EdWorkingPaper No. 20-205

Bias in the Air: A Nationwide Exploration of Teachers' Implicit Racial Attitudes, Aggregate Bias, and Student Outcomes

Mark J. Chin
Harvard University

David M. Quinn
University of Southern California

Tasminda K. Dhaliwal
University of Southern California

Virginia S. Lovison
Harvard University

Theory suggests that teachers' implicit racial attitudes affect their students, but we lack large-scale evidence on US teachers' implicit biases and their correlates. Using nationwide data from Project Implicit, we find that teachers' implicit White/Black biases (as measured by the implicit association test) vary by teacher gender and race. Teachers' adjusted bias levels are lower in counties with larger shares of Black students. In the aggregate, counties in which teachers hold higher levels of implicit and explicit racial bias have larger adjusted White/Black test score inequalities and White/Black suspension disparities.

VERSION: February 2020

Suggested citation: Chin, Mark J., David M. Quinn, Tasminda K. Dhaliwal, and Virginia S. Lovison. (2020). Bias in the Air: A Nationwide Exploration of Teachers' Implicit Racial Attitudes, Aggregate Bias, and Student Outcomes. (EdWorkingPaper: 20-205). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/n3fe-vm37

Bias in the Air: A Nationwide Exploration of Teachers' Implicit Racial Attitudes, Aggregate Bias, and Student Outcomes

Mark J. Chin[1]
David M. Quinn[2]
Tasminda K. Dhaliwal[2]
Virginia S. Lovison[1]

The research reported in this article was made possible by a grant from the Spencer Foundation (# 201800136). The views expressed are those of the authors and do not necessarily reflect the views of the Spencer Foundation.

[1] Harvard Graduate School of Education
[2] Rossier School of Education, University of Southern California

Abstract

Theory suggests that teachers' implicit racial attitudes affect their students, but we lack large-scale evidence on US teachers' implicit biases and their correlates. Using nationwide data from Project Implicit, we find that teachers' implicit White/Black biases (as measured by the implicit association test) vary by teacher gender and race. Teachers' adjusted bias levels are lower in counties with larger shares of Black students. In the aggregate, counties in which teachers hold higher levels of implicit and explicit racial bias have larger adjusted White/Black test score inequalities and White/Black suspension disparities.

Key words: *Implicit racial bias, teacher bias, school discipline disparities, achievement inequality*

Bias in the Air: A Nationwide Exploration of Teachers' Implicit Racial Attitudes, Aggregate Bias, and Student Outcomes

A vast literature in education shows that teachers treat students differently based on student race, and that such differential treatment can affect students' learning (Ferguson, 2003; Tenenbaum & Ruck, 2007). In a separate literature, social psychologists demonstrate that people hold "implicit racial biases," or biases that lie outside conscious awareness. Measures of implicit bias predict various biased behaviors (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), especially at the aggregate level (Payne, Vuletich, & Lundberg, 2017). Education researchers have thus begun measuring teachers' racial biases to better understand how they affect students, but these studies are few in number, small-scale, and mostly situated outside the US (Warikoo, Sinclair, Fei, & Jacoby-Senghor, 2016). As such, we lack the basic descriptive facts about teachers' implicit racial biases and their correlates that will help advance theory of implicit racial bias in education. In the present study, we use data from three large-scale nation-wide data sources to help fill this gap.

Background

Implicit bias is mediated by a process of implicit cognition (Greenwald & Krieger, 2006). Cognition is "implicit" when it takes place outside of one's conscious attentional focus (Greenwald & Krieger, 2006). Two forms of implicit cognition relevant to race include implicit attitudes (the tendency to like or dislike members of a racial group) and implicit stereotypes (the association of a group with a particular trait) (Greenwald & Krieger, 2006). Implicit attitudes and stereotypes can be automatically activated in one's mind (Devine, 1989), leading to implicit bias, or prejudicial behaviors or judgments (Greenwald & Krieger, 2006). Thus, people can exhibit implicit bias even when they do not consciously endorse the underlying attitude or stereotype (Devine, 1989; Dovidio, Kawakami, & Gaertner, 2002).

Because implicit attitudes elude conscious awareness, they require special methods of measurement. The most widely used measure of implicit racial bias is the implicit association test (IAT). The Black-White IAT assesses the relative strength of one's implicit associations between European Americans[1] and an attitude or stereotype, relative to the strength of one's associations for African Americans, through response times on a series of computerized categorization tasks (Greenwald et al., 2009). Numerous studies show IAT performance predicts racially-biased behaviors in individual-level and geographically aggregated data (Greenwald et al, 2009; Green et al., 2007; Hehman, Flake, & Calanchini, 2018; Leitner, Hehman, Ayduk, & Mendoza-Denton, 2016; McConnell & Leibold, 2001; but see Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013 for a different take on the evidence).

Implicit Racial Bias and Educators

Educators' implicit racial biases are of particular interest due to their potential consequences for students (Quinn, 2017; Warikoo et al., 2016). Findings from non-educational settings (Dovidio et al., 2002) lead us to expect teachers' negative implicit attitudes toward different racial groups to influence their demeanor and warmth when interacting with students and families from those groups. These cues are often detectable (Dovidio et al., 2002) and can communicate a lack of interest or confidence in students, in turn inhibiting the development of relationships conducive to learning (Babad, 1993).

Teachers with implicit biases are liable to provide biased evaluations of students' academic performance or potential, which can negatively impact Black students through self-fulfilling prophesies (Papageorge, Gershenson, & Kang, 2016) or by triggering stereotype threat

(Steele & Aronson, 1995). Students are generally good at perceiving teachers' expectations (McKown, Gregory, & Weinstein, 2010), and students as young as six can recognize when people hold stereotypes (McKown & Weinstein, 2003). This may not only impede performance in the short-term, but can also diminish learning in the long-term, either through stress (Taylor & Walton, 2011) or by inducing challenge avoidance, dis-identification with school, and rejection of teacher feedback (Perry, Steele, & Hillard, 2003; Steele & Aronson, 1995).

Educators' implicit biases may also contribute to the well-documented racial disparities in school discipline outcomes (Gregory, Skiba, & Noguera, 2010) by affecting the way in which educators interpret students' behaviors or the severity of the punishments they deliver. Evidence suggests that Black students are often disciplined for more subjective infractions, such as "disrespectful behavior" or acting "disruptively," while White students are often disciplined for more objective infractions such as smoking or vandalism (Skiba, Michael, Nardo, & Peterson, 2002). Educators with stronger implicit biases may be more likely to interpret Black students' behaviors as threatening and hence dispense discipline (Ferguson, 2000), which can negatively affect student learning and other life outcomes (Gregory et al., 2010; Lacoe & Steinberg, 2019).

Measuring implicit bias in education. Despite theoretical support for its influence in education, few researchers have directly measured teachers' implicit racial biases in the US. Studies from outside the US show that teachers' levels of implicit bias (as measured by the IAT) toward racial/ethnic minorities predict test score inequalities within teachers' classrooms (Peterson, Rubie-Davies, Osborne, & Sibley, 2016; van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010), and similar results have been found for gender bias (Carlana, 2019). In the US, teachers with higher levels of racial bias on the IAT were less likely to report that they promoted mutual respect among students in their classrooms (Kumar, Karabenick, & Burgoon, 2015). In an experimental study, Black - but not White - college students learned less when taught by a White college student with higher levels of implicit racial bias (as measured by a subliminal priming task), and this effect seemed to be mediated by instructor anxiety and instructional quality (Jacoby-Senghor, Sinclair, & Shelton, 2015).

Aggregate Implicit Bias

Several studies, mostly occurring in non-educational contexts, show implicit bias scores from the IAT to more strongly predict racial disparities when aggregated to the level of nation, US state, or county/metropolitan area. For example, researchers in the US have found that aggregated implicit (and explicit) bias scores predict county-level rates of cardiovascular disease among Black residents, greater Black-White disparities in infant health outcomes, and disproportionate use of lethal force by police (Blair & Brondolo, 2017). Aggregate implicit bias also explains some of the geographic variation in racial differences in economic mobility (Chetty, Hendren, Jones, & Porter, 2018). In the field of education, Nosek and colleagues (2009) showed that country-level implicit stereotypes dissociating women with science predicted country-level gender disparities on international math and science assessments (Nosek et al., 2009). In the most relevant study to our work, Riddle and Sinclair (2019) find that county-level estimates of White respondents' biases predict disciplinary disparities between Black and White students.

To interpret findings on aggregate bias, social psychologists have proposed the "bias of crowds" (Payne et al., 2017). In this perspective, implicit bias is not a stable trait of individuals. Instead, implicit bias is conceived of as "a social phenomenon that passes through the minds of individuals" which "exists with greater stability in the situations they inhabit" (Payne et al., 2017, p.5). The extent to which an individual exhibits bias will vary across contexts due to differential concept accessibility across those contexts (Payne et al., 2017). Concept accessibility is “the likelihood that a thought, evaluation, stereotype, trait, or other piece of information will be retrieved for use” in cognitive processing (Payne et al., 2017, p. 235). For racial bias in particular, this refers to the ease of accessing negative evaluations or associations when a racial category is activated in one’s mind. According to this theory, some portion of an individual’s IAT score reflects concept accessibility in the broader culture, some reflects influences encountered shortly before the test, and some reflects intermediate influence, or shared concepts that may be made more accessible in some contexts than others. When individuals’ bias scores are aggregated, the idiosyncratic influences wash away and variation in average scores will reflect the contextual influences with the most widely shared accessibility (Payne et al., 2017). Measures of implicit bias are therefore better measures of situations than of individuals and will consequently be more predictive in aggregate.

In our study, we build on limited previous work on aggregate implicit bias in education in two primary ways. First, we consider racial test score differences as outcomes. Despite growing evidence connecting disciplinary and achievement gaps (Pearman, Curran, Fisher, & Gardella, 2019), limited work has investigated the influence of racial bias on the latter outcome (Pearman [2020], working independently has considered similar test score models to ours in a recent working paper). Furthermore, unlike prior work, we disaggregate regional estimates of bias to specifically explore the biases of teachers. We identify the predictors of teachers’ biases and also their relationship to key disparities.

Summary and Research Questions

Theory from social psychology suggests that teachers’ implicit racial biases contribute to racial disparities in academic and school disciplinary outcomes. Initial studies demonstrate the potential value of greater incorporation of theory and measures of implicit biases into education research. Yet we lack a basic descriptive picture of teachers' implicit biases and their correlates. In this study, we therefore address the following research questions:

RQ1) How do teachers' implicit racial biases vary across the US? Do individual characteristics predict teacher implicit bias? Do contextual variables (such as racial composition and average SES) or instructional variables (such as racial differences in student/teacher ratios) predict teachers' implicit biases?

RQ2) Does county-level implicit and explicit Black/White bias (pooling teachers and non-teachers) predict racial disparities in test scores or disciplinary outcomes? Does teacher county-level bias predict disparities?

Methods

Data

We draw from several data sources to answer our research questions. A key data source is Project Implicit, an archive of internet volunteers who visited the Project Implicit website (Xu et al., 2014). The data include visitors' scores on the Black/White IAT and responses to survey items including explicit racial attitudes, demographics, and occupation[2]. The data file contains FIPS county identifiers, enabling us to merge individual- and county-level bias data with data from the Stanford Education Data Archive (SEDA; Reardon et al., 2019a) and the Civil Rights Data Collection (CRDC).

Project Implicit.

The Black/White IAT. The Black/White IAT provides *d*-scores indicating how much more strongly the respondent associates "African American" with a negative valence and "European American" with a positive valence, versus associating "African American" with a positive valence and “European American” with a negative valence. Positive scores indicate an implicit preference for European Americans, negative scores indicate the reverse, and a score of zero indicates neutrality. Cut scores of +/- .15, .35, and .65 are used to distinguish between “little or no,” “slight,” “moderate,” and “strong” biases (Project Implicit, n.d.). We use only IAT data from (self-reported) first-time test-takers so as to avoid including multiple measurements from the same individual, and to improve comparability of scores across respondents. We also include only respondents who visited the Project Implicit website during the academic years overlapping with our student outcome data (i.e., July 2008-June 2016).

Explicit bias. The Project Implicit website administers feeling thermometer items (11-point scale of how cold, neutral, or warm respondents feel towards particular racial groups). For each respondent, we created an explicit bias score by subtracting the respondent’s rating of Black people from their rating of White people.

Stanford Education Data Archive (SEDA). The SEDA test score dataset (v 3.0) contains average student standardized test scores for school districts across the US over the 2008-09 academic year through the 2015-16 academic year (Fahle, Shear, Kalogrides, Reardon, Chavez, & Ho, 2019). These data were assembled using the ED*Facts* data system, which contains math and ELA scores for 3rd through 8th graders, disaggregated by student race/ethnicity. For this study, we used estimates of the standardized mean difference in test scores between White and Black students, aggregated across grades, subjects, and school years to the county-level.

We merge test score data to measures from the SEDA covariate dataset (v 2.1) in order to include county-level controls in analyses. To maintain consistency with models used by Reardon, Kalogrides, and Shores (2019b) in their study explaining geographic variation in racial test score differences, we use an earlier version (v 2.1) of the SEDA covariate file. This version contains a wider range of covariates but, unlike the SEDA test score dataset, does not incorporate district data from the 2015-16 school year. For detail on covariates we use, see Appendix A; for detail on how variables were compiled and for which counties, see Fahle et al. (2019).

Civil Rights Data Collection (CRDC). We merge the Project Implicit data with data from the US Department of Education's Civil Rights Data Collection (CRDC) using county identifiers. The CRDC collects school-level data from all school districts in the U.S. The data contain school-level enrollment counts by race/ethnicity, along with counts by race/ethnicity of students who received at least one in-school or out-of-school suspension over the 2011-12, 2013-14, and 2015-16 school years. We aggregate these counts to the county level over the three school years, then merge the county-level suspension data with (a) county-level bias data from Project Implicit (described below) and (b) the aforementioned county-level covariates from SEDA.

Samples

For ease of comparison, we use the same sample to answer each research question. Specifically, when exploring the predictors of teachers' biases and the relationship between biases and student outcomes, we restrict our analyses to counties that meet the following criteria: have Project Implicit teacher respondents with demographic data and implicit bias scores; have county-level bias estimates; have SEDA test score gap data; have CRDC disciplinary gap data; and have all key county-level covariate data. After these restrictions, we preserve 73% of the 2282 counties with at least one K-12 teacher IAT respondent and 79% of the 2109 counties with both achievement and disciplinary gap data.[3] Furthermore, Tables C1, C2, and C3 in Appendix C show that our results for teacher bias are robust to alternative sample restrictions. In Table C4 of

Appendix C, we use American Community Survey data to show that though our sample counties are more populated, key demographic and economic indicators are similar to counties omitted; however, because of our sample restrictions, we caution against generalizing findings to the approximately 3000 counties in the US more broadly.

In Table 1, we present descriptive statistics for K-12 educators in our common sample (along with comparisons to national estimates when available). Sample teachers are slightly less likely to be female (71% vs 77%), more likely to be Black (9% vs. 7%), and more likely to hold a master's degree (59% vs. 57%) compared to national estimates.

<Table 1>

Analytic Plan

RQ1: Predicting teachers' implicit biases.

To address RQ1, we use responses from K-12 educators in the Project Implicit data to fit multilevel models of the form:

$$Y_{ics} = \alpha_{cs} + \alpha_s + X' + C' + \gamma + \epsilon_{ics} \ (1)$$

$$\alpha_{cs} \sim N(\mu_{cs}, \sigma_{cs}) \ \perp \ \alpha_s \sim N(\mu_s, \sigma_s) \ \perp \ \epsilon_{ics} \sim N(0, \sigma_\epsilon)$$

where Y_{ics} is the IAT score for teacher *i* in county *c* in state *s* (including Washington D.C.), α_{cs} and α_s are random intercepts for county and state respectively, X' is a vector of respondent-level predictor variables (including mutually-exclusive dummy variables for race/ethnicity, gender, age category, and education level), C' is the vector of contextual and instructional variables from the SEDA data similar to those used in Reardon et al. (2019b) described in Appendix A, and γ is a set of school-year fixed effects. To understand how educators' implicit biases vary across the US, we fit model 1 without X' and C' and report the county- and state-level intra-class correlations (ICCs). In Appendix F we include analyses comparing biases of educators and non-educators.

RQ2: Aggregate implicit (and explicit) Black/White biases predicting racial disparities in test scores and suspensions.

Test scores. To investigate the relationship between implicit racial bias and student achievement, we first obtain county-level empirical Bayes (EB) bias predictions adjusted based on: the (non)representativeness of the IAT respondent sample as compared to the actual population of counties; and the differences in reliabilities of predictions across contexts due to variation in the sample size of respondents. We specifically use a multilevel regression and post-stratification (MrP) approach (Hoover & Dehghani, 2019; for more detail see Appendix D) to perform this adjustment. In our MrP model, we use the county-level joint distributions for age, gender, and race from the American Community Survey (2015 5-year estimates) to adjust our pooled bias scores.

We are unaware of any single source that provides nationwide county-level data on teacher demographics, complicating the post-stratification of county-level estimates of teacher bias. We thus searched for this data online for each state, to varying degrees of success. With few states reporting joint distributions, we focused on identifying county-level breakdowns of teacher race (i.e., White, Black, or other race), as individuals' race significantly predicted their biases in our analyses. With the available data, we employ MrP and adjust the county-level teacher bias scores used in analyses. In Appendix Table D, we document the 20 states (including Washington, D.C.) for which we found this data.

To make coefficients more interpretable, we rescale adjusted EBs for bias as *z*-scores at the county level. We then include either pooled or teacher county-level EBs, $\hat{\delta}_j$, as predictors in the following model:

$$\widehat{Y}_j = \alpha + \beta\hat{\delta}_j + C_j' + \gamma + \varepsilon_j + \chi_j, \text{ (2)}$$

$$\varepsilon_j \sim N\left(0, \sigma_{\varepsilon_j}\right),$$

$$\chi_j \sim N\left(0, \widehat{\phi_j^2}\right)$$

In Equation 2, $\widehat{Y}_j$ represents the estimated standardized mean White-Black test score difference (across subjects and years) in county j (using the cohort standardized scale in SEDA). We fit this model using meta-analytic techniques to account for known variation in the precision of these estimated racial test score differences across counties; χ_j reflects the sampling error in $\widehat{Y}_j$ with known variance $\widehat{\phi_j^2}$. We include county covariates, C_j', similar to those used by Reardon et al. (2019b) to explain regional variation in White-Black test score disparities; γ represents a vector of state fixed effects. β thus captures our coefficient of interest—the relationship between county-level bias and test score disparities. Finally, we fit models replacing implicit bias EBs with explicit bias EBs.

Suspensions. Our preferred models for examining the relationship between geographic-area Black/White biases and Black/White school discipline disparities are logistic regression models of the form:

$$P(Y_{ic} = 1|\boldsymbol{X}_c) = \frac{1}{1+\exp(-(\beta_1 Black_i + \beta_2\widehat{\delta_j} + \beta_3(Black_i \times \widehat{\delta_j}) + X_c + \gamma))} \text{ (3)}$$

The outcome, Y_{ic} is an indicator for whether student *i* in county c was suspended one or more times in a given school year, with separate models for in-school and out-of-school-suspensions. $Black_i$ is an indicator for whether student *i* is Black (versus White; we exclude other racial groups), $\hat{\delta}_j$ again represents adjusted county-level EBs (rescaled as *z*-scores, for either pooled or teacher bias scores), and γ represents state fixed effects. We fit models with and without the SEDA county-level covariates, X_c. Note that the CRDC data are not student-level data; rather, we mimic student-level models by pooling suspension data within county across school years and summing the frequency counts, then applying these counts as frequency weights to the aggregated data (see Appendix B for detail).

The predictor of interest, β_3, expresses whether county-level Black-White bias—pooled or for teachers, specifically—is more predictive of suspension probability for Black students than for White students. We hypothesize these coefficients to be positive and statistically significant. Again, we fit additional models that replace implicit bias EBs with explicit bias EBs. In order to account for correlated errors across individuals within geographic regions, we cluster standard errors at the county level (see Appendix E for qualitatively similar results when clustering standard errors at the state level).

Results

Educators' Implicit Racial Biases

Geographic variation. In column 1 of Table 2, we present the results from the unconditional multilevel model predicting K-12 educators' IAT scores (conditional only on year fixed effects). On average, K-12 educators hold "slight" anti-Black implicit bias (*d*-score=.35 in the baseline year, see intercept). Most of the variation in these biases lies within-county, with approximately 2% lying between counties and .6% lying between states (see ICCs in bottom rows).

<Table 2>

Individual and contextual predictors. In column 2 of Table 2, we add dummy variables for teacher gender (female vs. not female), race/ethnicity, age range, and education level. Controlling for everything else, female teachers showed slightly lower levels of bias than non-females (-.023). In many cases, teachers of color showed lower average bias than White teachers (whose mean *d*=.38 [not shown]), with Black teachers showing the lowest levels (average *d*-score of approximately -.04 [not shown]). As a set, the teacher-level predictors reduced the county-level ICC to approximately 1 percentage point. Contextual variables (column 3) reduced county-level variation by a similar amount, with lower levels of teacher bias found in higher SES counties and counties with larger shares of Black students (controlling for other contextual factors). The instructional variables (i.e., expenditures and student-teacher ratio) did not generally predict teacher bias. As seen in column 4, coefficients for teacher-level variables were largely unaffected by the inclusion of the full set of contextual controls, while the magnitudes of the contextual predictors were generally reduced when controlling for teacher demographics.

Racial Bias and Student Achievement

In Table 3, we present results from models using county-level implicit bias (Panel A) and explicit bias (Panel B) to predict county-level test score inequality. As seen in column 1, we find significant negative unadjusted associations between test score inequality and pooled implicit or explicit bias scores (pooled across all Project Implicit site visitors). However, the adjusted associations are statistically significant and positive (b=.033 and b=.025, for implicit and explicit bias, respectively [column 2]). That is, controlling for contextual variables, White students score higher than Black students in counties with higher levels of pro-White/anti-Black implicit and explicit bias.

In columns 3 through 6, we present results on the set of counties for which we can adjust teacher bias scores for sample representativeness. First, we replicate the analyses from columns 1 and 2 (columns 3 and 4), again finding that pooled bias scores predict smaller test score differences when omitting contextual controls but predict slightly larger test score differences with their inclusion (though relationships are attenuated relative to those found in the full sample). For teacher biases in particular, we find similar patterns: significant negative unadjusted associations between White-Black test score inequalities and teachers' county-level implicit (b=-.057) and explicit (b=-.057) biases, but significant positive associations once we enter the set of control variables (column 6). Specifically, controlling for everything else in the model, a one SD-unit difference in county-level implicit bias of teachers predicts approximately a .08 SD unit difference in White-Black test score disparity (.07 SD adjusted association for explicit bias). For reference, this represents approximately 15% of the average disparity (.55 SDs) in our sample counties (see Table 1).

<Table 3>

Racial Bias and Discipline Outcomes

In our sample, Black students are more than twice as likely to receive one or more suspensions (in-school and out-of-school) than White students in the average county; for in-school suspensions, the rates are 14% and 6%, respectively, and for out-of-school suspensions, the rates are 13% and 5% (see Table 1). In Table 4, we present the more formal results from our logistic regression models predicting these disparities using county-level bias. In every model, the coefficient on being a Black student is always positive: regardless of controls, Black students are suspended at higher rates than white students. With regards to bias, without (column 1) and with (column 2) our key county-level covariates, we find patterns consistent with our hypotheses: Higher levels of pooled aggregate implicit and explicit bias again predict in- and out-of-school suspensions differentially for White students and Black students. Black/White disciplinary gaps are larger among counties with higher levels of bias; these relationships appear to be primarily driven by greater probabilities of suspensions for Black students in counties with stronger bias, and not necessarily by lower probabilities of suspensions for White students. When replicating models from columns 1 and 2 for the subset of counties for which we can adjust teacher bias scores (columns 3 and 4), we arrive at largely similar conclusions. Finally, our hypotheses are also supported when focusing on just teachers' biases: counties where teachers have a stronger preference for Whites have greater Black/White disciplinary gaps (columns 5 and 6), even after reweighting scores for representativeness and including covariates.

To help put these numbers into context, see Figure 1, where we plot predicted probabilities for suspension by race against bias (assuming mean values for all other covariates) using the coefficients from the models represented in Panels A and B, column 6. From the figure, we see that Black students in counties with average teacher bias scores on the original IAT *d*-score scale (.35) have respective predicted probabilities of in- or out-of-school suspensions of approximately 14% and 16%; for White students these are about 5% for both outcomes. For a county at the cutoff between "little or no bias" towards Whites and "slight bias" (.15), the analogous predicted probabilities for in- or out-of-school suspensions are closer: for Black students, they are 11% and 7%; for White students, they are 5% and 3%. Though no counties in our sample have implicit bias estimates of zero (i.e., no preference for either Whites or Blacks), extrapolation suggests that these disciplinary disparities would approximate zero.

<Table 4>

<Figure 1>

Discussion

Few studies have measured and predicted the implicit racial biases of educators in the US, and fewer have linked teachers' biases to student outcomes. In this study, we find that teachers' implicit Black/White biases vary depending on teacher gender and race/ethnicity: female teachers appear less biased than male teachers, and teachers of color appear less biased than White teachers. In general, our contextual and instructional variables have little predictive value for teachers' implicit biases, though teachers tend to show lower adjusted levels of bias in counties with larger shares of Black students. Overall, counties with higher aggregate levels of implicit and explicit bias tend to have larger adjusted White/Black test score inequalities and suspension disparities. These associations hold even when focusing only on teachers' biases and after accounting for a wide range of county-level covariates. Before further interpreting these results, we consider some data limitations.

Data Limitations

As noted earlier, one limitation of the study is the self-selection of respondents into the Project Implicit data. Even though we adjust county-level bias scores to account for the non-representativeness of the IAT respondent sample based on observable differences, if stratification weights fail to capture important unobserved determinants of implicit bias, any county-level estimates would still be biased. For example, people particularly aware of their own implicit racial biases may be taking the race IAT—this may bias estimates of implicit preferences towards Whites downwards (if awareness is correlated with lower bias). Another possibility is that school districts with especially significant inequality may be compelling their staffs to take the race IAT as a launching point for professional development targeting implicitly held attitudes and stereotypes. We therefore urge caution when interpreting or generalizing our findings regarding the implicit racial biases of educators. Additionally, the county identifiers we use to link Project Implicit data with SEDA and CRDC identify where teachers complete the IAT; we cannot confirm these are the counties in which they actually teach. With these limitations in mind, we proceed with interpreting our results.

Interpreting Descriptive Results

It is somewhat reassuring to see that teachers in counties with larger shares of Black students have relatively lower levels of implicit bias, as the reverse would be worrisome. Of course, the explanation for this association cannot be determined from these data. Teachers with lower levels of implicit anti-Black bias may be more interested in working in counties with more Black students, may be more likely to remain teaching in these counties over time, or may be more likely to be hired in these counties. Teachers may also become less biased over time by working in counties with more Black students.

For RQ2, where we investigate the relationships between bias and Black-White student disparities, our results are consistent with theory. As would be expected, test score differences are larger in counties with stronger preferences for Whites. These results depend on including county-level covariates in models, stressing the need to consider contextual differences across counties when relating bias to outcomes. We similarly find that Black students are suspended at higher rates than White students in counties with stronger preferences for Whites. With regards to prior empirical work, these results for discipline outcomes generally converge with those from Riddle and Sinclair (2019), the only existing study on this topic, despite analytic differences (e.g., we focus on the biases of all respondents and not just White respondents; we use slightly different covariates in our MrP model; we use data from all CRDC years).

Bias and Student Outcomes: Theoretical Implications

As noted, we are only able to examine, in an exploratory manner, the non-causal associations between aggregate bias and student outcomes. The self-selection of respondents into the Project Implicit data prevents us from confidently generalizing about the levels of implicit bias in particular counties. Additionally, our design does not allow us to describe the causal mechanisms behind any observed associations in the data. Instead, our results raise questions that should be explored in future research.

According to the bias of crowds theory, the racial context in which one is embedded influences one's automatic racial associations. The implicit bias scores of people within a county therefore provide information about the racial context of that county, rather than simply describing stable, independent attitudes of people who happen to reside in that county. Even though Project Implicit respondents are a self-selected group, the bias of crowds theory suggests that their aggregate biases proxy for structural forces that lead to unequal outcomes by race: their implicit bias is "a psychological marker of systemic prejudice in the environment" (Payne et al., 2017, p 239). In counties where Black residents face more discrimination and more formidable structural barriers (such as economic and housing opportunities, disproportionate policing), negative stereotypes of Black Americans will be more accessible in the minds of IAT test-takers. Implicit bias can then serve as a mechanism that converts systemic prejudice into individual acts of discrimination (Payne et al., 2017). Thus, observed associations between aggregate biases and student outcomes may arise partly from students' experiences of racial discrimination outside of school, and partly from the structural forces that jointly produce racial bias and inequalities in educational outcomes. At the same time, the vast majority of the variation in teachers' (and non-teachers') implicit biases occur within counties (Table 2). This may indicate that a level of analysis lower than the county is necessary when applying the bias of crowds theory. For example, teacher bias may vary more at the school level, and school-level teacher bias may more strongly predict school-level racial disparities in student outcomes.

Future Directions

One natural extension of our study would be look beyond this paper's focus on individuals' racial attitudes towards Black Americans and examine measures of bias towards other groups to understand how they influence other students' outcomes. Furthermore, because race is socially constructed and thus changes over time and across contexts (Haney López, 1994) the work of developing measures of bias and investigating their impacts need to be ongoing.

Future quantitative work should specifically seek exogenous sources of variation in the implicit racial bias of educators to help determine whether they have direct, indirect, or proxy effects on student outcomes, and help to uncover the level of analysis that is most meaningful for examining these questions. Finally, qualitative work (e.g., interviews with Black students and/or teachers) in particular can provide detailed insight unavailable from large quantitative studies on which of the theoretical mechanisms described in our literature review contribute most to relationships between teachers' bias and test score and/or disciplinary outcomes.

Conclusion

This study responds to calls from education researchers and social psychologists for incorporating theory and measures of implicit racial bias into education research (Quinn, 2017; Warikoo et al., 2016). These calls are particularly pressing given, among other reasons, the projected growth in the population of K-12 students of color and that present-day racist political rhetoric may even be counteracting years of improvement in explicit (if not implicit) racial attitudes (e.g., Schaffner, 2018). Our findings serve as a foundation for future research on teachers' implicit racial biases and they raise questions about the specific ways in which bias may contribute to racial disparities in educational outcomes, both at the interpersonal and the aggregate levels.

Notes

[1] The IAT uses the category labels “European American” and “African American.” We therefore use these terms when discussing components of the test specifically, and “White” and “Black” otherwise.

[2] Approximately 19% of Project Implicit site visitors did not respond to the occupation question. The occupation variable does not differentiate between public or private school teachers.

[3] As we describe in more detail in the Analytic Plan section, for RQ2 we adjust the county-level estimates of bias used to predict racial differences in outcomes to account for non-representativeness of the IAT respondent sample. For pooled bias scores (i.e., those using all respondents), we adjust scores using ACS data. For teacher bias scores, we adjust scores for fewer counties due to data limitations, described in Appendix C. These limitations restricted county coverage, resulting in a common sample of counties representing 33% and 35% of counties with K-12 teacher IAT data or student outcome gaps, respectively. In the results, we show that patterns of findings across the “Pooled” sample (i.e., all counties that we can adjust IAT scores using ACS data) and “Teacher” sample (i.e., the subset of counties that we can adjust IAT scores using available teacher demographic data) are similar.

References

Babad, E. (1993). Teachers' differential behavior. *Educational Psychology Review, 5(4),* 347-376.

Blair, I. V., & Brondolo, E. (2017). Moving beyond the individual: Community-level prejudice and health. *Social Science & Medicine*, *183*, 169-172.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *Quarterly Journal of Economics.* Advanced online publication. https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjz008/5368349

Chetty, R., Hendren, N., Jones, M.R., & Porter, S.R. (2018). Race and economic opportunity in the United States: An intergenerational perspective. NBER Working Paper 24441.

Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5-18.

Dovidio, J.F., Kawakami, K., & Gaertner, S.L. (2002). Implicit and explicit prejudice and interracial interactions. *J. Pers. Soc. Psychol.* 82:62–68

Fahle, E. M., Shear, B.R., Kalogrides, D., Reardon, S.F., Chavez, B. & Ho, A.D. (2019). Stanford Education Data Archive technical documentation. Version 3.0. Retrieved from: https://stacks.stanford.edu/file/druid:db586ns4974/SEDA_documentation_v30_09212019.pdf

Ferguson, A. A. (2000). *Bad boys: Public school and the making of Black masculinity*. Ann Arbor: University of Michigan Press.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban education*, *38*(4), 460-507.

Green, A.R., Carney, D.R., Pallin, D.J., Ngo, L.H., Raymond, K.L., Iezzoni, L.I. & Banaji, M.R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white parents. *Journal of General Internal Medicine, 22,* 1231-1238.

Greenwald, A.G., & Krieger, L.H. (2006). Implicit bias: Sci. foundations. *CA Law Rev, 94,* 945-967. Greenwald, A.G., Nosek, B.A., & Sriram, N. (2005). Consequential validity of the Implicit Association Test. Retrieved from: https://osf.io/preprints/psyarxiv/tnfgr/download

Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., & Banaji, M.R. (2009). Understanding and using the Implicit association test: III. Meta-analysis of predictive validity. *Jour Pers Soc Psy, 97,* 17-41.

Gregory, A., Skiba, R.J., & Noguera, P.A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher, 39(1),* 59-68.

Haney López, I. F. (1994). The social construction of race. In Richard Delgado (Ed.) *Critical race theory: The cutting edge.* Philadelphia, PA: Temple University Press. (pp. 191-203)

Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social psychological and personality science*, *9*(4), 393-401.

Hoover, J., & Dehghani, M. (2019). The big, the bad, and the ugly: Geographic estimation with flawed psychological data. *Psychological Methods*.

Jacoby-Senghor, D.S., Sinclair, S. & Nicole Shelton, J. (2015). A lesson in bias: The relationship between implicit racial bias and performance in pedagogical contexts, *Jour Exp Soc Psy*.

Kumar, R., Karabenick, S.A., & Burgoon, J.N. (2015). Teachers' implicit attitudes, explicit beliefs, and the mediating role of respect and cultural responsibility on mastery and performance-focused instructional practices. *Journal of Educational Psychology, 107,* 533-545.

Lacoe, J., & Steinberg, M. P. (2019). Do Suspensions Affect Student Outcomes?. *Educational Evaluation and Policy Analysis, 41*(1), 34-62.

Leitner, J.B., Hehman, E., Ayduk, O., & Mendoza-Denton, R.(2016b).Racial bias is associated with ingroup death rate for Blacks and Whites. *Soc Sci Med, 170,* 220-227.

McConnell, A.R., & Liebold, J.M. (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology, 37,* 435-442.

McKown, C., & Weinstein, R.S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development, 74(2),* 498-515.

doi:10.1016/j.jsp.2007.05.001
McKown, C., Gregory, A., Weinstein, R. (2010). Expectations, stereotypes, and self-fulfilling prophecies in classroom and school life. In Meece, J., & Eccles, J., Ed. *Handbook of Research on Schools, Schooling, and Human Development.* New York: Routledge.
Nosek, B., Smyth, F.L., Sriram, N., Linder, N.M., Devos, T., Ayala, A, …& Greenwald, A.G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *PNAS, 106(26),* 10593-10597.
Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of personality and social psychology*, *105*(2), 171.
Papageorge, N.W., Gershenson, S., & Kang, K. (2016). Teacher expectations matter. IZA Discussion Paper No. 10165. Retrieved from: http://ftp.iza.org/dp10165.pdf
Payne, B.K., Vuletich, H.A., & Lundberg, K.B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry, 28*, 233-248.
Pearman, F. A. (2020). *County-level rates of implicit bias predict black-white test score gaps in US schools. EdWorkingPaper No. 20-192.* Providence, RI: Annenberg Brown University.
Pearman, F. A., Curran, F. C., Fisher, B., & Gardella, J. (2019). Are Achievement Gaps Related to Discipline Gaps? Evidence From National Data. *AERA Open*, *5*(4), 2332858419875440.
Perry, T., Steele, C.M., & Hillard, A. (2003) *Young, gifted and black: Promoting high achievement among African-American students.* Boston, MA: Beacon Press.
Peterson, E.R., Rubie-Davies, C., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement. *Learning and Instruction, 42*, 123-140.
Project Implicit. (n.d.). Race Attitude. Retrieved from: https://implicit.harvard.edu/implicit/demo/background/raceinfo.html
Quinn, D. M. (2017). Racial attitudes of preK–12 and postsecondary educators: Descriptive evidence from nationally representative data. *Educational Researcher*, *46*(7), 397-411.
Reardon, S.F., Ho, A.D., Shear, B.R, Fahle, E. M., Kalogrides, D., Jang, H., Chavez, B., Buontempo, J., & DiSalvo, R.. (2019a). Stanford Education Data Archive (Version 3.0). http://purl.stanford.edu/db586ns4974.
Reardon, S.F., Kalogrides, D., & Shores, K. (2019b). The geography of racial/ethnic test score gaps. *American Journal of Sociology, 124,* 1164-1221.
Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proceedings of the National Academy of Sciences*, *116*(17), 8255-8260.
Schaffner, B. F. (2018). *Follow the racist? The consequences of Trump's expressions of prejudice for mass rhetoric.*
Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The urban review*, *34*(4), 317-342.
Steele, C.M. & Aronson, J. (1995) Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-811.
Taylor, V.J., & Walton, G.M. (2011). Stereotype threat undermines academic learning. *Personality and Social Psychology Bulletin, 37(8),* 1055-1067. DOI:10.1177/0146167211406506
Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*, 253-273.
van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R.W. (2010). The implicit prejudiced attitudes of teachers. *American Education Research Journal, 47,* 497-527.
Warikoo, N., Sinclair, S., Fei, J., & Jacoby-Senghor, D. (2016). Examining racial bias in education: A new approach. *Educational Researcher, 45,* 508-514.
Xu, K., Nosek, B, & Greenwald, A.G. (2014). Data from the race implicit association test on the Project Implicit demo website. *Jour Open Psy Data, 2(1): e3*

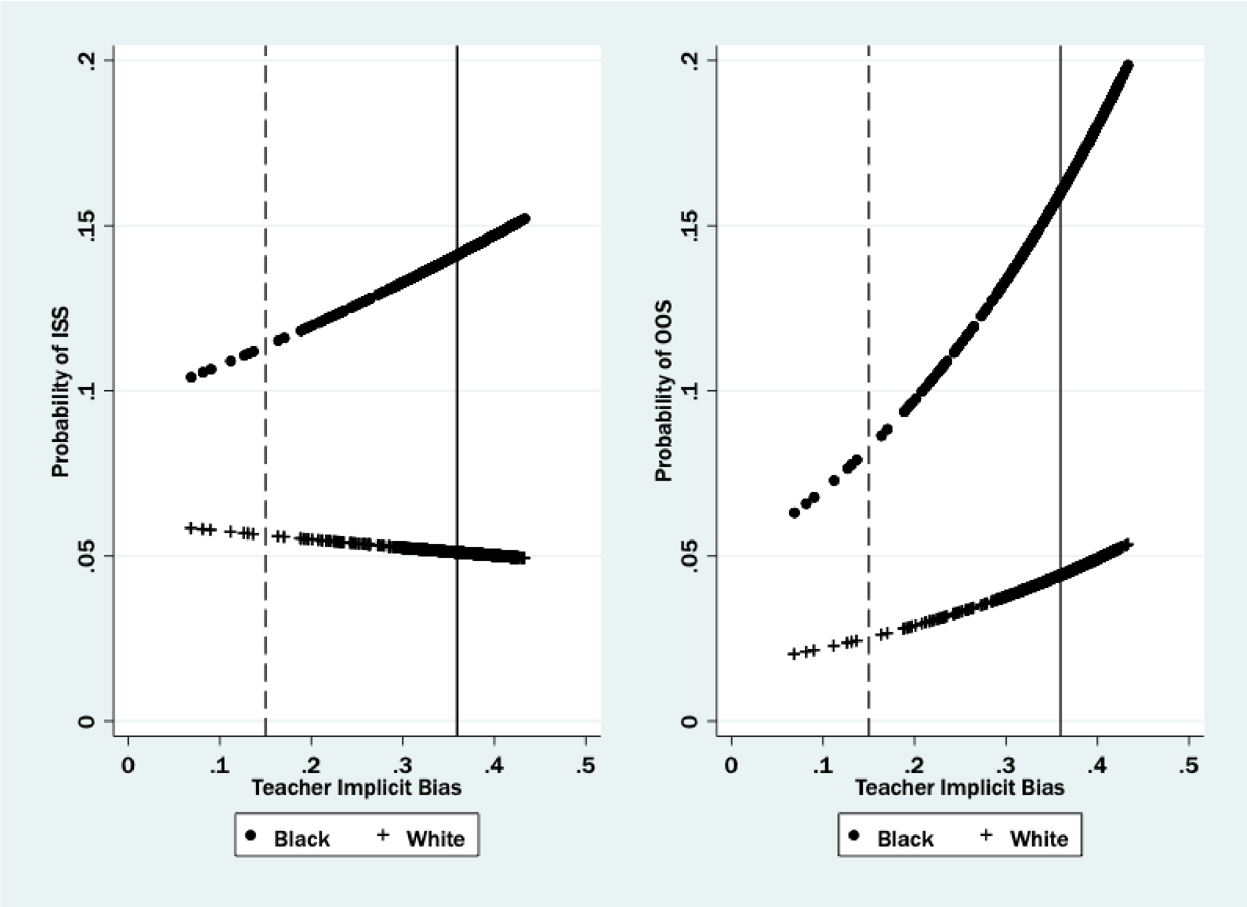

Figure 1. Predicted probabilities for in-school suspension (ISS) and out-of-school suspension (OOS) by race against county-level teacher implicit bias (original IAT d-scale), adjusted for representativeness with values for contextual controls set at the mean. The black solid line identifies the county-level mean for teacher implicit bias. The black dashed line identifies the IAT d-scale cutoff of .15 that distinguishes between "little or no" White bias versus "slight" White bias.

Table 1.
Descriptive Statistics for K-12 Educators

	Mean	SD	Nationwide
Respondent-level Project Implicit Data (N respondents = 39581, N counties = 1673)			
Age < 30	0.389		Average age: 42.4
Age 30-39	0.287		
Age 40-49	0.175		
Age 50-59	0.106		
Age 60-69	0.037		
Age 70+	0.006		
American Indian	0.004		0.004
White	0.803		0.801
Black	0.089		0.067
Black+White	0.013		
East Asian	0.010		0.025 (APIA)
Multi-racial	0.033		0.014
Native Hawaiian	0.003		
Other race (unspecified)	0.040		
South Asian	0.006		
Educ: Elem-some HS	0.006		
Educ: HS degree	0.008		
Educ: Some college/Assoc. deg.	0.086		
Educ: Bach degree	0.261		0.405
Educ: Master's degree	0.590		0.573
Educ: Advanced degree	0.049		
Female	0.714		0.766
IAT d-score	0.324	0.455	
2008-2009 School year	0.159		
2009-2010 School year	0.142		
2010-2011 School year	0.096		
2011-2012 School year	0.085		

2012-2013 School year	0.086	
2013-2014 School year	0.099	
2014-2015 School year	0.178	
2015-2016 School year	0.156	
County-level OCR Data (N counties = 1673)		
Student enrollment: Black	13435.500	43410.700
Student enrollment: White	40560.500	63568.300
Prob. In-school suspension: Black	0.145	0.080
Prob. Out-of-school suspension: Black	0.133	0.060
Prob. In-school suspension: White	0.062	0.039
Prob. Out-of-school suspension: White	0.047	0.029
County-level SEDA Test Data (N counties = 1673)		
White-Black achievement gap	0.547	0.222
County-level SEDA Covariate Data (N counties = 1673)		
SES composite (all)	-0.171	0.736
Proportion Black in public schools	0.155	0.197
Proportion Hispanic in public schools	0.125	0.159
Between-school free lunch/not free lunch segregation	0.040	0.051
Between-school Black/White segregation	0.072	0.085
Proportion attending charter schools	0.017	0.042
Per-pupil instructional expenditures in average student's school (in $10000)	0.597	0.152
Average student-teacher ratio	16.324	17.094
White-Black gap in SES composite	1.817	1.673
White-Black school free lunch rate difference	-0.039	0.062
White/Black relative student-teacher ratio	1.013	0.032
White-Black charter school enrollment rate difference	0.001	0.030

Note. Variables in rows without reported SD are binary indicator variables for the row name. Statistics for the Nationwide column come from the National Teacher and Principal Survey (2015-16) and include estimates for only public school teachers. SES=composite measure of socioeconomic status (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher).

Table 2.
Multilevel Models Predicting IAT Scores, K-12 Educators only

	1	2	3	4
Am. Indian		-0.0874*		-0.0898**
		(0.0341)		(0.0341)
East Asian		-0.00643		-0.00543
		(0.0221)		(0.0221)
South Asian		-0.0966***		-0.0933**
		(0.0293)		(0.0293)
Native Haw.		-0.125**		-0.127**
		(0.0408)		(0.0408)
Black		-0.435***		-0.429***
		(0.00791)		(0.00802)
Black+White		-0.220***		-0.219***
		(0.0192)		(0.0192)
Other multi-racial		-0.143***		-0.142***
		(0.0124)		(0.0124)
Race: Other/unknown		-0.0945***		-0.0941***
		(0.0114)		(0.0114)
Female		-0.0230***		-0.0231***
		(0.00488)		(0.00487)
Age: 30-39		-0.0158**		-0.0169**
		(0.00569)		(0.00569)
Age: 40-49		-0.0360***		-0.0381***
		(0.00665)		(0.00665)
Age: 50-59		-0.0220**		-0.0238**
		(0.00797)		(0.00797)
Age: 60-69		-0.0173		-0.0190
		(0.0122)		(0.0122)
Age: 70+		0.0415		0.0393
		(0.0291)		(0.0291)
Educ: HS degree		0.000670		0.000905
		(0.0383)		(0.0382)
Educ: Some college		0.0224		0.0218
		(0.0302)		(0.0301)
Educ: Bachelors		-0.0125		-0.0103
		(0.0296)		(0.0296)
Educ: Masters		-0.00881		-0.00587
		(0.0295)		(0.0295)
Educ: Advanced deg.		-0.0221		-0.0193

		(0.0310)		(0.0310)
SES Composite			-0.0181**	-0.00669
			(0.00623)	(0.00559)
Prop. Black			-0.299***	-0.0860**
			(0.0288)	(0.0264)
Prop. Hispanic			-0.0543*	0.0108
			(0.0267)	(0.0241)
Info index FRL/not FRL			0.00154	0.0215
			(0.0628)	(0.0538)
Info index White/Black			-0.0872	-0.0164
			(0.114)	(0.0990)
Prop. Charter			-0.0587	-0.166**
			(0.0609)	(0.0536)
PPE Instruction			0.0439	0.0162
			(0.0279)	(0.0263)
Stu/teach ratio			0.00000228	0.000217
			(0.000443)	(0.000421)
FRL: W-B			-0.109	-0.0121
			(0.112)	(0.0975)
Prop. Charter: W-B			0.0406	0.113
			(0.0952)	(0.0830)
Stu/Teach: W/B			-0.206	-0.163
			(0.121)	(0.108)
SES Composite: W-B			-0.00116	-0.00477
			(0.00296)	(0.00274)
Constant	0.352***	0.420***	0.598***	0.599***
	(0.00836)	(0.0303)	(0.121)	(0.111)
ICC County	0.0198	0.00957	0.00997	0.00796
ICC State	0.00580	0.00464	0.00462	0.00593

Note: All models include random intercepts for counties and states. Sample size for each column is 39581 respondents and 1673 counties. All models control for year fixed effects. SES=composite measure of socioeconomic status (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher); SES Composite: W-B= White/Black differences on the SES composite; FRL: W-B=White/Black differences in school free lunch rates; Prop. Black=proportion of Black students in public schools; Info index W/B=between-school White/Black segregation (measured by the Theil information index, which equals 0 when all schools in a district have the same racial composition as the district overall, and 1 when schools contain only one racial group); Info index FRL/not FRL=between-school free lunch/not free lunch segregation; PPE instruction=per-pupil instructional expenditures; Stud/teach ratio=average student-teacher ratio; Stu/Teach: W/B = White/Black ratio for student-teacher ratios; Prop in charters=proportion of public school students attending charter schools; Prop. Charter: W-B=White/Black differences in charter enrollment rates. $^{\sim}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.
Predictors of County-Level Racial Test Score Inequality

	1	2	3	4	5	6
Panel A. Implicit Bias						
All	-0.0493***	0.0334*	-0.0647***	0.00178		
	(0.00629)	(0.0168)	(0.00910)	(0.0254)		
Teacher					-0.0571***	0.0814***
					(0.0103)	(0.0162)
Panel B. Explicit Bias						
All	-0.0496***	0.0249~	-0.0674***	0.00743		
	(0.00611)	(0.0129)	(0.00923)	(0.0195)		
Teacher					-0.0569***	0.0697***
					(0.00987)	(0.0151)
Sample	Pooled	Pooled	Teacher	Teacher	Teacher	Teacher
N Counties	1673	1673	746	746	746	746
Covariates		Yes		Yes		Yes

Note. Standard errors in parentheses. All models include state fixed effects. Outcome is county's mean standardized White-Black test score difference, pooled across grades and subjects (cohort standardized scale). Bias predictors are county-level empirical Bayes predicted means, standardized to county-level SD of 1 and mean of 0. The Pooled sample consists of all counties used across analyses; the Teacher sample consists of these counties but subset to those that with data used to adjust teacher bias scores based on representativeness. Covariates include: SES composite (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher), W-B difference in SES composite, W-B difference in free lunch %, percent public school students Black, percent public school students Hispanic, per-pupil instructional expenditure, student/teacher ratio, percent public school students attending charter school, W-B difference % charter, W/B ratio of student/teacher ratio, segregation indices. Estimated from a meta-regression performed by methods of moments. $^{\sim}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.
Logistic Regression Models Predicting In- and Out-of-School Suspensions

	1	2	3	4	5	6
Panel A: In-school suspensions on implicit bias						
Black	1.070***	1.133***	1.042***	1.117***	1.068***	1.116***
	(0.0225)	(0.0179)	(0.0330)	(0.0220)	(0.0262)	(0.0242)
Bias (all)	0.0953***	0.0213	0.126**	-0.0281		
	(0.0270)	(0.0647)	(0.0432)	(0.133)		
Blk*Bias (all)	0.0800**	0.143***	0.0692~	0.130***		
	(0.0254)	(0.0226)	(0.0387)	(0.0356)		
Bias (tch)					0.133***	-0.0265
					(0.0387)	(0.0690)
Blk*Bias (tch)					0.0617~	0.0910**
					(0.0324)	(0.0305)
Panel B: Out-of-school suspensions on implicit bias						
Black	1.395***	1.451***	1.355***	1.422***	1.446***	1.417***
	(0.0266)	(0.0231)	(0.0365)	(0.0293)	(0.0325)	(0.0304)
Bias (all)	-0.0890***	-0.0549	-0.0600~	-0.0509		
	(0.0244)	(0.0534)	(0.0328)	(0.0742)		
Blk*Bias (all)	0.0357	0.101***	0.00234	0.0568		
	(0.0260)	(0.0258)	(0.0356)	(0.0393)		
Bias (tch)					0.0153	0.149**
					(0.0472)	(0.0571)
Blk*Bias (tch)					0.0375	0.0440
					(0.0323)	(0.0300)
Panel D: In-school suspensions on explicit bias						
Black	1.096***	1.140***	1.091***	1.135***	1.103***	1.132***
	(0.0207)	(0.0181)	(0.0289)	(0.0233)	(0.0279)	(0.0255)
Bias (all)	0.104***	0.0206	0.137***	0.0270		
	(0.0242)	(0.0458)	(0.0374)	(0.0840)		
Blk*Bias (all)	0.0901***	0.139***	0.104**	0.136***		
	(0.0249)	(0.0234)	(0.0348)	(0.0343)		
Bias (tch)					0.141***	0.0133
					(0.0343)	(0.0615)
Blk*Bias (tch)					0.0680*	0.0859**
					(0.0316)	(0.0296)
Panel C: Out-of-school suspensions on explicit bias						
Black	1.424***	1.461***	1.412***	1.443***	1.495***	1.442***
	(0.0237)	(0.0236)	(0.0313)	(0.0312)	(0.0356)	(0.0338)
Bias (all)	-0.0722**	-0.0140	-0.0472	0.0633		
	(0.0236)	(0.0357)	(0.0317)	(0.0494)		
Blk*Bias (all)	0.0530*	0.110***	0.0581	0.0935**		
	(0.0245)	(0.0248)	(0.0373)	(0.0362)		
Bias (tch)					-0.0104	0.118**
					(0.0364)	(0.0418)
Blk*Bias (tch)					0.0932**	0.0714*

					(0.0352)	(0.0322)
Sample	Pooled	Pooled	Teacher	Teacher	Teacher	Teacher
N	90335233	90335233	48988187	48988187	48988187	48988187
N counties	1673	1673	746	746	746	746
Covariates	No	Yes	No	Yes	No	Yes

Note. Standard errors clustered at the county level in parentheses. All models include state fixed effects. Models fit using aggregate county*race data pooled over the 2011-2012, 2013-2014, and 2015-2016 school years with frequency weights to mimic models with student data pooled across years. Bias predictors are county-level empirical Bayes predicted means standardized to mean=0, SD=1. The Pooled sample consists of all counties used across analyses; the Teacher sample consists of these counties but subset to those that with data used to adjust teacher bias scores based on representativeness. Covariates include: SES composite (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher), W-B difference in SES composite, W-B difference in free lunch %, percent public school students Black, percent public school students Hispanic, per-pupil instructional expenditure, student/teacher ratio, percent public school students attending charter school, W-B difference % charter, W/B ratio of student/teacher ratio, segregation indices. $^{\sim}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix A. Details on Data and Covariates

Civil Rights Data Collection

The CRDC includes variables related to student enrollment, demographics, and discipline for the 2011-12, 2013-2014, and 2015-16 school years. These data are collected by surveying all public local educational agencies (LEAs) in the U.S. on a biennial basis, and data fields are reported at the school level. While the CRDC data set includes a range of data fields, we utilize variables that are associated with school discipline and student demographics including: (a) student enrollment counts, (b) counts of students with one or more in-school suspension, (c) counts of students with one out-of-school suspension, and (d) counts of students with more than one out-of-school suspension. These data fields are disaggregated by race, and we use these counts to construct the following measures: (a) count of Black students enrolled in school, (b) count of White students enrolled in school, (c) count of Black students with one or more in-school suspension, (d) count of White students with one or more in-school suspension, (e) count of Black students with one or more in-school suspension, and (f) count of Black students with one or more out-of-school suspensions.

In constructing these count variables, we ultimately aggregate school-level data to the county level, but we first aggregated to the district level for descriptive purposes. When aggregating to the LEA-level, we drop LEAs where all schools do not report the data fields of interest due to privacy concerns or data reporting errors, which results in 28 LEAs dropped in 2011-12, 93 LEAs dropped in 2013-14, and 74 LEAs dropped in 2015-16. In the 2011-12 school year, 95,635 schools and 16,503 LEAs were recorded in the CRDC data set; the 2013-14 SY includes 95,507 schools and 16,758 LEAs; and the 2015-16 SY includes 96,360 schools and 17,337 LEAs.

We then aggregate LEA-by-school-year-level Black and White enrollment and suspension counts to the county level (pooled across school years). Using data from the Common Core of Data, we linked LEAs to their most recently assigned county between the 2008-09 and 2015-16 school year; 99% of LEAs had county links. In these county-level aggregated data, there were no instances of subgroup suspension proportions above 1. At the LEA-by-school-year level, less than 1% of LEA-by-school-year observations had subgroup suspension proportions above 1.

SEDA Covariates

The SEDA covariate (v 2.1) dataset contains information gathered from the Education Demographic and Geographic Estimates and the Common Core of Data (Fahle et al., 2019). In several analyses, described in more detail below, we included several predictors from this dataset similar to those used in Reardon et al.'s (2019) exploration of the geography of achievement gaps. These covariates included contextual variables such as: (a) a composite measure of socioeconomic status (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher), (b) White/Black differences on the SES composite, (c) White/Black differences in school free lunch rates, (d) proportion of Black students in public schools, (e) proportion of Hispanic students in public schools; and instructional variables such as (f) per-pupil instructional expenditures (g) average student-teacher ratio, (h) White/Black differences in student-teacher ratios, (i) proportion of public school students attending charter schools, (j) White/Black differences in charter enrollment rates, and (k) between-school White/Black (FRPL/Non FRPL) segregation (measured by the Theil information index, which equals 0 when all schools in a district have the same racial [FRPL] composition as the district overall, and 1 when schools contain only one racial [FRPL] group).

The SEDA covariate dataset is at the district level. Thus, we first linked LEAs to county identifiers; 99% of districts had identifiers between the 2008-09 and 2015-16 school years. We assign LEAs their most recent county. We then aggregated the above-referenced controls to the county level by taking the weighted means of the LEA-level data (weighting by the relevant demographic group for subgroup-specific variables).

Appendix B. Suspension analyses

We transformed the aggregate CRDC count data so that it could be analyzed at the student level.[1] In essence, we created a data set that mimicked a data set in which observations were at the student level, with each student observation assigned a 0/1 indicator for whether that student was suspended one or more times that school year (with indicators for in-school and out-of-school suspensions). Students also had race indicators (Black or White) and county identifiers, to match them to county-level bias EB scores and county-level covariates.

The student level data set can be created by expanding the county-by-race counts for the number of students suspended and not suspended. For example, across CRDC years, if a county has 100 Black students and 5 are suspended, this county would be given 95 rows of 0s for the "suspension" variable for Black students and 5 rows of 1s (though note that our counts are county-by-race). Then, one can fit a logistic regression model to the data. For computational efficiency, we created an equivalent data set using frequency weights rather than assigning each student observation a row in the data. We illustrate with Table B1 below.

[1] Several options are available for analyzing aggregate count-level data such as CRDC suspension data, but none is ideal for our purposes. In a meta-analytic framework, county-level suspension risk differences by race can be predicted by county-level bias scores. The drawback to this approach is that the magnitude of risk differences of rare events such as suspensions are difficult to interpret, and the metric erases variation in baseline risk rates. County-level racial differences in the log-odds of being suspended could also be modelled in a meta-analytic framework, but counties with zero values must be either dropped or have an arbitrary constant added to them, both of which introduce bias (Rucker, Schwarzer, Carpenter, & Olkin, 2009). The arcsine transformation can handle zero values, but meta-analyses with arcsine transformations yield biased estimates when groups are unbalanced (as in CRDC; Rucker et al., 2009). Another option is the fraction response model (Papke & Wooldridge, 1996), which can handle proportions of 0 or 1. In this approach, a single model is fit across the whole range of data. This raises issues of interpretation in cases such as ours, because it does not allow for an alternative model that generates the 0 (or 1) values. For example, counties may have no out-of-school suspensions because out-of-school suspensions are prohibited; or, school-level prohibition of suspensions within counties may be correlated with the racial make-up of the school. In this case, it may not make sense to use a single model to predict suspensions for counties with zero and non-zero suspension counts. A zero-inflated beta model (Cook, Kieschnick, & McCullough, 2008) separates the model for zeros and non-zero proportions, but does not allow for unity values. As sensitivity checks, we fit alternative models using some of these approaches and the general findings from the main text were replicated: we found significant adjusted interactions between aggregate county-level bias and student race.

In Table B1, "black" is an indicator for Black students (1=yes), "county_iat" is the county-level EB score for implicit bias (aggregated across years), and "susp" is the outcome variable, an indicator for students who were suspended (1=yes). The ISS and OOS variables are used as frequency weights for the in-school suspension and out-of-school suspension analyses, respectively. These variables represent the sum of the student observations with that row's combination on the *black* and *susp* variable. For example, row 1 in Table B1 shows that the sum of the number of student observations for White students in the county who received at least one in-school suspension was 1254; the sum for out-of-school suspensions was 319. In row 2, we see that the sum of the number of student observations for White students who did not receive an in-school suspension was 18,723.

Using a data file with data stored as in Table B1 along with the frequency weights, we fit logistic regression models of the form described in the main text:

$$P(Y_{ic} = 1|\boldsymbol{X}_c) = \frac{1}{1+\exp(-(\beta_1 Black_i + \beta_2 \widehat{\delta_j} + \beta_3 (Black_i \times \widehat{\delta_j}) + X_c + \gamma))} \quad (3)\ (B1)$$

The outcome, Y_{ic} is an indicator for whether student *i* in county c was suspended one or more times. This is the "susp" variable in Table B1. When analyzing in-school suspensions, we apply the "ISS" frequency weight; when analyzing out-of-school suspensions, we apply the "OOS" frequency weight. In equation B1, $Black_i$ is an indicator for whether student *i* is Black (versus White), and the county-level EB $\hat{\delta}_j$ is standardized at the county level (we fit models with and without the SEDA county covariates in X_c); γ represents the vector of state fixed effects. We fit these models in Stata 15.1 MP using the "logit" command.

Table B1.
Example of data set-up for frequency-weighted logistic regression models used to answer RQ2 in the main text.

countyid	black	susp	ISS	OOS	county_iat
1001	0	1	1254	319	0.355846
1001	0	0	18723	19658	0.355846
1001	1	1	1091	344	0.355846
1001	1	0	5920	6667	0.355846

Appendix C. Full Sample Analyses

As noted in our main text, we use a common sample for analyses across RQs. Here we replicate these analyses but make as few restrictions as possible on the sample, i.e., all observations without missing data for outcomes and controls are included regardless of their inclusion in other analyses. Table C1 below shows our results for RQ1 without sample restrictions, with Tables C2 and C3 showing our results for the test score and disciplinary gap analyses, respectively, of RQ2. Overall, we find our main findings robust to this sample restriction.

In Table C4, we use county-level data from the 2015 American Community Survey 5-year (ACS) to compare differences in population demographics across the following samples: "IAT" (any county that has at least one individual with implicit bias scores); "IAT K12" (any county that has at least one K-12 teacher with implicit bias scores); "SEDA test score" (any county that has Black-White test score gaps from SEDA); "CRDC" (any county with CRDC disciplinary data); "Pooled" (our primary analytic common sample); and "Teacher" (a subset of Pooled sample counties that we can adjust teacher bias scores for representativeness). Though the number of counties included in each sample varies, characteristics across samples are very similar, further assuaging concerns over our sample restrictions for analyses.

Table C1.
Multilevel Models Predicting IAT Scores, K-12 Educators only

	1	2	3	4
Am. Indian		-0.0888**		-0.0888**
		(0.0331)		(0.0338)
East Asian		-0.0145		-0.00522
		(0.0213)		(0.0220)
South Asian		-0.0880**		-0.0930**
		(0.0288)		(0.0293)
Native Haw.		-0.123**		-0.126**
		(0.0398)		(0.0406)
Black		-0.435***		-0.429***
		(0.00773)		(0.00801)
Black+White		-0.220***		-0.217***
		(0.0187)		(0.0192)
Other multi-racial		-0.148***		-0.142***
		(0.0121)		(0.0124)
Race: Other/unknown		-0.0987***		-0.0950***
		(0.0111)		(0.0113)
Female		-0.0242***		-0.0236***
		(0.00474)		(0.00484)
Age: 30-39		-0.0156**		-0.0169**
		(0.00554)		(0.00566)
Age: 40-49		-0.0342***		-0.0371***
		(0.00648)		(0.00661)
Age: 50-59		-0.0204**		-0.0228**
		(0.00777)		(0.00793)
Age: 60-69		-0.0135		-0.0171
		(0.0119)		(0.0121)
Age: 70+		0.0525		0.0477
		(0.0281)		(0.0286)
Educ: HS degree		0.0110		0.00249
		(0.0372)		(0.0378)
Educ: Some college		0.0269		0.0202
		(0.0295)		(0.0299)
Educ: Bachelors		-0.00709		-0.0118
		(0.0289)		(0.0294)
Educ: Masters		-0.00416		-0.00667
		(0.0288)		(0.0293)
Educ: Advanced deg.		-0.0174		-0.0201

		(0.0304)		(0.0308)
SES Composite			-0.0158**	-0.00624
			(0.00598)	(0.00548)
Prop. Black			-0.291***	-0.0888***
			(0.0277)	(0.0259)
Prop. Hispanic			-0.0477	0.0125
			(0.0258)	(0.0238)
Info index FRL/not FRL			-0.00637	0.0176
			(0.0606)	(0.0532)
Info index White/Black			-0.151	-0.0340
			(0.111)	(0.0982)
Prop. Charter			-0.0434	-0.161**
			(0.0593)	(0.0533)
PPE Instruction			0.0282	0.00892
			(0.0269)	(0.0258)
Stu/teach ratio			-0.0000729	0.000166
			(0.000441)	(0.000421)
FRL: W-B			-0.153	-0.0227
			(0.108)	(0.0965)
Prop. Charter: W-B			0.00997	0.0922
			(0.0928)	(0.0825)
Stu/Teach: W/B			-0.214	-0.157
			(0.114)	(0.104)
SES Composite: W-B			-0.00115	-0.00378
			(0.00269)	(0.00255)
Constant	0.353***	0.415***	0.616***	0.599***
	(0.00807)	(0.0296)	(0.113)	(0.108)
N	43455	41740	41692	40055
ICC County	0.0210	0.00986	0.00966	0.00785
ICC State	0.00551	0.00414	0.00448	0.00573
N Counties	2219	2200	1878	1868

Note: All models include random intercepts for counties and states. All models control for year fixed effects. SES=composite measure of socioeconomic status (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher); SES Composite: W-B= White/Black differences on the SES composite; FRL: W-B=White/Black differences in school free lunch rates; Prop. Black=proportion of Black students in public schools; Info index W/B=between-school White/Black segregation (measured by the Theil information index, which equals 0 when all schools in a district have the same racial composition as the district overall, and 1 when schools contain only one racial group); Info index FRL/not FRL=between-school free lunch/not free lunch segregation; PPE instruction=per-pupil instructional expenditures; Stud/teach ratio=average student-teacher ratio; Stu/Teach: W/B = White/Black ratio for student-teacher ratios; Prop in charters=proportion of public school students attending charter schools; Prop. Charter: W-B=White/Black differences in charter enrollment rates. $^{\sim}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C2.
Predictors of County-Level Racial Test Score Inequality

	1	2	3	4
Panel A. Implicit Bias				
All	-0.0398***	0.00687		
	(0.00544)	(0.0148)		
Teacher			-0.0579***	0.0862***
			(0.00997)	(0.0164)
N Counties	2111	1999	772	748
Panel B. Explicit Bias				
All	-0.0398***	0.0124		
	(0.00528)	(0.0119)		
Teacher			-0.0541***	0.0763***
			(0.00933)	(0.0150)
N Counties	2111	1999	782	757
Covariates		Yes		Yes

Note. Standard errors in parentheses. All models include state fixed effects. Outcome is county's mean standardized White-Black test score difference, pooled across grades and subjects (cohort standardized scale). Bias predictors are county-level empirical Bayes predicted means, standardized to county-level SD of 1 and mean of 0. Covariates include: SES composite (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher), W-B difference in SES composite, W-B difference in free lunch %, percent public school students Black, percent public school students Hispanic, per-pupil instructional expenditure, student/teacher ratio, percent public school students attending charter school, W-B difference % charter, W/B ratio of student/teacher ratio, segregation indices. Estimated from a meta-regression performed by methods of moments. $^{\sim} p < 0.10$, $^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Table C3.
Logistic Regression Models Predicting In- and Out-of-School Suspensions

	1	2	3	4
Panel A: In-school suspensions on implicit bias				
Bias (all)	0.0916***	0.0230		
	(0.0240)	(0.0600)		
Blk*Bias (all)	0.0801***	0.150***		
	(0.0233)	(0.0210)		
Bias (tch)			0.133***	-0.0234
			(0.0382)	(0.0683)
Blk*Bias (tch)			0.0605~	0.0910**
			(0.0322)	(0.0306)
N	98474405	95200012	50210541	49520093
Panel B: Out-of-school suspensions on implicit bias				
Bias (all)	-0.0988***	-0.0539		
	(0.0217)	(0.0501)		
Blk*Bias (all)	0.0455~	0.111***		
	(0.0243)	(0.0246)		
Bias (tch)			0.0117	0.152**
			(0.0464)	(0.0568)
Blk*Bias (tch)			0.0357	0.0436
			(0.0321)	(0.0299)
N	98474405	95200012	50210541	49520093
Panel D: In-school suspensions on explicit bias				
Bias (all)	0.0984***	0.0185		
	(0.0217)	(0.0433)		
Blk*Bias (all)	0.0892***	0.145***		
	(0.0227)	(0.0218)		
Bias (tch)			0.144***	0.0174
			(0.0344)	(0.0613)
Blk*Bias (tch)			0.0658*	0.0900**
			(0.0317)	(0.0303)
	98475184	95200012	50337027	49624095
Panel C: Out-of-school suspensions on explicit bias				
Bias (all)	-0.0810***	-0.0109		
	(0.0214)	(0.0341)		
Blk*Bias (all)	0.0593**	0.116***		
	(0.0227)	(0.0236)		
Bias (tch)			-0.0157	0.120**
			(0.0362)	(0.0421)
Blk*Bias (tch)			0.0913**	0.0742*
			(0.0343)	(0.0323)
N	98475184	95200012	50337027	49624095

Covariates	No	Yes	No	Yes

Note. Standard errors clustered at the county level in parentheses. All models include state fixed effects. Models fit using aggregate county*race data pooled over the 2011-2012, 2013-2014, and 2015-2016 school years with frequency weights to mimic models with student data pooled across years. Bias predictors are county-level empirical Bayes predicted means standardized to mean=0, SD=1. Covariates include: SES composite (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher), W-B difference in SES composite, W-B difference in free lunch %, percent public school students Black, percent public school students Hispanic, per-pupil instructional expenditure, student/teacher ratio, percent public school students attending charter school, W-B difference % charter, W/B ratio of student/teacher ratio, segregation indices. $^{\sim} p < 0.10$, $^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Table C4.
ACS County Comparisons across Samples

	IAT	IAT K12	SEDA test score	CRDC	Pooled	Teacher
Implicit Bias	0.339	0.342	0.329	0.339	0.333	0.342
ACS # Households	37550.883	50415.315	52766.068	37524.259	62963.527	80530.737
Median Household Income	46543.416	48207.785	47294.948	46494.528	48818.889	48761.777
Total Population	101508.259	136464.633	142951.151	101422.624	170518.537	218238.674
Prop. White	0.836	0.833	0.806	0.838	0.809	0.822
Prop. Black	0.091	0.094	0.123	0.091	0.118	0.106
Prop. Other Race	0.073	0.073	0.071	0.071	0.073	0.072
Prop. Male	0.500	0.498	0.497	0.500	0.496	0.495
Prop. Female	0.500	0.502	0.503	0.500	0.504	0.505
N Counties	3088	2216	2111	3086	1673	746

Note: County-level data come from the 2015 American Community Survey 5-year (ACS). Samples are: "IAT" (any county that has at least one individual with implicit bias scores); "IAT K12" (any county that has at least one K-12 teacher with implicit bias scores); "SEDA test score" (any county that has Black-White test score gaps from SEDA); "CRDC" (any county with CRDC disciplinary data); "Pooled" (our primary analytic common sample); and "Teacher" (a subset of Pooled sample counties that we can adjust teacher bias scores for representativeness).

Appendix D. Multilevel Regression and Post-stratification of Bias Scores

As noted in the main text, we used multilevel regression and post-stratification (MrP, Hoover and Dehghani, 2019) to adjust county-level pooled and teacher bias scores. For pooled scores, use county-level data from the 2015 American Community Survey 5-year (ACS) for this reweighting. Our MrP model included race, age, and gender variables, and used the ACS county-level population joint distribution for these variables to post-stratify scores. The process for reweighting pooled scores is as follows. First, we use the Project Implicit data to estimate county-level bias scores via the following i.i.d. Maximum Likelihood response model:

$$Y_{ijkl} = \beta + \alpha_j + \alpha_k + \alpha_{r[l]} + \alpha_{sXa[l]} + \varepsilon_{ijkl} \text{ (C1)}$$

Where Y_{ijkl} captures the bias scores of respondent i living in county j in state k in census division l. There are random effects for county α_j and state α_k, and random demographic effects by division for race $\alpha_{r[l]}$ and gender-by-age $\alpha_{sXa[l]}$. Following Hoover and Dehghani (2019), our categories for race are “black”, “white”, and “other”; for age our categories are “under 18”, “18 to 29”, “30 to 44”, “45 to 64”, and “over 65”. This model shrinks random effects to the mean for counties with fewer respondents to account for more uncertainty in estimates.

Using this trained model, predictions for bias were made for each cross-classification of race, age, and gender for each county. These predictions were reweighted and summed based on the proportion of the actual population represented by these cross-classifications to arrive at our final county-level MrP pooled bias score.

For teacher scores, we use publicly available data from states on teacher race—at the county-level or below. In Table D1, we document this state data.

Reweighting county-level teacher bias scores followed the same process for pooled scores, with a slightly different response model to account for the limited data on the marginal distributions of demographics within county. Specifically, we used Project Implicit teacher data to estimate county-level bias scores via the following i.i.d Maximum Likelihood response model:

$$Y_{ijk} = \beta + \alpha_j + \alpha_{r[k]} + \varepsilon_{ijk} \text{ (C2)}$$

Where Y_{ijk} captures the bias scores of respondent i living in county j in state k. There are random effects for county α_j and random demographic effects by state for race $\alpha_{r[k]}$. Following Hoover and Dehghani (2019), our categories for teacher race are "black", "white", and "other". Using this trained model, predictions for bias were made for each cross-classification of race for each county. These predictions were reweighted and summed based on the proportion of the actual teacher population represented by these cross-classifications to arrive at our final county-level MrP teacher bias score.

Table D1.
Publicly Available Data on County-Level Teacher Race

State	Publicly available data on teacher race at county, district, or school level?	Years used	Observation level	CCD Size Ranking 2017-18
CALIFORNIA	Yes	2009-2016	District	1
TEXAS	No			2
FLORIDA	Yes	2014-2016	District	3
NEW YORK	Yes	2009-2016	District	4
ILLINOIS	No			5
GEORGIA	No			6
PENNSYLVANIA	Yes	2016	County	7
OHIO	No			8
NORTH CAROLINA	Yes	2009-2016	District	9
MICHIGAN	Yes	2009-2016	District	10
NEW JERSEY	Yes	2009-2016	District	11
VIRGINIA	No			12
WASHINGTON	No			13
ARIZONA	Yes	2009-2016	County	14
INDIANA	No			15
TENNESSEE	Yes	2017	District	16
MASSACHUSETTS	Yes	2009-2016	District	17
MISSOURI	No			18
COLORADO	Yes	2019	District	19
MARYLAND	No			20
MINNESOTA	Yes	2009-2016	School	21
WISCONSIN	Yes	2009-2016	District	22
SOUTH CAROLINA	Yes	2016	District	23
ALABAMA	Yes	2015-2016	School	24
LOUISIANA	No			25
OKLAHOMA	No			26
KENTUCKY	Yes	2014-2015	District	27
UTAH	No			28
OREGON	No			29
CONNECTICUT	Yes	2009-2016	District	30
IOWA	No			31
ARKANSAS	Yes	2009-2016	County	32
KANSAS	No			33

NEVADA	No			34
MISSISSIPPI	No			35
NEW MEXICO	No			36
NEBRASKA	Yes	2012-2016	District	37
IDAHO	No			38
WEST VIRGINIA	No			39
HAWAII	No			40
NEW HAMPSHIRE	No			41
MAINE	No			42
MONTANA	No			43
RHODE ISLAND	No			44
SOUTH DAKOTA	No			45
DELAWARE	No			46
ALASKA	No			47
NORTH DAKOTA	No			48
WYOMING	No			49
DISTRICT OF COLUMBIA	Yes	2019	District	50
VERMONT	No			51

Appendix E. Clustering of Standard Errors

In the main text, we present results for the analyses predicting disciplinary outcomes using county-level bias scores with standard errors clustered at the county level—the level at which our bias scores are also estimated. Here we present results with more conservatively estimated standard errors, as recommended by some (Cameron & Miller, 2015; for a different perspective, see Abadie, Athey, Imbens, & Wooldridge, 2017). Specifically, we estimate our main logistic regression models with standard errors clustered at the state level to account for potential correlated errors across individuals within the same state. For the SEDA achievement gap analyses, we use the *metareg* command in Stata, which does not allow for the clustering of standard errors. As such, here we present results using traditional OLS regression and standard errors clustered at the state level. We show that our findings are generally robust to these changes.

Table E1

Predictors of County-Level Racial Test Score Inequality (OLS Regression)

	1	2	3	4
Panel A. Implicit Bias				
All	-0.0494**	0.0294		
	(0.0181)	(0.0286)		
Teacher			-0.0577*	0.0867**
			(0.0231)	(0.0269)
Panel B. Explicit Bias				
All	-0.0494**	0.0236		
	(0.0161)	(0.0202)		
Teacher			-0.0577*	0.0734*
			(0.0235)	(0.0319)
Standard Error	County	State	County	State
Sample	Pooled	Pooled	Teachers	Teachers
N Counties	1673	1673	746	746
Covariates	Yes	Yes	Yes	Yes

Note. Traditional OLS regression. Standard errors in parentheses. All models include state fixed effects. Outcome is county's mean standardized White-Black test score difference, pooled across grades and subjects (cohort standardized scale). Bias predictors are county-level empirical Bayes predicted means, standardized to county-level SD of 1 and mean of 0. The Pooled sample consists of all counties used across analyses; the Teacher sample consists of these counties but subset to those that with data used to adjust teacher bias scores based on representativeness. Covariates include: SES composite (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher), W-B difference in SES composite, W-B difference in free lunch %, percent public school students Black, percent public school students Hispanic, per-pupil instructional expenditure, student/teacher ratio, percent public school students attending charter school, W-B difference % charter, W/B ratio of student/teacher ratio, segregation indices. Estimated from a meta-regression performed by methods of moments. $^{\sim}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E2.

Logistic Regression Models Predicting In- and Out-of-School Suspensions

	1	2
Panel A: In-school suspensions on implicit bias		
Bias (all)	0.0213	
	(0.0642)	
Blk*Bias (all)	0.143***	
	(0.0244)	
Bias (tch)		-0.0265
		(0.0571)
Blk*Bias (tch)		0.0910**
		(0.0300)
Panel B: Out-of-school suspensions on implicit bias		
Bias (all)	-0.0549	
	(0.0495)	
Blk*Bias (all)	0.101**	
	(0.0377)	
Bias (tch)		0.149*
		(0.0709)
Blk*Bias (tch)		0.0440
		(0.0449)
Panel D: In-school suspensions on explicit bias		
Bias (all)	0.0206	
	(0.0369)	
Blk*Bias (all)	0.139***	
	(0.0244)	
Bias (tch)		0.0133
		(0.0785)
Blk*Bias (tch)		0.0859**
		(0.0277)
Panel C: Out-of-school suspensions on explicit bias		
Bias (all)	-0.0140	
	(0.0379)	
Blk*Bias (all)	0.110***	
	(0.0323)	
Bias (tch)		0.118*
		(0.0501)
Blk*Bias (tch)		0.0714
		(0.0471)
Sample	Pooled	Teacher
N	90335233	48988187
N counties	1642	738
Covariates	Yes	Yes

Note. Standard errors clustered at the state level in parentheses. All models include state fixed effects. Models fit using aggregate county*race data pooled over the 2011-2012, 2013-2014, and 2015-2016 school years with frequency weights to mimic models with student data pooled across years. Bias predictors are county-level empirical Bayes predicted means standardized to mean=0, SD=1. The Pooled sample consists of all counties used across analyses; the Teacher sample consists of these counties but subset to those that with data used to adjust teacher bias scores based on representativeness. Covariates include: SES composite (composed of log median income, poverty rates, unemployment rates, proportion households receiving SNAP, proportion single-mother-headed households, proportion 25+ with bachelor's degree or higher), W-B difference in SES composite, W-B difference in free lunch %, percent public school students Black, percent public school students Hispanic, per-pupil instructional expenditure, student/teacher ratio, percent public school students attending charter school, W-B difference % charter, W/B ratio of student/teacher ratio, segregation indices. $^{\sim}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix F: Comparing Educators' Implicit Biases to those of Non-educators.

We fit models similar to model 1 in the main text, except that we include educators and non-educators along with a binary indicator for whether the respondent was a K-12 teacher at the time when taking the IAT**.** The only sample restriction imposed on this analysis is requiring non-missingness of predictor and outcome data; as such, the analysis includes additional counties not appearing for the RQ1 analysis given that some counties had non-educators but not educators who met the common sample inclusion criteria.

In Table F1, we compare the implicit racial biases of K-12 educators to those of non-educators. Unadjusted, K-12 educators show nearly identical levels of implicit bias compared to non-educators (column 2). When controlling for individual demographic data, however, teachers show slightly less anti-Black/pro-White bias than non-teachers (by -.008 IAT *d*-scores, column 3). Contextual variables have hardly any effect on teacher/non-teacher bias differences once individual-level demographic variables have been accounted for (column 4).

Our finding of no unadjusted difference in the implicit racial biases of teachers and non-teachers is somewhat surprising given previous research with nationally representative data showing that teachers held more positive or less negative explicit racial attitudes compared to non-educators (Quinn, 2017). Teachers showed only slightly lower levels of implicit bias compared to demographically similar non-teachers, which may indicate that teachers' implicit racial attitudes lag behind their explicit racial attitudes. Alternatively, the contrasting patterns for implicit and explicit racial attitudes may simply indicate that the process of selection into the Project Implicit data set differs for teachers and non-teachers in ways that prevent us from generalizing these results to the broader population.

Table F1.
Multilevel Models Predicting IAT Scores

	1	2	3	4	5
K-12		-0.000291	-0.00756***	-0.0000204	-0.00746**
		(0.00228)	(0.00227)	(0.00228)	(0.00227)
Am. Indian			-0.105***		-0.105***
			(0.00530)		(0.00530)
East Asian			-0.0221***		-0.0222***
			(0.00279)		(0.00279)
South Asian			-0.0802***		-0.0803***
			(0.00326)		(0.00326)
Native Haw.			-0.0995***		-0.0997***
			(0.00572)		(0.00572)
Black			-0.455***		-0.455***
			(0.00139)		(0.00140)
Black+White			-0.230***		-0.230***
			(0.00312)		(0.00312)
Other multi-racial			-0.134***		-0.134***
			(0.00198)		(0.00198)
Race: Other/unknown			-0.122***		-0.122***
			(0.00193)		(0.00194)
Female			-0.0287***		-0.0286***
			(0.000864)		(0.000864)
Age: 30-39			-0.0244***		-0.0244***
			(0.00125)		(0.00125)
Age: 40-49			-0.0297***		-0.0298***
			(0.00153)		(0.00153)
Age: 50-59			-0.0260***		-0.0261***
			(0.00191)		(0.00191)
Age: 60-69			-0.00462		-0.00461
			(0.00297)		(0.00297)
Age: 70+			0.0436***		0.0436***
			(0.00646)		(0.00646)
Educ: HS degree			0.0228***		0.0231***
			(0.00192)		(0.00192)
Educ: Some college			0.0392***		0.0397***
			(0.00143)		(0.00143)
Educ: Bachelors			0.0349***		0.0351***
			(0.00170)		(0.00170)
Educ: Masters			0.0122***		0.0124***
			(0.00188)		(0.00188)
Educ: Advanced deg.			0.0204***		0.0207***
			(0.00220)		(0.00220)
SES Composite				0.00664**	0.00809***
				(0.00221)	(0.00146)
Prop. Black				-0.289***	0.00514

				(0.0102)	(0.00707)
Prop. Hispanic				-0.0457***	0.0212**
				(0.0102)	(0.00697)
Info index FRL/not FRL				-0.0103	0.00975
				(0.0263)	(0.0166)
Info index White/Black				-0.0547	0.00954
				(0.0474)	(0.0295)
Prop. Charter				0.110***	-0.0252
				(0.0286)	(0.0174)
PPE Instruction				0.0133	-0.0189*
				(0.0120)	(0.00913)
Stu/teach ratio				0.000119	0.0000912
				(0.000106)	(0.0000916)
FRL: W-B				-0.156***	0.0142
				(0.0467)	(0.0291)
Prop. Charter: W-B				-0.0851*	0.00359
				(0.0395)	(0.0254)
Stu/Teach: W/B				-0.0554	-0.000734
				(0.0405)	(0.0282)
SES Composite: W-B				0.00123	0.000147
				(0.000744)	(0.000563)
Constant	0.349***	0.349***	0.401***	0.430***	0.407***
	(0.00565)	(0.00565)	(0.00307)	(0.0408)	(0.0284)
ICC County	0.0184	0.0184	0.00279	0.00604	0.00311
ICC State	0.00669	0.00669	0.00167	0.00129	0.00215

Note: All models include random intercepts for counties and states. All models control for year fixed effects. Respondent-level n=991962; county-level n=2362; state-level n=51 (includes DC).
$^{\sim}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References (Appendices)

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* (No. w24003). National Bureau of Economic Research.

Cameron, A.C., & Miller, D.L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50,* 317-372.

Cook, D.O., Kieschnick, R., & McCullough, B.D. (2008). Regression analysis of proportions in finance with self selection. *Journal of Empirical Finance, 15,* 860-867.

Fahle, E. M., Shear, B.R., Kalogrides, D., Reardon, S.F., Chavez, B. & Ho, A.D. (2019). Stanford Education Data Archive technical documentation. Version 3.0. Retrieved from: https://stacks.stanford.edu/file/druid:db586ns4974/SEDA_documentation_v30_09212019.pdf

Papke, L.E., & Wooldridge, J.M. (1996). Econometric methods for fractional response variables with an application to 401(K) plan participation rates. *Journal of Applied Econometrics, 11,* 619-632.

Quinn, D. M. (2017). Racial attitudes of preK–12 and postsecondary educators: Descriptive evidence from nationally representative data. *Educational Researcher*, *46*(7), 397-411.

Reardon, S.F., Kalogrides, D., & Shores, K. (2019b). The geography of racial/ethnic test score gaps. *American Journal of Sociology, 124,* 1164-1221.

Rucker, G., Schwarzer, G., Carpenter, J., & Olkin, I. (2009). Why add nothing to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Statistics in Medicine, 28,* 721-738.