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

Ying Shi
Syracuse University[†]

Maria Zhu
Syracuse University[‡]

Abstract

Racial disparities in school discipline may arise from differences in hard-to-observe student behavior or from bias, in which treatment for the same behavior differs by student race or ethnicity. We provide evidence for the presence of bias using statewide administrative data from the United States. Two complementary strategies identify bias in disciplinary outcomes. The first uses within-incident variation in outcomes by student race. The second employs individual fixed effects to examine how consequences vary across disciplinary incidents based on the race of the peer involved. Both approaches find that Black students receive harsher punishment than Hispanic or White students but show no evidence of Hispanic-White disparities.

JEL: I24, J15

1 Introduction

Disciplinary practices such as suspensions and expulsions are prevalent in K-12 schools in the United States (Heitzeg, 2009; Steinberg, 2016). These practices are controversial, as recent research documents long-lasting negative effects of severe disciplinary punishments on students' educational 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 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 students of color, who are disproportionately represented in K-12 disciplinary infractions and face pervasive racial disparities upon entering the criminal justice system (Anderson & Ritter, 2017; Bekkerman & Gilpin, 2015, 2016; R. J. Skiba, Michael, Nardo, & Peterson, 2002).[1]

These potentially stark consequences of harsh discipline motivate efforts to curb racially disparate use of exclusionary discipline practices such as suspension and expulsion. Doing so requires a more comprehensive understanding of the origins of such gaps. One possibility is that disparities result from differences in hard-to-observe student characteristics and behavior. Another possibility is that some of this gap arises 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 school discipline using administrative K-12 data from North Carolina and two complementary identification strategies.

---

[1]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 approaches examine racial gaps in school discipline for student pairs who are jointly 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 are 0.05 days longer on average. Black students are also suspended 0.04 days longer than Hispanic students in incidents involving a Black and Hispanic student. 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 robustness checks and find our point estimates are robust to these additional controls. Racial differences in the severity of school discipline cannot be explained by previous disciplinary history or student characteristics, such as relative age. To further address the concern that even within a disciplinary incident, variation in individual behavior may be driving differential disciplinary outcomes, we employ a second empirical strategy. This approach uses student-level fixed effects instead of incident-level fixed effects to assess how disciplinary consequences vary for a given student across incidents based on the race of other students involved in the same incident. Results show similar patterns to the results using the first strategy: 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, and we again find no Hispanic-White differences. The proximity of these estimates to those using the first approach, coupled with the ability of individual fixed effects models to account for unobserved student behavior common across disciplinary infractions, provide further evidence that measured racial differences in discipline are not driven by behavioral differences.

To understand contexts that mitigate or exacerbate racial biases in school 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 students of color 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 disciplinary gaps are smaller in schools with Black principals, implying that the presence of same-race principals may benefit students of color through disciplinary channels, in addition to academic channels found in previous studies.

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). A recent paper by Barrett et al. (2019) uses statewide suspension data from Louisiana to study disciplinary gaps between Black and White students involved in the same *inferred* fights together.[2] By using an incident fixed effects approach on a pairwise sample of interracial fights, they are able to show significant racial gaps in total days suspended. We contribute to and advance this literature in several ways. First, we observe data on all student *referrals* instead of a censored dataset containing only suspensions, as in Barrett et al. (2019). This has the advantage of allowing for analyses of racial differences in the likelihood of suspension. In addition to the new dependent variable, we note that selection into the sample of suspended students may depend on factors correlated with both race and severity of disciplinary outcomes. By juxtaposing our results with those of a sample that conditions on suspensions only, we show that failing to account for this selection may bias estimates of racial differences in suspension severity. Our referral sample overcomes this concern by including all students involved in an incident regardless of whether they receive a suspension.

Second, our empirical approach supplements the incident fixed effects model with a student fixed effects model that exploits variation in disciplinary outcomes and racial composition across pairwise disciplinary incidents. This approach alleviates further concerns over unobservable behavioral differences across students within incidents. This in part distinguishes our work from the

---

[2]The authors use date and school information to infer students involved together in the same fight. Our data contains incident identifiers, which create a more precise linkage of students to incidents.

only other paper that to our knowledge uses administrative referral data to examine racial differences in exclusionary discipline. Liu, Hayes, and Gershenson (2021) rely on incident fixed effects models to document greater suspension rates and length for under-represented students relative to White students in a diverse California urban school district.

Another advantage of this study is that we broaden our investigation to other types of infractions beyond fights and to additional racial and ethnic groups such as Hispanic students. This allows us to assess whether racial differences are more pronounced in certain types of infractions. Additionally, we are able to investigate interracial incidents involving only students of color from different groups. Both empirical approaches find substantive gaps in discipline outcomes for Black and Hispanic students. 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 the academic and socioeconomic similarities between Hispanic and Black students.

Finally, the focus on bias in school discipline recalls the growing body of literature that examines how discrimination among police officers and judges can contribute 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).[3]

---

[3]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 (i.e. automobile crash investigations or traffic stops under the "veil of darkness" around dusk).

2 Data and Descriptive Statistics

2.1 North Carolina Education Data

Data for this project comes from the North Carolina Education Research Data Center (NCERDC). We observe statewide administrative records on disciplinary information for all elementary and secondary public school students in the state. In the disciplinary records, each reported offense contains information on the type of infraction, individual(s) involved, and the disciplinary consequences each individual received. Additionally, NCERDC data tracks students across grades and schools over time and provides information on students' socio-demographic characteristics and standardized test score performance. This paper focuses on students in grades K-12 from 2008-2018.

The data bears two key advantages for our analysis: First, disciplinary records contain unique incident identifiers, allowing us to precisely identify 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.[4] 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. These infraction types likely permit more discretion or subjective assessment among the school staff making the referral, relative to offenses such as weapons possession or skipping class. The primary disciplinary consequence we focus on is school suspension, which includes both in-school and out-of-school suspensions, although we also provide supplemental analyses separating the two types.

---

[4]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.

2.2 Descriptive Statistics

Table 1 presents descriptive statistics for both the full sample at the student-year level, as well as a student-incident sample for reported infractions. Slightly over half of students in the full sample are White, 26% are Black, and 14% are Hispanic. Black individuals have historically been the largest non-White racial group in the state, although the Hispanic population is growing at a faster rate. While 17% of students in the sample are involved in a disciplinary incident each year that led to an office referral, these percentages vary greatly by race and ethnicity. In a given year, 27% of Black students received a referral, compared to 13% of White students and 14% of Hispanic students. Similar patterns appear in suspension outcomes—21% of Black students receive a suspension per year, while only 8% of White students and 9% of Hispanic students receive suspensions. Students of color are more likely to be categorized as economically disadvantaged, while White students have significantly higher average lagged academic achievement relative to Black and Hispanic students.

The bottom panel of Table 1 summarizes disciplinary infractions at the student-incident level for the infraction types in our sample. Black students are more likely to be suspended for disciplinary incidents than White students, and their duration of suspension is also longer. 63% of Black students with an office referral eventually receive a suspension, while the corresponding share is only 56% for White students. Hispanic students fall somewhere in between, with a suspension propensity of 59%. Similarly, Black students receive an average of 1.74 days of suspension in a disciplinary incident, while this number is 1.33 days for White students and 1.46 days for Hispanic students. These racial discrepancies in days suspended are driven by differences in out-of-school suspension days.

3 Empirical Strategy

Our empirical approach aims to identify racial disparities in disciplinary outcomes that arise for reasons distinct from behavioral differences. We interpret bias as differential disciplinary outcomes

reflect bias, conditional on the same behavior. This interpretation is inclusive of race and its corre-lates, 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 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, given research documenting negative consequences of harsher punishment on student outcomes.

Crucially for 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 school discipline. First, we use a within-incident approach to examine differential outcomes for students of different racial and ethnic groups involved in the same disciplinary incident:

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

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: 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

---

[5]A related example under our definition of racial bias is 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.

at the school-grade-year level that may affect disciplinary outcomes.

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 dataset precisely identifies incidents, and we compare outcomes to students in the same incident who 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 control for the history of offenses directly and restrict our analyses to a group of students who have no previous infraction record to ensure that racial differences we find are not driven by disciplinary history.

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. We therefore include a vector of student covariates, $\mathbf{X}'_{ijgst}$ in some specifications to allow us to examine the role of attributes such as economic disadvantage on discipline gaps.

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 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:

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

where Y_{ijgst} denotes the difference in the suspension lengths received by student i and their peer involved in the same incident. The vector of student covariates, $\mathbf{X}_{ijgst}$, can contain information on a student's disciplinary history and time-varying sociodemographic attributes. 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.

Student fixed effects, α_i, 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 involved in an incident with a White peer, as opposed to with 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 other-race peers involved in the same disciplinary incidents.

4 Results

4.1 Results from Within-Incident Approach

Table 2 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. Results are presented for three samples of student pairs: incidents in which students are Black and White, Hispanic and White, and Black and Hispanic, respectively.[7]

---

[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 reduces observations significantly from the original sample.

[7]For all estimates, we censor total days suspended at 20 for suspensions exceeding 20 days to ensure results are not being driven by extreme observations. Results are very similar when we estimate the model with uncensored observations.

9

Column 1 shows that Black students are on average 7.6 percentage points more likely to be suspended relative to their White peers. 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. Since these unadjusted differences likely reflect a variety of classroom-, school-, or district-level factors ranging from student composition to disciplinary practices, we add school-grade-year fixed effects in column 2. With this addition, Black individuals are only 1.5 percentage points more likely to be suspended relative to White students. This specification still permits reasons other than discrimination to attribute to existing differences. For one, different teachers or administrators may handle office referrals involving Black students due to tracking within a given school and grade, and they may have stricter standards about what constitutes "disruptive" behavior. Alternatively, it is at least possible that differences in disciplinary outcomes may reflect differences in behavior. We consider the more tightly controlled setting of interracial incidents to better distinguish between these possibilities.

Column 3 in Table 2 restricts the sample to only incidents involving two students of different racial or ethnic groups and includes school-grade-year fixed effects. The resulting Black-White gap indicates that Black students are 0.5 percentage points more likely to be suspended relative to White students in the sample. Column 4 in Table 2 adds pair-specific incident fixed effects such that we identify racial differences using within-pair variation in suspension probability.[8] 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, relative to a sample suspension probability of 69 percent. Notably, there is no longer a statistically significant Hispanic-White nor Black-Hispanic gap in the probability of suspension.

The remainder of Table 2 examines racial differences using suspension length as an outcome and the same sequence of specifications. Black students were suspended 0.4 more days on average than White students. This gap shrinks when including school-grade-year fixed effects and limiting to pairwise interracial incidents. In the preferred specification in column 8, Black students are

---

[8]Note that the number of observations decrease slightly relative to Column 3 due to dropped singletons.

suspended 0.051 more days than their White counterparts. This is equivalent to 2 percent of the average suspension length of 2.10 days for students in this pairwise sample of incidents involving Black and White students. 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.[9]

We place our main results in the context of previous literature by emphasizing several differences from Barrett et al. (2019). The availability of referral data allows us to examine racial differences in suspension probability, which was not feasible in their administrative data. Conditioning on only suspensions can also lead to biased estimates of racial differences in suspension length. The first two columns of Table A4 display the point estimates from above, juxtaposed with those using a suspensions-only sample. While the interaction term in Column 3 shows no statistically significant differences between the two samples, the positive coefficients for the Black-White and Black-Hispanic samples suggest that the absence of referral data may lead to overestimates of racial differences. The remaining specifications use the alternative dependent variable of whether a student was suspended for longer. We find that restricting only to suspensions yields an estimate for the Black-Hispanic sample that is statistically larger than the corresponding estimate using referral data.

Next, we probe potential explanations for the more severe punishment of Black students in interracial incidents. One is that they may have been referred or suspended more frequently in the past. Even if Black students exhibit the same behaviors as their White or Hispanic peers, differences in disciplinary history may lead to Black students receiving more severe punishment based on administrator discretion or as mandated by district or school disciplinary policies. Table 3 examines the sensitivity of our main findings to controls for disciplinary history. Column 1 replicates the racial differences estimated in Table 2, and column 2 controls for the number of

---

[9]One potential concern is that for incidents with large suspension gaps for the students involved, the assumption that underlying behavior is the same may be less plausible. We provide evidence that our results are robust to the exclusion of these incidents in section A.1 of the appendix.

cumulative referrals at the school up to the present disciplinary incident. The Black-White gap in suspension length becomes 0.041 days, while the Black-Hispanic gap is 0.028 days. Neither magnitude is statistically different from the original estimates in Table 2. Column 3 limits the sample to students who have no history of disciplinary incidents in the school. This specification may be preferable if Black students act out in response to perceived discrimination from prior incidents, such that subsequent referrals and suspensions are endogenously determined. Sample sizes shrink considerably, but the magnitudes of both coefficients are very similar to before and significantly different from zero. The Black-White gap in suspension length is now 0.058 days, and the corresponding Black-Hispanic gap is 0.041 days.

The final two columns of Table 3 examine racial gaps separately for in- and out-of-school suspensions.[10] 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 racial differences reported thus far do not adjust for individual characteristics. 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 4 explores the role of individual characteristics in explaining observed racial differences in disciplinary outcomes, with a focus on gender, age, economic disadvantage, special education, and limited English proficiency. The inclusion of these covariates reduces the Black-White gap in suspension length from 0.051 to 0.034 days. We view this as consistent with administrators using race as a proxy for disadvantage, such that the adjusted difference reflects the racial gap apart from correlates of disadvantage including lower family income or having an emotional or physical disability. Strikingly, the Black-Hispanic gap remains exactly the same as before even after accounting for individual characteristics, since Hispanic stu-

---

[10]In-school suspensions are usually short-term suspensions served in a suspension classroom, while out-of-school suspensions are served at a non-school location. While both types must be reported under state and federal statutes, local variation in disciplinary policy leads to differences in implementation at the district and school levels.

dents are similarly disadvantaged. The lingering 0.038 day gap shows that Black students are systemically suspended for longer even when compared to members of another under-represented group. Similar findings emerge when controlling for lagged math and reading scores. The inclusion of both student characteristics and lagged scores attenuates the Black-White suspension gap in the full sample. In contrast, the suspension gap between Black and Hispanic students remains nearly unchanged in a fully saturated model (Table A5).[11]

Finally, we explore heterogeneity in the magnitude of racial differences across disciplinary and school contexts. First, we examine whether the magnitude of racial gaps in school discipline vary across incident types. In particular, in section A.2 of the appendix, we explore whether effects differ across more subjective infractions (i.e. "disrespect"), compared to more objective infractions (i.e. "skipping class"). Findings indicate racial gaps in discipline are driven by subjective infractions. These results suggest that situations in which more individual discretion in judgment is involved may invite more bias. These findings motivate the next set of analyses on the characteristics of individuals making disciplinary decisions.

Specifically, we next examine whether racial bias is more muted in schools with administrators from under-represented groups. We focus on principals over other school staff, given their outsized role in shaping schools' disciplinary climate and influencing the severity of disciplinary outcomes (R. J. Skiba et al., 2014; Sorensen et al., 2021).[12] The exploration of same-race congruence is informed by literature documenting benefits that accrue to students of color after exposure to a same-race teacher or administrator (Bartanen & Grissom, 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 A6 shows that across the three samples, the interaction terms between the

---

[11]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.

[12]We examine principals over teachers because the former determine disciplinary outcomes and many referrals may not involve teachers. For instance, only half of documented North Carolina disciplinary incidents in 2011 and 2012 took place in the classroom.

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.[13] While findings suggest a role for principal race, more work is necessary to determine 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 alternative approach to accounting for hard-to-observe discipline-related behavior is to incorporate individual fixed effects. This complementary strategy allows us to net out behavior that is common across incidents for each student. Table 5 shows coefficients corresponding to a student fixed effects model. We restrict to pairwise incidents in which a Black or Hispanic student faced another student of the same race (i.e. incidents involving both Black students) or a different race (i.e. Black student in the same incident as a White student). The outcome variable is the difference in days suspended between the focal student and their peer. 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 also incorporates school-grade-year fixed effects and produces racial differences 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

---

[13]Corresponding analyses using whether suspended as an outcome instead of suspension length suggest that Black students are less likely to be suspended in Black-Hispanic incidents when the principal is Black. We find no significant different among the Black-White sample.

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 percent of the raw Black-White suspension gap of nearly 0.4 days. Notably, the estimates are nearly unchanged when we include time-varying student characteristics such as economic disadvantage, special education, and limited English proficiency (column 3), or the history of referrals (column 4). 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.

Taken together, our findings suggest that observable factors such as 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. For the purposes of interpretation, we clarify that we may be underestimating the magnitude of racial bias. Some of the overall racial differences in disciplinary outcomes may be due to Black students sorting into schools with stricter 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 at the institution- or system-level.

Another reason we may be underestimating the magnitude of racial bias is that we are focused on 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 a lower bound, with more scope for bias in incidents involving only students of the same race.

15

5 Conclusion

Disparities in school discipline are well-established empirically, but scholarship is still lacking on the role of bias in inducing these gaps. This paper provides new evidence on the existence and magnitude of racial bias in school discipline by using a unique statewide administrative dataset containing student referrals across incident categories. Two identification strategies show similar and complementary findings on the role of racial bias in school discipline.

We rely on a rich dataset with information on both referrals and suspensions to show that Black students are 0.5 percentage points more likely to be suspended than White peers in the same incident. Furthermore, Black students receive suspensions that average 0.05 and 0.04 days longer than White and Hispanic incident peers, respectively. In contrast, there are no Hispanic-White gaps in suspension probability or length.

The second empirical strategy using student fixed effects finds that Black students who are engaged in the same incident with a White student are suspended for over 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 similarity of these results to the within-incident estimates, coupled with the ability of individual fixed effects models to account for unobserved student behavior common across disciplinary infractions, provide further evidence that racial disparities are not merely capturing behavioral differences.

While we provide evidence on the existence of racial bias, our analyses are agnostic on underlying reasons. We stop short of concluding whether these unexplained racial differences are driven by taste-based or statistical discrimination (Arrow, 1973; Becker, 1971; Phelps, 1972). Notably, 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 to inform policies aimed at curbing these behaviors. One promising finding in our research is that the presence of Black administrators may mitigate Black-White suspension gaps. Future research on the nature of these administrator-student interactions and the link between a diversified

school workforce and longer-run student outcomes can inform equity-minded policies that address disproportionality in school discipline.

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.

Bekkerman, A., & Gilpin, G. (2015). On understanding inconsistent disciplinary behavior in schools. *Applied Economics Letters*, *22*(10), 772-776.

Bekkerman, A., & Gilpin, G. (2016). Crime and punishment: The role of student body character-

istics in schools' disciplinary behaviours. *Applied Economics*, *48*(15), 772-776.

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.

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

	All (1)	White (2)	Black (3)	Hispanic (4)
Panel A: Full Sample				
Disciplinary Incidents				
$\geqslant$ 1 Incident Referral	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
Student Characteristics				
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
Female	0.49	0.48	0.49	0.49
Econonmic Disadvantage	0.48	0.29	0.71	0.75
Reading Z-score $(t-1)$	0.00	0.30	-0.45	-0.36
	(1.00)	(0.93)	(0.92)	(0.94)
Math Z-score $(t-1)$	0.00	0.29	-0.49	-0.24
	(1.00)	(0.95)	(0.89)	(0.91)
N	16,315,145	8,416,472	4,312,032	2,271,910
Panel B: Disciplinary Sample				
Suspension Occurred	0.60	0.56	0.63	0.59
Total Days Suspended	1.57	1.33	1.74	1.46
	(3.49)	(3.16)	(3.75)	(3.04)
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)
N	5,088,532	1,624,333	2,677,399	444,031

Observations in Panel A denote student×year units for all students in grades K-12, 2008-2018. Variables measuring disciplinary occurrence 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 potentially endogenous contemporaneous test scores. Observations in Panel B 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. We restrict the sample to the following infractions: fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, and disrespect of faculty/staff.

Table 2: 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	4,233,880	4,158,504	68,231	66,222	4,233,880	415,8504	68,231	66,222
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	2,050,579	2,002,117	19,169	18,436	205,579	2,002,117	19,169	18,436
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	2,999,155	31,647	30,474	3,054,929	2,999,155	31,647	30,474
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 fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, and disrespect of faculty/staff. 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 3: Racial Differences for Students by Incident History and Suspension Type

| | Interracial Pairs | | | Days by Suspension Type | |
| | (1) | (2) | (3) | (4) | (5) |
	Total Days Susp.	Total Days Susp.	Total Days Susp.	In-School	Out-of-School
Black/White Sample					
Black	0.051***	0.041***	0.058***	0.009*	0.063***
	(0.007)	(0.007)	(0.016)	(0.005)	(0.014)
N	66,222	66,222	9,820	66,222	66,222
Hispanic/White Sample					
Hispanic	0.000	0.002	0.010	0.015	-0.027
	(0.013)	(0.013)	(0.026)	(0.014)	(0.039)
N	18,436	18,436	3,486	18,436	18,436
Black/Hispanic Sample					
Black	0.038***	0.028**	0.041*	0.002	0.036*
	(0.011)	(0.012)	(0.021)	(0.006)	(0.021)
N	30,474	30,474	4,706	30,474	30,474
School-grade-year FE	Y	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y	Y
Previous referrals		Y			
No previous referrals			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. Previous referrals is the total number of referrals a student accumulates at the school before a given disciplinary incident, entered linearly. Standard errors are clustered at the school level.

Table 4: Racial Differences in Days Suspended,
Conditional on Student Characteristics

	Interracial Pairs	
	(1)	(2)
	Total Days Susp.	Total Days Susp.
Black/White Sample		
Black	0.051***	0.034***
	(0.007)	(0.008)
N	66,222	62,458
Hispanic/White Sample		
Hispanic	0.000	0.014
	(0.013)	(0.019)
N	18,436	16,054
Black/Hispanic Sample		
Black	0.038***	0.038***
	(0.011)	(0.014)
N	30,474	26,732
School-grade-year FE	Y	Y
Incident FE	Y	Y
Student characteristics		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. Student characteristics include gender, and indicators for birth year and month, economic disadvantage, special education, and limited English proficient. Sample size differences due to missing data on student covariates. Standard errors are clustered at the school level.

Table 5: Racial Differences in Suspension Outcomes -
Student FE Model

	Dependent Variable: Diff. in Days Susp.		
	(1)	(2)	(3)
Black Students in Incidents with White or Black Student			
Other-Race Student	0.074***	0.079***	0.078***
	(0.017)	(0.017)	(0.018)
N	165,840	164,043	158,512
Hispanic Students in Incidents with White or Hispanic Student			
Other-Race Student	0.016	0.022	0.000
	(0.069)	(0.069)	(0.077)
N	7,545	7,448	6,360
Black Students in Incidents with Hispanic or Black Student			
Other-Race Student	0.060**	0.060**	0.060**
	(0.024)	(0.023)	(0.024)
N	150,970	149,249	144,075
Student FE	Y	Y	Y
School-grade-year FE	Y	Y	Y
Student characteristics		Y	
Previous referrals			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. Student characteristics include gender, and indicators for birth year and month, economic disadvantage, special education, and limited English proficient. Sample size differences due to missing data on student covariates. Standard errors are clustered at the school level.

A   APPENDIX

A.1   Robustness to Removal of Large Suspension Gaps

A key assumption for identifying racial bias in this paper is that our estimates are not conflating racial bias with unobserved differences in behavior across racial groups. One potential concern is that this assumption may be less plausible for incidents with large suspension gaps for the students involved. In our setting, over 95 percent of incidents involving two students have a suspension gap of three or fewer days. Furthermore, over 98 percent of incidents have a suspension gap of five or fewer days. As a robustness check, Table A1 re-estimates our preferred specification on samples that drop incidents with large suspension gaps.[14] Results in Table A1 show that even conditioning on incidents in which the gap in days suspended between the two students is no more than three or five days, we still find sizable and significant gaps in the probability of receiving a suspension and total days suspended between Black and White students involved in the same incident. Similarly, we also still find significant gaps in total days suspended between Black and Hispanic students. These results provide further support of evidence of racial bias in school discipline decisions, independent of underlying behavioral differences between students.

---

[14]Note that we do not expect estimates in the restricted to be the identical to original estimates because we are altering our sample in a way that mechanically restricts estimates of suspension gaps. However, if we still find significant and sizable suspension gaps with these restrictions, this provides support that our findings are not driven by incidents that are more prone to having differences in underlying behavior.

Table A1: Racial Differences in Disciplinary Outcomes: Robustness to Removal of Large Suspension Gaps

	Dependent Var.: Was Suspended			Dependent Var.: Total Days Susp.		
	Full Sample	⩽5 Days Susp. Gap	⩽3 Days Susp. Gap	Full Sample	⩽5 Days Susp. Gap	⩽3 Days Susp. Gap
	(1)	(2)	(3)	(4)	(5)	(6)
Black/White Sample						
Black	0.005***	0.004***	0.004***	0.051***	0.039***	0.030***
	(0.001)	(0.001)	(0.001)	(0.007)	(0.005)	(0.004)
N	66,222	65,456	64,242	66,222	65,456	64,242
Hispanic/White Sample						
Hispanic	0.002	0.002	0.002	0.000	0.005	0.006
	(0.002)	(0.002)	(0.002)	(0.013)	(0.009)	(0.007)
N	18,436	18,252	17,940	18,436	18,252	17,940
Hispanic/Black Sample						
Black	0.001	0.001	0.001	0.038***	0.026***	0.025***
	(0.002)	(0.002)	(0.002)	(0.011)	(0.008)	(0.006)
N	30,474	30,082	29,488	30,474	30,082	29,488
School-grade-year FE	Y	Y	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y	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 fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, and disrespect of faculty/staff. Columns 1 and 4 report results using the full sample from our preferred specification. Columns 2 and 5 restrict the sample to incidents in which the suspension gap between the two students is ⩽5 days. Columns 3 and 6 restrict the sample to incidents in which the suspension gap between the two students is ⩽3 days. Standard errors are clustered at the school level.

A.2 Heterogeneity by Infraction Type

Table A3 assesses whether effects differ across different types of infractions. In particular, we classify infractions as subjective or objective infractions. In our main analysis, we focus on the most commonly occurring infractions involving multiple individuals. For this analysis, we expand to the full set of infractions. Table A2 shows a detailed breakdown of infraction categorizations. We broadly define objective infractions as infractions that can be concretely quantified or classified, such as "skipping class", while subjective infractions are those that are less easily certified, such as "disrespect." We acknowledge that under this definition there is still ambiguity in how some infractions should be categorized. However, we feel that overall, the set of infractions classified as objective compared to those classified as objective provides a good representation of their respective categories.

Table A2: Categorization of Infractions: Objective vs. Subjective

Subjective	Disruptive behavior, insubordination, aggressive behavior, inappropriate language/disrespect, bus misbehavior, fighting, disrespect of faculty/staff, other school-defined offenses, disorderly conduct, bullying, communicating threats, assault on student, misuse of school technology, harassment (verbal), excessive display of affection, falsification of information, repeat offender, affray, assault on school personnel not resulting in injury, gang activity, assault (other), assault on student without weapon not resulting in injury, violent assault not resulting in serious injury, hazing, assault on non-student without weapon not resulting in injury, sexual assault not involving rape or sexual offense, discrimination, assault resulting in serious injury
Objective	Excessive tardiness, late to class, skipping class, cell phone use, skipping school, theft, dress code violation, leaving school without permission, leaving class without permission, truancy, being in an unauthorized area, use of tobacco, inappropriate items on school property, property damage, honor code violation, possession of tobacco, mutual sexual contact between students, possession of marijuana, possession of weapon (excluding firearms/explosives), possession of a firearm or powerful explosive, cutting class, possession of chemical or drug paraphernalia, use of controlled substances, possession of controlled substance (other), alcohol possession, immunization, use of alcoholic beverages, possession of counterfeit items, false fire alarm, possession, sale of marijuana, unlawfully setting fire, possession of student's own prescription drug, gambling, use of narcotics, physical exam, sale of controlled substance (other), bomb threat, assault involving use of a weapon, extortion, use of counterfeit items

Table A3 displays the results of our heterogeneity analysis. Findings indicate racial gaps in

discipline are driven by subjective infractions: For objective infractions, the coefficient estimate for Black students is 0.1 percentage points and not statistically significant for the outcome of whether a student is suspended. Similarly, the coefficient estimate on total days suspended is 0.012 for Black students and also statistically insignificant. In contrast, results indicate Black students are 0.5 percentage points more likely to be suspended than White peers involved in the same incident and suspended for 0.050 days longer on average in subjective infractions.

Table A3: Heterogeneity in Racial Differences in Disciplinary Outcomes: Objective vs. Subjective Infractions

	Dependent Var.: Was Suspended			Dependent Var.: Total Days Susp.		
	Objective	Subjective	Full Sample	Objective	Subjective	Full Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Black/White Sample						
Black	0.001	0.005***	0.001	0.012	0.050***	0.012
	(0.003)	(0.001)	(0.003)	(0.015)	(0.007)	(0.015)
Black×Subjective			0.004			0.038**
			(0.003)			(0.017)
N	13,962	75,658	89,620	13,962	75,658	89,620
Hispanic/White Sample						
Hispanic	0.002	0.001	0.002	-0.018	-0.003	-0.018
	(0.004)	(0.002)	(0.004)	(0.025)	(0.012)	(0.025)
Hispanic×Subjective			-0.000			0.015
			(0.005)			(0.028)
N	7,524	22,536	30,060	7,524	22,536	30,060
Black/Hispanic Sample						
Black	-0.002	0.001	-0.002	0.023	0.032***	0.023
	(0.005)	(0.002)	(0.005)	(0.027)	(0.010)	(0.027)
Black×Subjective			0.003			0.009
			(0.005)			(0.028)
N	6,534	34,482	41,016	6,534	34,482	41,016
School-grade-year FE	Y	Y	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y	Y	Y

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample restricted to interracial pairs. All samples span grades K-12, 2008-2018. Table A2 provides a breakdown of objective and subjective classifications by infraction. Columns (1) and (4) use only the sample of objective infractions, columns (2) and (5) use only the sample of subjective infravtions, and columns (3) and (6) use the full sample. Standard errors are clustered at the school level.

A.3 Additional Tables

Table A4: Comparison with Barrett et al. (2019) Sample

	Dependent Var.: Total Days Susp.			Dependent Var.: Suspended for Longer		
	Preferred Sample: Referrals, Incl. Suspensions	Barrett et al.: Suspensions Only	Interacted:	Preferred Sample: & Referrals, Incl. Suspensions	Barrett et al.: Suspensions Only	Interacted:
	(1)	(2)	(3)	(4)	(5)	(6)
Black/White Sample						
Black	0.051***	0.057***	0.039***	0.020***	0.022***	0.015***
	(0.007)	(0.009)	(0.012)	(0.003)	(0.003)	(0.004)
Black×Barrett et al.			0.017			0.008
			(0.014)			(0.005)
N	66222	44618	66222	66222	44618	66222
Hispanic/White Sample						
Hispanic	0.000	-0.007	0.010	0.004	0.003	0.003
	(0.013)	(0.017)	(0.021)	(0.004)	(0.006)	(0.007)
Hispanic×Barrett et al.			-0.015			0.001
			(0.028)			(0.010)
N	18436	12142	18436	18436	12142	18436
Black/Hispanic Sample						
Black	0.038***	0.046***	0.019	0.014***	0.018***	0.004
	(0.011)	(0.013)	(0.019)	(0.004)	(0.005)	(0.007)
Black×Barrett et al.			0.027			0.014*
			(0.024)			(0.008)
N	30474	21490	30474	30474	21490	30474
School-grade-year FE	Y	Y	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y	Y	Y

*** p<0.01, ** p<0.05, * p<0.1. All samples are restricted to only incidents involving two individuals of different racial and ethnic groups. 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.

Table A5: Racial Differences in Days Suspended, Cond. on Lagged Achievement

	(1) Total Days Susp.	(2) Total Days Susp.	(3) Total Days Susp.
Black/White Sample			
Black	0.051***	0.018*	0.058**
	(0.009)	(0.010)	(0.023)
N	39,570	39,570	6,574
Hispanic/White Sample			
Hispanic	-0.002	-0.002	0.041
	(0.017)	(0.021)	(0.049)
N	10,816	10,816	2,114
Black/Hispanic Sample			
Black	0.035**	0.033**	0.030
	(0.017)	(0.017)	(0.035)
N	16,622	16,622	2,714
School-grade-year FE	Y	Y	Y
Incident FE	Y	Y	Y
Student characteristics		Y	Y
Lagged math and reading z-scores		Y	Y
No previous referrals			Y

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample limited to grades 4-9 with non-missing lagged test score data. Student characteristics include gender, and indicators for birth year and month, economic disadvantage, special education, and limited English proficient. The second and third specifications also control for cubics of lagged math and reading achievement. Standard errors are clustered at the school level.

Table A6: 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.