Equal Time for Equal Crime? Racial Bias in School Discipline

Ying Shi
Syracuse University

Maria Zhu
Syracuse University

Well-documented racial disparities in rates of exclusionary discipline may arise from differences in hard-to-observe student behavior or bias, in which treatment for the same behavior varies by student race or ethnicity. We provide evidence for the presence of bias using statewide administrative data that contain rich details on individual disciplinary infractions. Two complementary empirical strategies identify bias in suspension outcomes. The first uses within-incident variation in disciplinary outcomes across White and under-represented minority students. The second employs individual fixed effects to examine how consequences vary for students across incidents based on the race of the other student involved in the incident. Both approaches find that Black students are suspended for longer than Hispanic or White students, while there is no evidence of Hispanic-White disparities. The similarity of findings across approaches and the ability of individual fixed effect models to account for unobserved characteristics common across disciplinary incidents provide support that remaining racial disparities are likely not driven by behavior.

Equal Time for Equal Crime? Racial Bias in School Discipline [*]

Ying Shi

Syracuse University[†]

Maria Zhu

Syracuse University[‡]

Abstract

Well-documented racial disparities in rates of exclusionary discipline may arise from differences in hard-to-observe student behavior or bias, in which treatment for the same behavior varies by student race or ethnicity. We provide evidence for the presence of bias using statewide administrative data that contain rich details on individual disciplinary infractions. Two complementary empirical strategies identify bias in suspension outcomes. The first uses within-incident variation in disciplinary outcomes across White and under-represented minority students. The second employs individual fixed effects to examine how consequences vary for students across incidents based on the race of the other student involved in the incident. Both approaches find that Black students are suspended for longer than Hispanic or White students, while there is no evidence of Hispanic-White disparities. The similarity of findings across approaches and the ability of individual fixed effect models to account for unobserved characteristics common across disciplinary incidents provide support that remaining racial disparities are likely not driven by behavior.

JEL: I24, J15

1   Introduction

The use of exclusionary discipline practices is prevalent in K-12 education in the United States (Heitzeg, 2009; Steinberg, 2016). These practices are increasingly controversial, as recent research document long-lasting negative effects of harsher disciplinary punishments on students' educational attainment and achievement (Bacher-Hicks, Billings, & Deming, 2019; R. Skiba, Arredondo, & Williams, 2014; Sorensen, Bushway, & Gifford, 2021). A related concern is that these policies facilitate the school-to-prison pipeline, a phenomenon in which increasingly harsh school policies expose students to the criminal justice system at a young age (Bacher-Hicks et al., 2019; Heitzeg, 2009; Owens, 2017; Weisburst, 2019). This situation is especially troubling for minority students, who are disproportionately represented in K-12 disciplinary infractions and face pervasive racial disparities upon entering the criminal justice system (Anderson & Ritter, 2017; R. J. Skiba, Michael, Nardo, & Peterson, 2002).[1]

These potentially stark consequences of harsh discipline on later life outcomes motivate efforts to curb the disparate use of exclusionary discipline. Doing so requires a more comprehensive understanding of the origins of such gaps. For instance, one possibility is that disproportionality results from differences in hard-to-observe student characteristics and behavior. Another possibility though is that some of this disproportionality may arise from bias, in which students exhibiting the same behaviors are treated differently by race. While an increasing number of studies show evidence for racial bias in criminal justice settings such as federal sentences (Rehavi & Starr, 2014), bail decisions (Arnold et al., 2018), and criminal trials (Anwar, Bayer, & Hjalmarsson, 2012), less is known about the prevalence of disciplinary-related racial bias during primary and secondary school, before most individuals formally encounter the criminal justice system. This paper provides evidence of racial bias in exclusionary discipline using administrative K-12 data from North Carolina and two complementary identification strategies.

---

[1]These inequalities have been found from police encounters all the way through the system to judge sentencings: Abrams, Bertrand, and Mullainathan (2012); Arnold, Dobbie, and Yang (2018); Fryer (2019); Goncalves and Mello (2021); Grogger and Ridgeway (2006); Horrace and Rohlin (2016); Knowles, Persico, and Todd (2001); Rehavi and Starr (2014).

Our empirical approach examines racial gaps in exclusionary discipline for students of different races who are involved in the same disciplinary incident for the same type of infraction. The first identification strategy leverages within-incident variation in suspension outcomes across students of different races. We show that in disciplinary incidents consisting of one Black student and one white student, Black students are 0.5 percentage points more likely to be suspended, and receive suspensions that average 0.05 days longer. Black students are also suspended 0.04 days longer than Hispanic students in the same incident. In contrast, there are no Hispanic-White gaps in suspension probability or length.

The key identifying assumption underlying this approach is that within an incident, student race is not correlated with hard-to-observe differences in student behavior. We subject our findings to a number of checks, including limiting the analysis to students with no history of office referrals, and find our point estimates are robust to these additional controls. Racial differences in the severity of exclusionary discipline cannot be explained by previous disciplinary history or student characteristics such as relative age. To further address the concern that variation in individual behavior may be driving differential outcomes in exclusionary discipline, even within the same disciplinary incident, we employ a second empirical strategy. This approach uses student-level fixed effects instead of incident-level fixed effects and looks at how disciplinary consequences vary for a given student across incidents based on the race of other students involved in the same incident. Results using this approach show similar patterns to the results using the first strategy: Black students are suspended for longer than their peers when they are involved in an incident with a White or Hispanic student, as opposed to when they are involved in an incident with another Black student. Specifically, Black students who are engaged in the same incident with a White student are suspended for 0.07 additional days, relative to when they are in a same-race incident. The analogous difference for Black-Hispanic incidents is 0.06 days, while we again find no Hispanic-White differences. The proximity of these magnitudes to the earlier set of estimates and ability of individual fixed effects models to account for unobserved student behavior common across disciplinary infractions provide further evidence that measured racial differences in exclusionary difference are not driven by

2

behavior.

To the extent that these racial differences capture discrimination in exclusionary discipline, we might expect them to be correlated with other measures of bias in the academic setting. The presence of subjective classroom teacher assessment in our data allow us to look at the relationship between racial gaps in academic assessment and racial gaps in disciplinary outcomes. We construct a district-level variable that captures teacher assessment bias towards Black students (relative to White students), and relate it to within-incident Black-White differences in suspension length aggregated at the district-level, as described using the first empirical strategy. We find that districts in which teachers systematically under-evaluate Black students in math are associated with relatively longer suspension times for Black students relative to their White peers. These findings provide suggestive evidence that discrimination manifests in multiple school settings, indicating analyses focusing on singular dimensions may overlook the full extent of these effects.

To understand contexts that mitigate or exacerbate racial biases in exclusionary discipline, we examine heterogeneity in effects by the race of school administrators. This analysis is motivated by literature documenting the role of school principals as key players in disciplinary decisions (R. J. Skiba et al., 2014; Sorensen et al., 2021) and research demonstrating academic benefits to exposing under-represented minorities to race-congruent teachers or administrators (Bartanen & Grissom, 2021; Dee, 2005; Gershenson, Hart, Hyman, Lindsay, & Papageorge, 2021; Gershenson, Holt, & Papageorge, 2016). We find suggestive evidence that Black-White and Black-Hispanic gaps in exclusionary discipline are smaller in schools with Black principals, implying that the presence of minority principals may benefit students from under-represented minority groups through disciplinary channels as well.

This paper relates to a growing body of studies examining racial gaps in exclusionary discipline in the education system (Barrett, McEachin, Mills, & Valant, 2019; Kinsler, 2011; R. Skiba et al., 2014). Using North Carolina data as well, Kinsler (2011) finds significant statewide gaps in suspensions between Black and White students, conditional on referral and infraction type (e.g. fighting, truancy, property damage). These effects are largely generated by cross-school variation

3

in punishment. A limitation of this study is the use of earlier data that do not contain incident identifiers, making it difficult to disentangle whether gaps arise from underlying situational and behavioral variances across incidents as opposed to disparate racial standards in treatment. Barrett et al. (2019) use administrative data from Louisiana to look at suspension gaps between Black and White students involved in fights together, finding significant gaps in total days suspended. The authors only observe students who were suspended in the data, and they use date information to infer which students were involved in a given incident. One key advantage of this study is that our data have explicit incident identifiers in our data, which creates a more precise linkage of students to incidents. This information also allows us to expand our analysis to other types of infractions beyond fights, since we do not have to infer incidents by using infractions that necessarily involve multiple students. Additionally, we observe data on student referrals, not just incidents resulting in suspensions. This is an important piece of information, given that we find racial differences in the propensity to receive a suspension, conditional on the incident, a dimension that would be unobserved in suspensions-only data. We are aware of one other paper that uses referral data to examine racial differences in exclusionary discipline. Liu, Hayes, and Gershenson (2021) identify intentional discrimination using rich administrative data from a diverse large urban school district in California. They use an approach akin to our first identification strategy to document greater suspension rates and length for minority relative to White students. The overarching congruence of our results is notable given the significant differences in context in terms of population density and demographic composition. Our paper is unique in implementing a second identification strategy that exploits variation in racial compositions across peers using a student fixed effects approach. This approach further alleviates concerns that there are unobservable behavioral differences across students within incidents that are influencing suspension outcomes.

The availability of incident-level data in a state with substantial populations of under-represented racial and ethnic groups allows us to examine the experiences of Hispanic students in addition to Black students. This is important for external validity, given that Hispanics are the second largest racial or ethnic group behind Whites in the US. In particular, we are able to investigate inter-

racial incidents involving only minority students of different groups. We find substantive gaps in exclusionary discipline outcomes for Black and Hispanic students, even with the inclusion of incident-level fixed effects. Our finding of similar magnitudes of Black-Hispanic gaps in suspension length relative to Black-White gaps is especially interesting. It suggests that the more severe punishment of Black students is unlikely driven by differences in perceived disadvantage or test performance given that academically and socioeconomically, Hispanic students look much more like Black students than they do White students. Our findings of gaps in disciplinary outcomes between Black and Hispanic students shed light on the complexity of relationships among minority racial and ethnic groups and the need for a nuanced and comprehensive approach to understanding these dynamics.

Finally, the focus on school discipline-based bias recalls the growing body of literature studying the role discrimination among police officers and judges plays in contributing to racial disparities in the criminal justice system (Antonovics & Knight, 2009; Anwar & Fang, 2006; Fryer, 2019; Goncalves & Mello, 2021; Grogger & Ridgeway, 2006; Horrace & Rohlin, 2016; Knowles et al., 2001; West, 2018). Challenges to identifying bias in the K-12 context involving the selection of students into schools and unobserved behavior echo the difficulty of pinpointing bias under endogenous police or judge encounters and imperfect data on individual behavior. While some strategies addressing these challenges are not applicable in the K-12 setting, our use of incident fixed effects and juxtaposition across individuals of different races parallels some approaches in the literature on criminal justice and bias (see, for example, West (2018) and use of automobile crash fixed effects).[2]

In the remainder of the paper, Section 2 presents the data used in this study alongside descriptive statistics. Section 3 discusses the empirical strategy used to identify racial biases in exclusionary discipline. Section 4 presents our results and Section 5 concludes.

---

[2]The selection of students into schools and classrooms, and the repeated interactions between students with teachers and school administrators over time distinguishes the K-12 context from settings that may provide more plausible examples of random encounters with law enforcement (e.g. automobile crash investigations or traffic stops under the "veil of darkness" around dusk). Unobserved differences in student behavior in categories such as insubordination are not easily quantifiable, thus it is difficult to identify discrimination at the individual level as done in the context of police officers using driving speed as an objective measure of individual behavior (Goncalves & Mello, 2021).

2   Data and Descriptive Statistics

2.1   North Carolina Education Data

Data for this project come from the North Carolina Education Research Data Center (NCERDC). We observe statewide administrative records on academic and disciplinary information for the universe of elementary and secondary public school students in the state. In this paper, we focus on students in grades K-12 from 2008-2018. The NCERDC data track students across grades and schools over time and contain information on students' background and demographic characteristics, standardized test score performance, and disciplinary records. In the disciplinary records, for each reported offense, we observe information on the type of infraction, individual(s) involved, and the disciplinary consequences each individual received.

These data contain two key advantages for our analysis: First, the disciplinary records contain unique incident identifiers, so we are able to identify the exact individuals involved in an event. Second, we observe the individuals involved in each reported offense regardless of the consequences of referral, which is an advantage over many studies that only observe students in an incident if it resulted in a suspension.[3] The data contain variety of different types of infractions in the data, and we restrict our focus to the most commonly occurring infractions involving multiple people: fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, and disrespect of faculty/staff. The primary disciplinary consequence we focus on is school suspension, which include both in-school and out-of-school suspensions, although we provide supplemental analyses that separate the two as well.

Another unique feature of our data is the presence of subjective teacher evaluations of student performance, in addition to standardized end-of-grade test scores. Concurrently with the standardized tests at the end of the year, teachers are asked to provide their assessment of each student's mastery in math and reading comprehension on a scale of 1 to 4, denoting to insufficient mas-

---

[3]State and federal statutes obligate North Carolina to report particular classes of incidents regardless of consequences. These infraction categories include more severe offenses such as fights, assault, possession of a firearm, and sexual assault. Other commonly occurring but less serious infraction categories are subject to less regulatory oversight.

tery, inconsistent mastery, consistent mastery, and superior performance, respectively.[4] Teachers submit evaluations before knowing standardized test results, and we use this information to assess systematic differences in teacher assessments by student race. After accounting for underlying achievement as measured by standardized test scores, we interpret these racial differences as a measure of teacher bias.

2.2 Descriptive Statistics

Table 1 presents descriptive statistics for the full sample as well as separately by racial and ethnic group. Slightly over half of students in the sample are White, 26% are Black, and 14% are Hispanic. While Black individuals have been the largest minority group in the state, the Hispanic population is growing at a faster rate. Approximately 17% of students in the sample are involved in a disciplinary incident each year, although these percentages vary greatly by race. In a given year, 27% of Black students are involved in a disciplinary incident, compared to 13% of White students and 14% of Hispanic students. Similar patterns appear in suspension outcomes—21% of Black students are suspended in a given year, while only 8% of White students and 9% of Hispanic students are suspended. The gender composition of the sample is approximately 49% female and does not vary much across groups. While 29% of White students categorized as economically disadvantaged, this number is 71% and 75% for Black and Hispanic students, respectively. Finally, White students have significantly higher average academic achievement as measured by lagged standardized math and reading test scores than Black and Hispanic students.

---

[4]A detailed description of each achievement level is as follows:

1. Students performing at this level do not have sufficient mastery of knowledge and skills in this subject area to be successful at the next grade level.

2. Students performing at this level demonstrate inconsistent mastery of knowledge and skills in this subject area and are minimally prepared to be successful at the next grade level.

3. Students performing at this level consistently demonstrate mastery of grade level subject matter and skills and are well prepared for the next grade level.

4. Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level work

Next, Table 3 displays descriptive information on disciplinary incidents. Observations denote student×incident units. We restrict the sample to incidents involving one of the following infractions: fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, and disrespect of faculty/staff. Black students are more likely to be suspended for disciplinary incidents than White students, and their total duration of suspension is also longer. The propensity for a Black student involved in a disciplinary incident to receive a suspension is 0.63, while this number is only 0.56 for White students. Hispanic students fall somewhere in between, with a suspension propensity of 0.59. Similarly, Black students receive an average of 1.68 days of suspension in a disciplinary incident, while this number is 1.29 days for White students and 1.42 days for Hispanic students. When we break out days of suspension into in-school and out-of-school suspension, we see that racial discrepancies in days suspended is driven by differences in out-of-school suspension days. When separating incidents by infraction type, we observe some differences in the distribution of incidents across racial and ethnic groups. For example, 10% of disciplinary incidents observed from Black students arise from fights, while only 7% of incidents observed from White students arise from fights. Conversely, 13% of disciplinary incidents observed from White students arise from bus misbehavior, while this number is only 9% for Black students. The most common infractions we observe in the data is disruptive behavior, followed by insubordination.

3   Empirical Strategy

Our empirical approach aims to identify racial disparities in disciplinary outcomes that arise for reasons distinct from behavioral differences. We advance that conditional on the same behavior, differential disciplinary outcomes reflect bias. This interpretation of bias is inclusive of race and its correlates such as socioeconomic status and test scores. To illustrate, if an economically disadvantaged Black student gets a more severe punishment relative to a wealthy White student after exhibiting the same behavior, this falls under our relatively broad conception of racial bias, which

8

permits the possibility that administrators use race as a proxy for socioeconomic disadvantage and vice versa.[5] While we cannot discern the precise intent of administrators who make disciplinary decisions in our data, we argue that any disparate impact across students of different racial and ethnic groups who otherwise behave the same is problematic due to research documenting the consequences of harsher punishment on student outcomes. Key to our identification strategy, then, is the ability to control for behavioral differences.

We use two complementary identification strategies to account for possible differences in behavior across racial and ethnic groups that may confound attempts to causally identify racial bias in exclusionary discipline. The first approach examines differential outcomes for students of different racial and ethnic groups involved in the *same* disciplinary incident. The specification below illustrates this within-incident approach:

$$Y_{ijgst} = \mathbf{Race}'_{ijgst}\beta + \mathbf{X}'_{ijgst}\Gamma + \delta_j + \theta_{gst} + \varepsilon_{ijgst} \tag{1}$$

where Y_{ijgst} is the outcome of interest for student i involved in disciplinary incident j in grade g, school s, and year t. We focus on two main outcomes of interest: an indicator for whether student i is suspended for incident j, as well as total number of days suspended for incident j (equal to zero if the student is not suspended). The variable $\mathbf{Race}_{ijgst}$ denotes the race of student i, and β is the coefficient of interest, capturing the relationship between student race and disciplinary outcomes.

Key to our analysis is δ_j, a set of disciplinary incident j fixed effects. The inclusion of these fixed effects means that β is identified off of incidents involving multiple students, using within-incident variation in student race. In our preferred specifications, we restrict the sample to incidents involving two different-race individuals. We furthermore include θ_{gst} to capture common shocks at the school-grade-year level that may affect disciplinary outcomes. For incidents involving two students in the same grade, this effect is subsumed by incident fixed effects, which are school- and

---

[5]A related example is differential involvement among parents that correlate with race and ethnicity. If White parents are more likely to contest disciplinary outcomes and their actions result in administrators either reducing the severity of punishment ex-post or preemptively choosing a lighter punishment, then this would also fall under our definition of racial bias.

year-specific by definition.

A central assumption for identifying racial bias in discipline is that student race is not correlated with unobservable differences in behavior. We argue that this is a reasonable assumption, given that our data precisely identifies incidents, and we restrict the sample to incidents in which students are charged with the same type of infraction. One potential concern is that students of certain racial and ethnic groups may have a more extensive history of disciplinary incidents, and this may influence the severity of punishment. We repeat our analyses on a group of students who have no previous infraction record to ensure that any racial differences we find are not driven by disciplinary history. We also look for racial differences by the category of infraction, such as disruptive behavior or fights, to verify that our findings are present across multiple disciplinary contexts.

Our preferred specification relies on a relatively broad interpretation of racial bias that is inclusive of race and its correlates. However, in some instances a racial gap adjusted for select individual attributes may be independently illuminating. The model therefore includes a vector of student covariates, $\mathbf{X}'_{ijgst}$, so we can condition on individual attributes such as socioeconomic status and test scores.

The empirical strategy in Equation 1 focuses on identifying differences in student outcomes for the same incident using an incident fixed-effects approach. Even with the robustness checks mentioned above, one potential concern is that students display underlying behavioral differences within these incidents that are unobserved in the data and correlated with race. To further address these concerns, we also use an alternative empirical strategy that relies on within-student variation in peer race across disciplinary incidents. For a given student, we examine differences in outcomes across incidents when the student is involved in an incident with an other-race peer, compared to a same-race peer. We restrict our sample to incidents involving two individuals and estimate the following:

$$Y_{ijgst} = \pi \mathbf{OtherRace}_{ijgst} + \mathbf{X}'_{ijgst}\Gamma + \alpha_i + \theta_{gst} + \varepsilon_{ijgst} \tag{2}$$

In this approach, Y_{ijgst} denotes the difference in the suspension length student i receives and the

suspension length the peer involved in the same incident receives. The vector of student covariates, $\mathbf{X}_{ijgst}$, can contain information on a student's disciplinary history and sociodemographic attributes, and θ_{gst} is a grade-by-school-by-year fixed effect. While the first empirical strategy focuses on incidents involving two students of different races, this approach uses incidents involving both same-race and other-race peers. The variable **OtherRace**$_{ijgst}$ is an indicator variable that takes on a value of one if the peer involved in the incident with student i is a different race from student i and a value of zero if the peer is the same race.

Crucially, we include student fixed effects (α_i) in the model. This absorbs both observable and unobserved student attributes common across incidents and time that may affect suspension outcomes in Equation 1.[6] For instance, the specification accounts for uniformly aggressive behavior for a given student across disciplinary incidents involving other peers. The coefficient of interest π captures the difference in number of days a student is suspended for an incident with an other-race peer as opposed to when confronting a same-race peer. Focusing on Black students engaged in incidents with Black or White peers, a positive and statistically significant π shows that a given Black student receives longer suspensions when he is involved in an incident with a White peer, as opposed to a Black peer. The inclusion of student fixed effects in this model addresses the concern that Black students who get involved in incidents with White students are negatively selected along unobservable behaviors relative to peers involved in disciplinary incidents.

Taken together, we view the estimation strategies in Equations 1 and 2 to be complementary to each other. We interpret estimation results that are consistent and persistent across both specifications to provide evidence on the state of racial disparities in disciplinary outcomes.

---

[6]One limitation of this approach is that we are only able to identify effects from students involved in multiple disciplinary incidents with students of both same and different races, which drops a significant number of interracial disciplinary incidents from the original sample.

4 Results

4.1 Results from Within-Incident Approach

Table 3 begins by describing raw differences in suspension outcomes by race before adopting the specification in Equation 1. We examine two outcomes: 1) whether a student was suspended following an office referral and 2) the number of days suspended, and present results across three samples: referred incidents in which the offending student is Black or White, Hispanic or White, and Black or Hispanic, respectively. We chose these samples to juxtapose the treatment of Black and Hispanic students in the disciplinary context. Previous work on teacher expectations document similar under-assessment of Black and Hispanic students relative to White counterparts with the same academic performance (Shi & Zhu, 2021). We explore whether analogous Black-White and Hispanic-White gaps exist in exclusionary discipline.

The first column shows that among the universe of referrals from 2008-2018, Black students are on average 7.6 percentage points more likely to be suspended relative to their White peers.[7] Since these unadjusted differences likely reflect a variety of classroom-, school-, or district-level factors ranging from student composition to disciplinary practices, we shift to variation within a school-grade for a given year. When including school-grade-year fixed effects in Column 2, the coefficient on Black for the Black and White sample decreases significantly. Black individuals are only 1.5 percentage points more likely to be suspended relative to White students.

Even though the second specification absorbs the impact of factors such as school-grade level changes to disciplinary practices in a given year, reasons other than discrimination may still attribute to existing differences. For one, different teachers or administrators may handle office referrals involving Black students due to tracking or segregation within a given school and grade, and they may have stricter standards about what constitutes "disruptive" or "aggressive" behavior. Alternatively, it is at least possible that students from certain racial groups have more severe infractions on average, and differences in disciplinary outcomes merely reflect differences in behavior.

---

[7]Analogous estimates for Hispanic students in the Hispanic and White sample is 3.5 percentage points, and 4.0 percentage points for Black students in the Black and Hispanic sample.

To better distinguish between these possibilities, we consider the more tightly controlled setting of interracial incidents.

Column 3 in Table 3 restricts the sample to only incidents involving two individuals of different racial and ethnic groups. The inclusion of school-grade-year fixed effects further absorbs hard-to-observe factors common to students in each cell. The resulting Black-White gap indicates that Black students are 0.5 p.p. more likely to be suspended relative to White students in the sample. Notably, there is no longer an economically or statistically significant Hispanic-White nor Black-Hispanic gap in the probability of suspension. The last column in Table 3 adds pair-specific incident fixed effects such that we identify racial differences using within-pair variation in suspension probability. In this preferred specification, Black students engaged in the same disciplinary incident as their White peers are 0.5 percentage points more likely to be suspended. We document differences in Hispanic-White and Black-Hispanic suspension probabilities that are estimated with relative precision at zero.

The remainder of Table 3 examines racial differences using suspension length as an outcome and the same sequence of specifications. In the full sample of Black and White students observed in the disciplinary dataset, Black students were suspended 0.4 more days on average. This gap shrinks to approximately one-quarter of the original magnitude when including school-grade-year fixed effects, and further attenuates when limiting to pairwise interracial incidents. In the preferred specification in Column 8, Black students are suspended 0.051 more days than their White counterparts. This is equivalent to 3% of the average suspension length of 1.68 days for Black students in the full disciplinary sample. No corresponding difference in suspension length exists between Hispanic and White students engaged in the same disciplinary incident. Notably, there is a Black-Hispanic suspension length gap of 0.038 days, despite no measurable differences in the probability of being suspended.

One potential explanation for these cross-group differences is that Black students in interracial incidents have a disciplinary record that predisposes them to more severe punishment. Even if Black students exhibit the same behaviors as their White or Hispanic peers, administrators may be

more inclined to hand students with past infractions a longer suspension. Since the number and length of previous suspensions may themselves be the product of discriminatory practice, we do not include them as controls in the regression. Instead, Column 2 in Table 4 limits the sample to students who have no history of disciplinary incidents. Sample sizes shrink considerably relative to Column 1, but the magnitudes of coefficients are stable across specifications. The Black-White gap in suspension length is 0.065 days, which is actually larger in magnitude, although not statistically different than the original estimate of 0.051 days. The corresponding Black-Hispanic gap is similar to before, at 0.032 days. However, the loss of statistical power renders the coefficient insignificant.

Table 4 then separately examines racial gaps by suspension type. Out-of-school suspensions comprise approximately three-quarters of overall days suspended. This category also appears to drive the racial and ethnic gaps in suspension outcomes, with significant Black-White and Black-Hispanic gaps of 0.063 and 0.036 additional out-of-school days, respectively. In contrast, Black students only average in-school suspensions that are longer by 0.009 days relative to White peers.

The persistence of Black-White and Black-Hispanic suspension gaps still leaves open the possibility that disciplinary outcomes reflect differences in behavior. Since we do not observe detailed individual actions, one possible remaining concern is that Black students misbehaved more than their White or Hispanic peers involved in the same incident. However, one benefit of our dataset is that we are able to identify multiple types of infractions involving interracial pairs of students. Some categories of infractions, such as fights, may contain more unobserved behavioral differences if we worry for example, that these incidents usually involve one party instigating. However, other categories are likely less susceptible to these concerns. As a robustness check, Table A1 in the appendix shows corresponding racial gaps for several infraction types involving pairs of individuals.[8] We cannot reject that the magnitudes of racial gaps are the same across incidents involving fighting, disruptive behavior, and aggressive behavior, the top three most common categories of infractions.

The racial differences reported thus far do not adjust for individual characteristics. However,

---

[8]This approach departs from Barrett et al. (2019), which uses only fights for identifying racial differences.

doing so may inform an understanding of the extent to which attributes such as socioeconomic status are used as proxies for race and ethnicity. Table 5 explores the role of individual characteristics in explaining observed racial differences in disciplinary outcomes. It begins by controlling for gender, after observing that nearly one-third of Black students in the disciplinary sample are girls compared to only one-fifth for White and Hispanic students. We find little to no difference in racial gaps when including gender in the model, suggesting that suspension patterns do not systematically differ by gender. The next specification accounts for relative age, under the presumption that older students may be perceived as more aggressive or more deserving of blame even when holding constant behavior. We add birth year and month interactions to the model, with resulting coefficients largely unchanged. Finally, Column 4 includes an indicator for economic disadvantage. Doing so reduces the Black-White gap in suspension length from 0.051 to 0.032 days. We view this as consistent with the view that administrators may use race as a proxy for disadvantage, such that the adjusted difference of 0.032 reflects the racial gap apart from factors correlated with disadvantage. Strikingly, the Black-Hispanic gap remains nearly the same as before even after accounting for socioeconomic status, since Hispanic students are similarly disadvantaged. The lingering 0.039 day gap shows that Black students are systemically suspended for longer even when compared to members of another under-represented minority group. Similar findings emerge when controlling for lagged math and reading scores. While their inclusion slightly attenuates the Black-White suspension gap from 0.049 to 0.034 days, the gap between Black and Hispanic students remain nearly unchanged (Table A2).[9]

Finally, we explore heterogeneity in the magnitude of racial differences across school contexts. Specifically, we examine whether racial bias is more muted in schools with minority administrators. This focus is motivated by two strands of literature; one underscores the role of principals and school administrators in influencing the severity of disciplinary outcomes (R. J. Skiba et al., 2014; Sorensen et al., 2021), while the other documents benefits that accrue to under-represented minority students after exposure to a same-race teacher or administrator (Bartanen & Grissom,

---

[9]Note that this is estimated on the sample of students in grades 4-9 for whom we have available grade 3-8 End-of-Grade test scores from the previous year.

2021; Dee, 2005; Gershenson et al., 2021, 2016). We limit the analyses to Black and White principals only, given the dearth of principals of other races and ethnicities in our sample. Slightly over one-quarter of incident-level observations involve students enrolled in a school with a Black principal, while the remainder have White principals. Table A3 shows that across the three samples, the interaction terms between the race indicator and Black principal carry the opposite sign from the race indicator itself, suggesting attenuated racial differences in suspension severity under Black principals. However, the coefficients are not significant using the full sample (Column 2). Columns 3 and 4 expand to a broader group of school administrators and exclude observations with missing race data on assistant principals. The evidence point to significantly smaller Black-White and Black-Hispanic suspension gaps in schools with Black principals, with coefficients robust to further accounting for the share of Black assistant principals. While findings suggest a role for principal race, more work is necessary to understand whether differences are attributable to principal discretion on referral cases or school-level correlates of principal race, such as the presence of alternative disciplinary practices in place of exclusionary discipline.

4.2 Results from Within-Student Approach

An alternate approach to accounting for hard-to-observe discipline-related behavior is to incorporate individual fixed effects. While we cannot fully rule out the possibility that individual behavior might vary depending on the race of the other student, this complementary strategy allows us to net out behavior that is common across incidents for each student. Table 6 shows coefficients for a model that uses only student fixed effects, and another model that also includes school-grade-year fixed effects. We restrict to pairwise incidents in which a Black or Hispanic student faced another student of the same race (e.g. incidents involving both Black students) or a different race (e.g. Black student in the same incident as a White student). The outcome variable is the difference in days suspended between the focal and peer student. The "Other-Race" coefficient, then, captures any gaps in differential suspension lengths between interracial incidents and those involving same-race students.

Our preferred specification in Column 2 produces coefficients that are close in magnitude to the within-incident identification strategy. Black students are suspended for 0.074 more days when engaged in the same incident with a White student, relative to when they are in a same-race incident. There are no analogous Hispanic-White differences, while Black students are suspended for 0.06 days longer when the incident involves a Hispanic student instead of another Black student. To place these magnitudes in context, the additional days suspended for Black students translate to approximately 15-19% of the raw Black-White suspension gap of nearly 0.4 days. The robustness of these findings to the inclusion of student fixed effects provides support that results are not driven by negative selection of Black students on hard-to-observe characteristics into incidents with other-race peers. This suggests that observable factors such as differences in the type of behavioral infraction and school- or district-level disparities in disciplinary practices explain the majority of the gap. Yet even after accounting for observable attributes and unobserved individual characteristics common across incidents, Black students are still penalized more harshly. We interpret these modest yet meaningful residual differences as racial bias.[10]

Another consideration in our interpretation of the magnitude is that we are estimating racial bias in select incidents involving students of different racial and ethnic groups. In these contexts, race is potentially more salient for school administrators making disciplinary decisions. To the extent that they are more cognizant of the potential for bias and careful to demonstrate equitable treatment of all students, we would expect the magnitude of bias to be attenuated in our analytic samples. If we can demonstrate the existence of bias in incidents where they are least likely to be observed, we may expect greater racial bias among incidents involving only students of the same race.

---

[10]Some of the overall racial differences in disciplinary outcomes may be due to Black students sorting into schools with harsher disciplinary practices. This can result in Black students disproportionately bearing the cost of harsher punishment. Our interpretation of racial bias focuses on the differential responses of school administrators and therefore is not inclusive of this form of disparate impact. However, a more expansive definition of racial bias that considers institution- or system-level factors may include these types of examples. In this sense we may be underestimating the magnitude of racial bias.

4.3 District-Level Measures of Racial Differences

The evidence consistently shows Black-White and Black-Hispanic gaps in suspension severity under circumstances involving similar student behavior. If racial bias is present in disciplinary practices, one may expect these biases to arise in other aspects of students' schooling experiences as well. We investigate this possibility by examining another dimension of the school setting, teacher assessments of students, which permits identification of potential bias.

An advantage of North Carolina administrative data is the availability of blind-scored standardized tests in math and reading alongside contemporaneous teacher assessments of student mastery in the same skills. Teachers are asked to assess students on 1-4 achievement scale, with 1 denoting insufficient mastery and 4 denoting superior performance. Teachers do not know students' test performance at the time they make these assessments. These subjective student assessments are used to benchmark achievement level cutoffs at the state level, and are not used as an input into any teacher- or school-based accountability system. We derive a measure of teacher racial bias by juxtaposing the teacher assessments of Black and White students who share the same academic record and belong to the same classroom. Specifically, we estimate the following model:[11]

$$NB_{icsd} = \sum_{d=1}^{D} \rho_d Black_{icsd} + \alpha f(A_{icsd}) + \eta_c + \xi_{icsd} \tag{3}$$

The outcome NB_{icsd} is the teacher's rating of mastery on a scale of 1-4 for a student i in class c and district d, separately for subject $s \in \{Math, Reading\}$. The model controls flexibly for raw test scores $f(A_{icd})$ using a set of year- and subject-specific test score indicator variables so that we compare students with the same academic performance in a given subject. The specification also includes classroom fixed effects η_c, such that ρ_d identifies district-specific racial differences in teacher assessments for Black and White students in the same class who exhibit the same academic mastery. We construe this coefficient as a measure of potential teacher bias.

We estimate an analogous district-specific measure based on racial differences in disciplinary

---

[11]We limit the sample to students in grades 3-8 for the years in which teacher assessment data was available (2008-2013).

outcomes. Equation 4 estimates district-level measures of Black-White suspension gaps using a sample of only pairwise Black-White disciplinary incidents. The outcome variable is total days suspended and the inclusion of incident fixed effects δ_j enables the identification of β_d, our district-specific measure of the racial suspension gap, using within-incident racial variation in outcomes.

$$Y_{ijd} = \sum_{d=1}^{D} \beta_d Black_{ijd} + \delta_j + \varepsilon_{ijd} \tag{4}$$

Table 7 shows the three district-level measures of Black-White differences estimated above. On average, districts penalize Black students by an additional 0.06 days suspended relative to their White peers in the same incident. Districts also rate Black students' math mastery 0.06 points less than their White counterparts in the same classroom on a scale of 1-4, even after controlling for underlying test scores. The analogous Black-White difference in reading assessment is -0.12. Other district-level measures in Table 7 include aggregated response data from the Race Implicit Association Test (IAT) taken from individuals who visited the Project Implicit website.[12] The IAT scale is centered at 0, indicating no bias, while positive numbers describe a pro-White bias. District-level estimates of implicit bias have mean 0.42 with a standard deviation of 0.05. On average, districts are majority White, with 61% of students qualifying for free and reduced lunch.

Next, we explore associations between these alternative measures of bias along multiple dimensions of students' school experiences. Table 8 shows a significant relationship between the suspension-based and math assessment measures, but no accompanying relationship between the suspension-based measure and assessments of reading mastery. For every point in which teachers under-rate Black students relative to White students who share the same math test scores, there is an associated increase in relative suspension length of 0.422 days for Black students who are engaged in the same incident with White students. This correlation is at least as large when we condition on other district-level attributes such as racial and socioeconomic composition. Notably, we do not observe a significant relationship between district-level average IAT scores and our measure of

---

[12]We restrict to respondents who self-identify as White, reside in a North Carolina county, and took the survey between 2008 and 2018. We furthermore limit the IAT sample to the 100 school districts with boundaries that are largely coterminous with the boundaries of the 100 counties in the state.

racial differences in suspension lengths. Overall, these results provide some suggestive evidence that racial biases in discipline may reflect broader patterns of bias that manifest in multiple school settings. This also indicates that a multi-dimensional analysis is important in capturing the full extent of bias that Black students face in the education system.

5   Conclusion

Disproportionality in exclusionary discipline is well-established empirically, but scholarship is still lacking on its origins. Racial gaps in suspension outcomes may arise from differences in behavior or differential treatment of students from different racial groups who exhibit the same behavior. This paper uses uniquely precise statewide administrative data to provide evidence on the existence and magnitude of the latter channel. We use two complementary identification strategies and an alternate measure of bias based on juxtaposition of blind and non-blind teacher assessments of student achievement to show the role of discrimination in the racial suspension gap.

The first identification strategy leverages within-incident variation in suspension outcomes for minority and White students. Black students in the same incident as White peers are 0.5 percentage points more likely to be suspended, and receive suspensions that average 0.05 days longer. Strikingly, this additional penalty for Black students is not only apparent in the context of incidents involving White students. Black students are also suspended 0.04 days longer than Hispanic students in the same incident. In contrast, there are no Hispanic-White gaps in suspension probability or length.

Concerns about variation in individual behavior even within the controlled context of disciplinary incidents lead to a second empirical strategy that examines within-student differences in suspension outcomes. Black students who are engaged in the same incident with a White student are suspended for 0.07 additional days, relative to when they are in a same-race incident. The analogous difference for Black-Hispanic incidents is 0.06 days, while we again find no Hispanic-White differences. The proximity of these magnitudes to the earlier set of estimates and ability of individ-

20

ual fixed effects models to account for unobserved student behavior common across disciplinary infractions provide further evidence that racial differences are not merely capturing behavioral differences.

To the extent that these racial differences capture discrimination in exclusionary discipline, we might expect them to be correlated with other measures of bias in an academic setting. We construct a district-level variable that captures teacher assessment bias towards Black students (relative to White students), and relate it to district-level Black-White gaps in suspension lengths. The finding that districts in which teachers under-evaluate Black students in math also observe increases in the relative suspension length of Black students provides suggestive evidence that racial disparities in disciplinary outcomes may be symptomatic of broader patterns of bias manifesting in multiple school settings.

While we provide evidence on the existence of racial bias, our analyses are agnostic on underlying reasons. We cannot decisively conclude whether the bulk of these unexplained racial differences are driven by taste-based or statistical discrimination, in which race is used to make inferences about individuals in a limited information environment (Arrow, 1973; Becker, 1971; Phelps, 1972). Moreover, we do not discount the interpretation of unintentional, implicit bias in addition to these two well-known theories (Bertrand, Chugh, & Mullainathan, 2005). More work is necessary to establish the conditions under which individuals consciously or unconsciously discriminate in the K-12 setting. Findings can inform the design of interventions aimed at curbing these behaviors.

References

Abrams, D. S., Bertrand, M., & Mullainathan, S. (2012). Do Judges Vary in Their Treatment of Race? *The Journal of Legal Studies*, *41*(2), 347–383.

Anderson, K. P., & Ritter, G. W. (2017, May). Disparate Use of Exclusionary Discipline: Evidence on Inequities in School Discipline from a U.S. State. *Education Policy Analysis Archives*, *25*(49).

Antonovics, K., & Knight, B. G. (2009, February). A New Look at Racial Profiling: Evidence from the Boston Police Department. *The Review of Economics and Statistics*, *91*(1), 163–177.

Anwar, S., Bayer, P., & Hjalmarsson, R. (2012, April). The Impact of Jury Race in Criminal Trials. *The Quarterly Journal of Economics*, *127*(2), 1017–1055.

Anwar, S., & Fang, H. (2006, March). An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence. *American Economic Review*, *96*(1), 127–151.

Arnold, D., Dobbie, W., & Yang, C. S. (2018, November). Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, *133*(4), 1885–1932.

Arrow, K. J. (1973, March). The Theory of Discrimination. In O. Aschenfelter & A. Rees (Eds.), *Discrimination in Labor Markets* (pp. 1–33). Princeton University Press.

Bacher-Hicks, A., Billings, S., & Deming, D. (2019, September). The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime. *NBER Working Paper 26257*.

Barrett, N., McEachin, A., Mills, J. N., & Valant, J. (2019, September). Disparities and Discrimination in Student Discipline by Race and Family Income. *Journal of Human Resources*, 0118–9267R2.

Bartanen, B., & Grissom, J. A. (2021, February). School Principal Race, Teacher Racial Diversity, and Student Achievement. *Journal of Human Resources*, 0218.

Becker, G. S. (1971). *The Economics of Discrimination* (2nd ed.). University of Chicago Press.

Bertrand, M., Chugh, D., & Mullainathan, S. (2005, May). Implicit Discrimination. *American Economic Review*, *95*(2), 94–98.

Dee, T. S. (2005). A Teacher like Me: Does Race, Ethnicity, or Gender Matter? *The American*

Economic Review, *95*(2), 158–165.

Fryer, R. (2019, April). An Empirical Analysis of Racial Differences in Police Use of Force. *Journal of Political Economy*, *127*(3), 1210–1261.

Gershenson, S., Hart, C. M. D., Hyman, J., Lindsay, C., & Papageorge, N. W. (2021, November). *The Long-Run Impacts of Same-Race Teachers* (Tech. Rep. No. w25254). National Bureau of Economic Research.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016, June). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209–224.

Goncalves, F., & Mello, S. (2021). A Few Bad Apples? Racial Bias in Policing. *American Economic Review (Forthcoming)*.

Grogger, J., & Ridgeway, G. (2006, September). Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness. *Journal of the American Statistical Association*, *101*(475), 878–887.

Heitzeg, N. A. (2009). Education or Incarceration: Zero Tolerance Policies and the School to Prison Pipeline. *Forum on Public Policy Online*(2).

Horrace, W. C., & Rohlin, S. M. (2016, May). How Dark Is Dark? Bright Lights, Big City, Racial Profiling. *Review of Economics and Statistics*, *98*(2), 226–232.

Kinsler, J. (2011, December). Understanding the black–white school discipline gap. *Economics of Education Review*, *30*(6), 1370–1383.

Knowles, J., Persico, N., & Todd, P. (2001, February). Racial Bias in Motor Vehicle Searches: Theory and Evidence. *Journal of Political Economy*, *109*(1), 203–229.

Liu, J., Hayes, M. S., & Gershenson, S. (2021). From Referrals to Suspensions: Unpacking the Origins of Exclusionary Discipline. *Working Paper*.

Owens, E. G. (2017, January). Testing the School-to-Prison Pipeline: Testing the School-to-Prison Pipeline. *Journal of Policy Analysis and Management*, *36*(1), 11–37.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic*

Review, *62*(4), 659–661.

Rehavi, M. M., & Starr, S. B. (2014, December). Racial Disparity in Federal Criminal Sentences. *Journal of Political Economy*, *122*(6), 1320–1354.

Shi, Y., & Zhu, M. (2021). Model Minorities in the Classroom? Positive Bias towards Asian Students and its Consequences. *Working Paper*.

Skiba, R., Arredondo, M., & Williams, N. (2014, October). More Than a Metaphor: The Contribution of Exclusionary Discipline to a School-to-Prison Pipeline. *Equity & Excellence in Education*, *47*.

Skiba, R. J., Chung, C.-G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014, August). Parsing Disciplinary Disproportionality: Contributions of Infraction, Student, and School Characteristics to Out-of-School Suspension and Expulsion. *American Educational Research Journal*, *51*(4), 640–670.

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.

Sorensen, L. C., Bushway, S. D., & Gifford, E. J. (2021). Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes. *Education Finance and Policy*, 1–74.

Steinberg, M. P. (2016, October). *What Do We Know About School Discipline Reform?*

Weisburst, E. K. (2019, March). Patrolling Public Schools: The Impact of Funding for School Police on Student Discipline and Long-term Education Outcomes: Patrolling Public Schools. *Journal of Policy Analysis and Management*, *38*(2), 338–365.

West, J. (2018). Racial Bias in Police Investigations. *Working Paper*.

Table 1: Descriptive Statistics: Full Sample

	All (1)	White (2)	Black (3)	Hispanic (4)
Disciplinary Incidents				
Disciplinary Incident Occurred	0.17	0.13	0.27	0.14
Suspended	0.12	0.08	0.21	0.09
Suspended (In-School)	0.07	0.06	0.12	0.06
Suspended (Out-of-School)	0.06	0.04	0.12	0.04
Race				
White	0.52	1.00	0.00	0.00
Black	0.26	0.00	1.00	0.00
Hispanic	0.14	0.00	0.00	1.00
Other	0.08	0.00	0.00	0.00
Characteristics				
Female	0.49	0.48	0.49	0.49
Econonomic Disadvantage	0.48	0.29	0.71	0.75
Reading Z-score $(t-1)$	0.00 (1.00)	0.21 (0.94)	-0.32 (1.04)	-0.21 (0.95)
Math Z-score $(t-1)$	0.00 (1.00)	0.20 (0.94)	-0.34 (1.04)	-0.13 (0.93)
N	16,315,145	8,416,472	4,312,032	2,271,910

Observations denote student×year units for all students in grades K-12, 2008-2018. Variables measuring disciplinary occurrence and suspensions are indicator variables equaling one if the student had a disciplinary incident or was suspended in a given year, respectively. Economic disadvantage variables are only available for grades 3-12, and lagged test scores are available for grades 4-9. We report lagged test scores rather than contemporaneous test scores since contemporaneous scores may be endogenous with disciplinary outcomes.

Table 2: Descriptive Statistics: Disciplinary Incidents

	All (1)	White (2)	Black (3)	Hispanic (4)
Suspension Occurred	0.60	0.56	0.63	0.59
Total Days Suspended	1.52	1.29	1.68	1.42
	(2.16)	(1.94)	(2.29)	(2.05)
Total Days In-School Suspension	0.50	0.51	0.50	0.52
	(1.18)	(1.08)	(1.22)	(1.18)
Total Days Out-of-School Suspension	1.07	0.82	1.25	0.95
	(3.34)	(2.99)	(3.60)	(2.90)
Infraction Type				
Fight	0.09	0.07	0.10	0.09
Disruptive Behavior	0.31	0.30	0.32	0.29
Aggressive Behavior	0.12	0.12	0.12	0.11
Bus Misbehavior	0.11	0.13	0.09	0.12
Inappropriate Language/Disrespect	0.11	0.13	0.10	0.11
Insubordination	0.18	0.17	0.18	0.19
Disrespect of Faculty/Staff	0.08	0.08	0.08	0.08
N	5,088,532	1,624,333	2,677,399	444,031

Observations denote student×incident units, indicating some students may appear in the data multiple times or not at all, depending on how many incidents they were involved in. Sample includes students in grades K-12, 2008-2018. We restrict the sample to incidents involving one of the following infractions: fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, and disrespect of faculty/staff. These represent the most common infractions for which we observe multiple students involved in a given incident. The number of days suspended are censored at 20 for suspensions exceeding 20 days.

Table 3: Racial Differences in Disciplinary Outcomes

	Dependent Var.: Was Suspended				Dependent Var.: Total Days Susp.			
	Full Sample		Interracial Pairs		Full Sample		Interracial Pairs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black/White Sample								
Black	0.076***	0.015***	0.005***	0.005***	0.398***	0.108***	0.055***	0.051***
	(0.008)	(0.001)	(0.001)	(0.001)	(0.028)	(0.005)	(0.008)	(0.007)
N	4233880	4158504	68231	66222	4233880	4158504	68231	66222
Hispanic/White Sample								
Hispanic	0.035***	0.007***	0.001	0.002	0.135***	0.058***	-0.004	0.000
	(0.008)	(0.002)	(0.003)	(0.002)	(0.025)	(0.005)	(0.014)	(0.013)
N	2050579	2002117	19169	18436	2050579	2002117	19169	18436
Black/Hispanic Sample								
Black	0.040***	0.005***	0.001	0.001	0.263***	0.040***	0.048***	0.038***
	(0.007)	(0.002)	(0.002)	(0.002)	(0.023)	(0.006)	(0.012)	(0.011)
N	3054929	2999155	31647	30474	3054929	2999155	31647	30474
School-grade-year FE		Y	Y	Y		Y	Y	Y
Incident FE				Y				Y

*** p<0.01, ** p<0.05, * p<0.1. All samples span grades K-12, 2008-2018. Columns 1, 2, 5, and 6 include all student-incident observations for infraction types described in Table . Columns 3, 4, 7, and 8 restrict the sample to only incidents involving two individuals of different racial and ethnic groups. Standard errors are clustered at the school level.

Table 4: Racial Differences for Students without Incident History and by Suspension Type

	Interracial Pairs			
	(1)	(2)	(3)	(4)
	Total Days Susp.	Total Days Susp.	In-School Susp. Days	Out-of-School Susp. Days
Black/White Sample				
Black	0.051***	0.065***	0.009*	0.063***
	(0.007)	(0.024)	(0.005)	(0.014)
N	66222	4854	66222	66222
Hispanic/White Sample				
Hispanic	0.000	0.007	0.015	-0.027
	(0.013)	(0.035)	(0.014)	(0.039)
N	18436	1866	18436	18436
Black/Hispanic Sample				
Black	0.038***	0.032	0.002	0.036*
	(0.011)	(0.031)	(0.006)	(0.021)
N	30474	2180	30474	30474
No history of disc. incidents		Y		
School-grade-year FE	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y

*** p<0.01, ** p<0.05, * p<0.1. Sample spans grades K-12, 2008-2018. All student-incident observations involve fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, or disrespect of faculty/staff. Standard errors are clustered at the school level.

Table 5: Racial Differences in Days Suspended, Conditional on Student Characteristics

	Interracial Pairs			
	(1)	(2)	(3)	(4)
	Total Days Susp.	Total Days Susp.	Total Days Susp.	Total Days Susp.
Black/White Sample				
Black	0.051***	0.051***	0.053***	0.032***
	(0.007)	(0.007)	(0.007)	(0.008)
N	66222	66222	64824	63002
Hispanic/White Sample				
Hispanic	0.000	-0.000	0.001	-0.000
	(0.013)	(0.013)	(0.013)	(0.015)
N	18436	18436	18052	17652
Black/Hispanic Sample				
Black	0.038***	0.039***	0.042***	0.039***
	(0.011)	(0.011)	(0.011)	(0.011)
N	30474	30474	29978	29262
Student covariates:				
Female		Y		
Birth year and month interactions			Y	
Economic disadvantage				Y
School-grade-year FE	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y

*** p<0.01, ** p<0.05, * p<0.1. Sample spans grades K-12, 2008-2018. All student-incident observations involve fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, or disrespect of faculty/staff. Standard errors are clustered at the school level.

Table 6: Racial Differences in Suspension Outcomes - Student FE Model

	Dependent Variable: Diff. in Days Susp.	
	(1)	(2)
Black Students in Incidents with Another White or Black Student		
Other-Race Student	0.075***	0.074***
	(0.014)	(0.017)
N	176684	165840
Hispanic Students in Incidents with Another White or Hispanic Student		
Other-Race Student	0.011	0.016
	(0.042)	(0.069)
N	11137	7545
Black Students in Incidents with Another Hispanic or Black Student		
Other-Race Student	0.055***	0.060**
	(0.020)	(0.024)
N	160738	150970
Student FE	Y	Y
School-grade-year FE		Y

*** p<0.01, ** p<0.05, * p<0.1. Sample spans grades K-12, 2008-2018. All samples include students who are in pairwise incidents involving another student, in which the other student is either from the same or a different racial or ethnic group. The dependent variable is the difference in suspension length between the focal student and their peer. The coefficient on Other-Race Student therefore captures any differential suspension length when the student is involved in an interracial incident, relative to differences in suspension length when the student is involved in an incident with a same-race peer. Standard errors are clustered at the school level.

Table 7: District-Level Summary Statistics

	Mean	Standard Deviation
	(1)	(2)
Black-White Diff in Days Suspended, Interracial Fights	0.06	0.14
Black-White Diff in Teacher Math Rating (1-4 Scale)	-0.06	0.05
Black-White Diff in Teacher Reading Rating (1-4 Scale)	-0.12	0.08
Average IAT Score	0.42	0.05
Share White	0.53	0.21
Share Black	0.28	0.21
Share Hispanic	0.13	0.08
Share Asian	0.02	0.02
Share American Indian	0.01	0.05
Share Other	0.04	0.02
Share FRL	0.61	0.14
N	102	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. All variables have 102 observations with the exception of IAT scores, which is available for 88 school districts. Restricted to districts with at least 50 Black students. Suspension data uses grades K-12, 2008-2018. Teacher rating data uses grades 3-8, 2008-2013.

Table 8: Relationship between Suspension-Based and Teacher Rating-Based Bias Measures

	Black-White Diff in Days Suspended, Interracial Pairs			
	(1)	(2)	(3)	(4)
Black-White Diff in Tch Math Rating (1-4 Scale)	-0.422**	-0.463**		
	(0.210)	(0.229)		
Black-White Diff in Tch Reading Rating (1-4 Scale)			-0.064	0.011
			(0.170)	(0.183)
Average IAT Score		0.155		0.144
		(0.249)		(0.269)
District Student Enrollment:				
Share Black		0.008		0.028
		(0.086)		(0.082)
Share Hispanic		-0.150		-0.130
		(0.172)		(0.165)
Share Asian		0.352		0.212
		(0.495)		(0.447)
Share American Indian		-0.096		-0.184
		(0.121)		(0.130)
Share Other		-0.100		-0.325
		(0.883)		(0.842)
Share FRL		0.008		-0.059
		(0.101)		(0.098)
N	102	88	102	88

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Units of observation are at the district level. All specifications are weighted by district enrollment. SE clustered at district level. Restricted to districts with at least 50 Black students. Suspension data uses grades K-12, 2008-2018. Teacher rating data uses grades 3-8, 2008-2013.

APPENDIX

Table A1: Racial Differences in Days Suspended - By Infraction Type

	Dependent Variable: Total Days Suspended				
	Full Sample (1)	Fights (2)	Disruptive Behavior (3)	Aggressive Behavior (4)	Other (5)
Black/White Sample					
Black	0.051***	0.068***	0.046***	0.080***	-0.006
	(0.007)	(0.016)	(0.011)	(0.015)	(0.012)
N	66222	22226	17342	11590	11316
Hispanic/White Sample					
Hispanic	0.000	0.006	-0.012	0.001	0.004
	(0.013)	(0.030)	(0.017)	(0.028)	(0.025)
N	18436	6082	4534	3218	3896
Black/Hispanic Sample					
Black	0.038***	0.070***	0.016	0.033	0.014
	(0.011)	(0.023)	(0.017)	(0.023)	(0.015)
N	30474	11776	7192	4880	5248
School-grade-year FE	Y	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y	Y

*** p<0.01, ** p<0.05, * p<0.1. Sample spans grades K-12, 2008-2018 and includes only incidents involving two students of different racial or ethnic groups. Student-incident observations in the Other category include bus misbehavior, inappropriate language/disrespect, insubordination, and disrespect of faculty/staff. Standard errors are clustered at the school level.

Table A2: Racial Differences in Days Suspended, Cond. on Lagged Achievement

	(1) Total Days Susp.	(2) Total Days Susp.
Black/White Sample		
Black	0.049***	0.034***
	(0.009)	(0.009)
N	41210	41210
Hispanic/White Sample		
Hispanic	-0.008	-0.012
	(0.015)	(0.016)
N	12226	12226
Black/Hispanic Sample		
Black	0.033**	0.029**
	(0.013)	(0.013)
N	18926	18926
School-grade-year FE	Y	Y
Incident FE	Y	Y
Lagged math and reading z-scores		Y

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample limited to grades 4-9 with non-missing lagged test score data. The second specification controls for cubics of lagged math and reading achievement. Standard errors are clustered at the school level.

Table A3: Racial Differences by School Administrator Race

	Dep. Variable: Total Days Suspended			
	(1)	(2)	(3)	(4)
Black/White Sample				
Black	0.051***	0.055***	0.053***	0.051***
	(0.008)	(0.009)	(0.009)	(0.011)
Black × Black principal		-0.014	-0.033*	-0.033*
		(0.018)	(0.019)	(0.019)
Black × Share of Black assistant principals				0.010
				(0.022)
N	64760	64760	51592	51592
Hispanic/White Sample				
Hispanic	0.002	0.011	0.002	-0.009
	(0.014)	(0.015)	(0.017)	(0.019)
Hispanic × Black principal		-0.045	-0.057	-0.061
		(0.036)	(0.041)	(0.041)
Hispanic × Share of Black assistant principals				0.052
				(0.047)
N	18090	18090	14546	14546
Black/Hispanic Sample				
Black	0.040***	0.041***	0.046***	0.040*
	(0.012)	(0.014)	(0.016)	(0.021)
Black × Black principal		-0.003	-0.049*	-0.050*
		(0.025)	(0.028)	(0.028)
Black × Share of Black assistant principals				0.016
				(0.033)
N	29700	29700	24312	24312
School-grade-year FE	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y

*** p<0.01, ** p<0.05, * p<0.1. Sample spans grades K-12, 2008-2018 and includes only incidents involving two students of different racial and ethnic groups. All specifications exclude the very small number of observations involving a principal who is not White or Black. The specifications in Columns 3 and 4 exclude observations with missing data on the race of assistant principals. Incident types include fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, and disrespect of faculty/staff. Standard errors are clustered at the school level.