroyed First year students also asked to consent for a follow-up when they get to their third year No mention of IRB	Strength: large sample size (first-year n = 239 and third-year n = 74 teacher education students) Strength: reports about all hypotheses (third-year students have lower implicit bias but that there was not a statistically significant difference in intelligence mindset between the groups)
Tropp and Rucinski (2022)	Survey Design	-National online survey with Teaching Tolerance and Perception Institute re: concerns of appearing racist, intention to engage in race talk, confidence in engaging in race talk, social desirability, support from school, motives to control prejudice) -IAT -Second survey at an urban school district	Online consent form $20 gift card to teachers in second survey IRB not specified Teachers recruited via email	Strength: random selection from their larger sample (all teachers that used the Teaching Tolerance website) Weakness: teachers that use the Teaching Tolerance website are already more likely to have more of a social justice mindset Strength: large sample (1,187) and (1,013) Strength: teachers in the first survey represented all geographic regions of the country (W, S, NE, Midwest) with a fairly even split
Ura and d'Abreu (2022)	Experimental design	-Survey data -Vignette responses (using Likert scale) -Demographic Information -Social-Emotional Competency questionnaire	IRB obtained Even though their hypothesis was not supported, they present the data honestly	Weakness: focus only on K–6 teachers in the south Strength: random assignment between groups ("White names" in vignettes vs. "Black names") Weakness: only 18% response rate (email invitation to 920 teachers and only 164 participated) Strength: controlled for gender differences in vignette situations presented
Whitford and Emerson (2019)	Experimental Design	-IAT (pre and posttest) -Demographic info	IRB was completed Participants were given extra credit in an education	Weakness: the sample was all White women, so the results are not as generalizable Strength: random assignment between control group and experimental group

			class for participating	Strength: considered variables of age, whether they have ever lived with someone of another race, where they were raised, and what they plan to do Weakness: The students only read about scenarios and were asked to empathize. Not sure if that proves that they would think or act differently in an actual interaction with a person of color.
Yeung (2022)	Phenomenological study (ethnography?)	15 open-ended questions on a questionnaire	Superintendent approved study and so did the Equity Development Opportunity Department Got informed consent from participants IRB reviewed	Strength: criterion sampling to choose participants Weakness: only 11 participants (they cite a researcher who says phenomenological studies should have between 10 to 30) Strength: chose questionnaire instead of face-to-face because of sensitive questions (people more likely to be honest) Strength: provides example of the questionnaire in the paper