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 unobservable student behavior or from bias, in which treatment for the same behavior varies by student race or ethnicity. We provide evidence for the presence of bias in school discipline decisions using statewide administrative data that contain rich details on individual disciplinary infractions. Two complementary empirical strategies identify racial differences suggestive of bias in suspension outcomes. The first uses within-incident variation in disciplinary outcomes across White, Black, and Hispanic 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 face higher suspension probabilities and longer suspensions than White students, and are suspended for longer than Hispanic students. 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 unlikely to be driven by differences in behavior.

JEL: I24, J15

*We are grateful to Bob Bifulco, Jennifer Doleac, Jing Liu, Amy Schwartz, and seminar participants at AEFP for their feedback and contributions to this paper.

†Shi: Department of Public Administration and International Affairs, 426 Eggers Hall, Syracuse, NY 13244-1020 (Email: yshi78@syr.edu; Phone: 315-443-9442)

‡Zhu: Department of Economics, 110 Eggers Hall, Syracuse, NY 13244-1020 (Email: mzhu33@syr.edu; Phone: 315-443-9043)

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 controversial, as recent research document 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; R. J. Skiba, Michael, Nardo, & Peterson, 2002).[1]

These potentially stark consequences of harsh discipline on later life outcomes motivate efforts to curb racially disparate use of exclusionary discipline. Doing so requires a more comprehensive understanding of the origins of such gaps. One possibility is that disparities result from differences in unobservable student characteristics and behavior. Another possibility is that some of this disproportionality arises from bias, in which students exhibiting the same behaviors are treated differently by race. While an increasing number of studies provide 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 uses administrative K-12 data from North Carolina and two complementary identification strategies to document findings that are consistent with racial bias driving some of the gaps in exclusionary discipline.

[1] Inequalities have been found from police encounters all the way through the criminal justice 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 exclusionary discipline for students of different races who are jointly involved in the same disciplinary incident. 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 0.4 percentage points more likely to be suspended, which corresponds to 6% of the overall gap in the propensity for White and Black students to receive a suspension, conditional on being reported for a disciplinary infraction. Furthermore, Black students receive suspensions that are 0.05 days longer on average than White peers involved in the same incident, and this magnitude is 12% of the overall gap between Black and White students in number of days suspended for an incident. Black students are also suspended 0.03 days longer than Hispanic students in the same incident, which is 8% of the gap between Black and Hispanic students in suspension length. 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 unobservable differences in student behavior. We subject our findings to multiple robustness 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 even within a disciplinary incident, variation in individual behavior may be driving differential outcomes in exclusionary discipline, 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 0.5 percentage points more likely to be suspended and have suspensions averaging 0.07 additional days, relative to when they are in a same-race incident. The analogous difference for Black-Hispanic incidents is 0.03 additional days. The proximity of these magnitudes to estimates 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 exclusionary are not driven primarily by behavioral differences.

To understand contexts that mitigate or exacerbate racial gaps in exclusionary discipline, we examine heterogeneity by infraction type. School administrators have more discretion over punishment severity for some types of incidents based on locally administered student codes of conduct. We abide by these district guidelines to classify incidents into objective vs. subjective infractions, with the former capturing misbehavior such as skipping school and dress code violations, and the latter encompassing categories such as insubordination and disruptive behavior. Findings indicate that racial differences are almost entirely driven by subjective infractions. Put differently, contexts involving more school administrator discretion may invite more bias.

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 receiving an office referral and infraction type. A limitation of this study is that earlier data 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 and school information to infer students involved together in the same fight.

We contribute to and advance the 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). Selection into the sample of suspended students may depend on factors correlated with both race and severity of disciplinary outcomes. Failing to account for this selection may bias estimates of racial differences. Our referral sample overcomes this concern by including all students involved

in an incident regardless of whether they receive a suspension. 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. The paper uses an approach akin to our first identification strategy to document greater suspension rates and length for under-represented students relative to White students. 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 influence suspension outcomes.

Another advantage of this study is that our data contains incident identifiers, which create a more precise linkage of students to incidents relative to Barrett et al. (2019). This enables us to expand our analysis to other types of infractions beyond fights, so that we can assess whether racial differences are more pronounced in infraction types involving more or less discretion on the part of school administrators. The scope of North Carolina data furthermore allows us to examine the experiences of Hispanic students in addition to Black and White students. In particular, we are able to investigate interracial incidents involving only students of color from 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 solely 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 under-represented racial and ethnic groups.

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 approaches in the literature on criminal justice and bias (see, for example, West (2018) and use of automobile crash fixed effects).[2]

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 disciplinary information for all elementary and secondary public school students in the state. In the disciplinary records, we observe information on the type of infraction, individual(s) involved, and the disciplinary consequences everyone received for each reported offense. NCERDC data track students across grades and schools over time and contain information on students' socio-demographic characteristics and standardized test score performance in addition to disciplinary records. In this paper, we focus on students in grades K-12 from 2008-2018.

These data contain two key advantages for our analysis: First, disciplinary records contain unique incident identifiers, allowing us to identify the exact individuals involved in an event as well

[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).

as distinguish between incidents with one or multiple students. 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] We use the universe of infraction types in the data. The most commonly occurring infractions involving more than one student in this setting are fights, disruptive behavior, aggressive behavior, bus misbehavior, inappropriate language/disrespect, insubordination, and disrespect of faculty/staff. The disciplinary consequences we focus on are i) whether a student receives a school suspension and ii) total suspension length.

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. 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. 17% of students in the sample are involved in a disciplinary incident each year that led to an office referral, although 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 in a year, compared to only 8% of White students and 9% of Hispanic students. Poverty is more concentrated among students of color, with 29% of White students categorized as economically disadvantaged compared to 71% and 75% for Black and Hispanic students, respectively. Finally, 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, across all incident types. Black students appear more likely to be suspended for disciplinary inci-

[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.

dents than White students, and their duration of suspension is also longer. 62% of Black students with an office referral eventually receive a suspension, while the corresponding share is only 53% for White students. Hispanic students fall somewhere in between, with a suspension propensity of 57%. Similarly, Black students receive an average of 1.71 days of suspension in a disciplinary incident, while this statistic is 1.31 days for White students and 1.43 days for Hispanic students.

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 students displaying 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 permits the possibility that some of the disparity by race can be explained by differences in socioeconomic status.[4] While we cannot discern the 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. 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. First, the within-incident approach examines differential outcomes for students of various racial and ethnic groups involved in the same disciplinary incident:

[4] A related example is that differential involvement among parents may 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.

$$Y_{ijst} = \mathbf{Race}'_{ijst}\beta + \mathbf{X}'_{ijst}\Gamma + \delta_j + \varepsilon_{ijst} \quad (1)$$

where Y_{ijst} is the outcome of interest for student i involved in disciplinary incident j in 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 $\mathbf{Race}_{ijst}$ 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. Since incidents occur within school and year, these fixed effects also subsume any shocks at the school-year level that may affect disciplinary outcomes. 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 specification, we restrict the sample to incidents involving two different-race individuals.

A central assumption for interpreting β, which captures disparate outcomes across race, as evidence of 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 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 infractions directly and restrict our analyses to a group of students who have no previous infraction record to ensure that any 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. The model therefore includes a vector of student covariates, $\mathbf{X}'_{ijgst}$, so we can condition on individual attributes such as whether a student is economically disadvantaged or has a disability.

Even with the robustness checks mentioned above, one potential concern is that students dis-

play underlying behavioral differences within 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:

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

In this approach, Y_{ijst} denotes the difference in suspension probability or length received by student i and their peer involved in the same incident. The vector of student covariates, $\mathbf{X}_{ijst}$, can contain information on a student's disciplinary history and 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 $\mathbf{OtherRace}_{ijst}$ 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.[5] 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 the 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

[5]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.

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.

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. While we cannot definitively rule out unobserved factors correlated with race as contributing to these disparities, we perform several analyses in the next section to assess the claim that these disparities in large part reflect bias.

4 Results

4.1 Racial Differences Using Within-Incident Approach

Table 2 describes raw racial differences in suspension outcomes before adopting the specification in Equation 1. We examine two outcomes: i) whether a student was suspended following an office referral and ii) the number of days suspended. Results are presented for three samples: Black and White, Hispanic and White, and Black and Hispanic students, respectively.

We begin with the full sample of incidents involving one or two students. Column 1 shows that Black students are on average 7.5 percentage points more likely to be suspended relative to their White peers. Analogous estimates are 3.7 percentage points for Hispanic students in the Hispanic and White sample, and 3.7 percentage points for Black students in the Black and Hispanic sample. Since these unadjusted differences likely reflect a variety of school-, district-, and time-specific factors ranging from student composition to disciplinary practices, we add school-year fixed effects in column 2. Even with these additions, 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 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 reflect differences in behavior. To better distinguish between

these possibilities, we consider the more tightly controlled setting of interracial incidents.

Column 3 in Table 2 restricts the sample to only incidents involving two students of different racial or ethnic groups, and includes pair-specific incident fixed effects such that racial differences are identified using within-pair variation in suspension probability. Black students are 0.4 p.p. more likely to be suspended relative to White students using this specification. This is equivalent to 6% of the overall gap in the propensity for White students and Black students to receive a suspension, conditional on being reported for an incident, as shown in Table 1. Differences in Hispanic-White and Black-Hispanic suspension probabilities are estimated with relative precision near zero.

The remainder of Table 2 examines racial differences using suspension length as an outcome and the same sequence of specifications. Black students are suspended 0.3 more days on average than White students in the full sample. This gap shrinks with the inclusion of school-year fixed effects and limiting to pairwise interracial incidents with incident fixed effects. In this specification, Black students are suspended 0.047 more days than their White counterparts.[6] This is equivalent to 12% of the overall gap in the propensity for White students and Black students in total days suspended for an incident. No corresponding difference in suspension length exists between Hispanic and White students engaged in the same disciplinary incident. Notably, there exists a significant Black-Hispanic suspension length gap, with Black students receiving a suspension of 0.030 days longer on average, despite no measurable differences in the probability of being suspended. This magnitude is 8% of the overall gap in the propensity for Hispanic students and Black students in total days suspended for an incident.

The preferred specification using within-incident variation in suspension outcomes trades off the more controlled context of a shared disciplinary incident with a significantly smaller sample. Table A1 examines whether this pairwise sample differs from the broader set of incidents in meaningful ways. One observation from this table is that incidents involving two individuals occur

[6]For context on the economic significance of this gap, Bacher-Hicks et al. (2019) find that students assigned to middle schools that have an average of 0.38 more days of suspensions are 15% more likely to drop out of school, 11% less likely to attend a 4-year college, and 20% more likely to ever be incarcerated as an adult. Our estimates indicate the disparity in number of days suspended between Black and White students is about 12% of 0.38 days.

much less frequently than single-person incidents. For example, the number of two-person incidents with both a Black and White student is approximately 2% of all infractions involving one Black or one White student. Notably, there are consistently larger unadjusted racial differences across both outcomes among single-person infractions, compared to the pairwise sample.[7] This evidence suggests that our main results may in fact be underestimating the magnitude of racial differences. For incidents involving students of multiple races, 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 expect the magnitude of bias to be a lower bound, with more scope for bias in single- or multi-person incidents involving only students of the same race.

Even within incidents, one concern is that the finding that Black students are punished more severely than White or Hispanic peers in the same incident may be consistent with explanations other than bias, namely that Black students have longer disciplinary records. Administrators are often obligated by school or district policies to hand students with past infractions a longer suspension. Tables 3 and 4 use the racial differences in disciplinary outcomes estimated in Table 2 as a starting point before accounting for the role of referral history. We next adjust for student characteristics such as gender, age, economic disadvantage, special education status, and limited English proficiency to assess the role of these factors in explaining measured disciplinary gaps between students of different races.

We account for student disciplinary history in several ways. Column 2 in Table 3 controls for the number of referrals a student has accumulated at the school up to a given disciplinary incident, while column 3 distinguishes between the number of referrals in the academic year and those that came before. Black students are 0.3 and 0.2 p.p. more likely to be suspended than White peers, compared to the original coefficient of 0.4 p.p. Next, we include a vector of individual covariates, which does not affect the coefficient but sufficiently decreases precision so that the Black-White

[7] In contrast, we do not observe large and systematic differences across pairwise and single-person incidents in student gender, socioeconomic status, and whether the student has a disability or limited English proficiency. One exception is that there are more male Black students in incidents involving two students.

difference in suspension probability is no longer significant. The final specification further restricts the sample to only students who have never been referred for any disciplinary incidents at the school. While restrictive, this avoids including an input that may itself be the subject of racial bias. The much-reduced sample size, however, inflates the standard error and we are unable to reject the null of no difference across Black and White students. There are no significant differences for the Hispanic-White and Black-Hispanic samples, as before.

Table 4 shifts from suspension probability as an outcome to examining the number of days suspended. Using samples with non-missing referral history and student covariate data, we document a Black-White gap in suspension length of 0.046 days, and a Black-Hispanic gap of 0.031. The finding that Black students are disciplined more severely across these two samples persists after accounting for past referrals using the total number of referrals and separately by cumulative referrals during current year and earlier. While somewhat attenuated, the coefficient is statistically the same as before. The final specification, which restricts the sample to never-referred students and controls for student covariates, shows Black-White and Black-Hispanic suspension gaps of 0.040 and 0.023 days, respectively.

Notably, the inclusion of only student covariates into the model in column 4 of Table 4 reduces the Black-White gap from 0.046 to 0.033 days, but the analogous Black-Hispanic disparity increases from 0.031 to 0.037 days. The latter reflects the comparable prevalence of disadvantage as measured by limited English proficiency, family income, and special education status between Hispanic and Black students. While these factors cannot explain why Black students are suspended more frequently when engaged in the same incident with members of another under-represented group, they shed some light on the Black-White disparity. Table A2 shows the coefficients on all student covariates corresponding to columns 4 and 5. The coefficient on economic disadvantage is large and highly significant for the Black-White sample for the outcome of whether suspended. The coefficient is similarly significant across all three samples for the outcome of days suspended, suggesting that lower socioeconomic status predicts greater suspension severity. This variable is positively correlated with the race indicator for Black, as White students are much less likely to be

identified as economically disadvantaged, in contrast to the comparable rates among Black and Hispanic students (Table 1). Strikingly, the contribution of economic disadvantage disappears when we restrict to students with no previous referrals, suggesting that students with clean disciplinary records are more comparable in socioeconomic status.[8][9]

4.2 Racial Differences Using Within-Student Approach

The within-incident empirical strategy further alleviates the possibility of variation in unobserved behavior that is correlated with race. To better account for these potential unobserved differences, we employ a complementary approach using individual fixed effects. This strategy allows us to control for unobserved behavior that is common across incidents for each student. Tables 5 and 6 show corresponding racial disparities in suspension probability and length.

Table 5 restricts to pairwise incidents in which a Black or Hispanic student are involved with 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 binary outcome variable takes on a value of 1 if the student was suspended, but their peer in the same incident is not. The "Other-Race" coefficient, then, captures differences in suspension probability between interracial incidents and those involving same-race students.

The first specification incorporates both student and school-year fixed effects. It documents racial differences that are close in magnitude to the within-incident identification strategy. Black students are 0.5 p.p. more likely to be suspended when engaged in an incident with a White peer, compared to an incident with another Black student. There is suggestive evidence that Hispanic students are more likely to be suspended than White students, in contrast to the first identification strategy, while there are no analogous Black-Hispanic disparities in the likelihood of suspension.

[8]Table A3 includes lagged math and reading scores in addition to student characteristics among the sample of never-referred students. Column 3 shows no differences in suspension probability across the three samples. The suspension gap between Black and White students is 0.042 in this fully saturated model, while the corresponding Black-Hispanic gap attenuates to 0.013 and is no longer significant (column 6).

[9]To get a sense for how differences in unobservable characteristics may be driving racial gaps in disciplinary outcomes within incidents, we also implemented a bounding exercise put forth by Oster (2019). Our finding that there exist significant racial differences in disciplinary outcomes are robust to estimated bounds.

Notably, estimates are nearly unchanged when we include time-varying attributes such as the cumulative number of previous referrals and student characteristics including economic disadvantage, special education status, and limited English proficiency. The robustness of these findings to the inclusion of student fixed effects suggests that results are not driven by negative selection of Black students on unobservable characteristics into incidents with other-race peers.

Table 6 repeats the exercise for suspension length, with the new outcome variable as the difference in days suspended between the focal and peer student. Black students are suspended for 0.069 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 0.032 days longer when the incident involves a Hispanic student relative to when the incident involves another Black student. To place these magnitudes in context, the additional days suspended for Black students translate to approximately one-sixth of the raw Black-White suspension gap of nearly 0.4 days. Again, coefficients are almost entirely stable to the addition of time-varying covariates.[10]

Overall, we find even after accounting for observable attributes and unobserved individual characteristics common across incidents, Black students are still penalized more harshly relative to White and Hispanic peers. We interpret these modest yet economically meaningful disparities as consistent with the presence of some racial bias. As mentioned previously, our findings may underestimate the true magnitude because our identification rests on interracial incidents, which exhibit smaller raw racial differences than single-person incidents. Another reason for underestimating the scope of bias is that 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 is not inclusive of this form of disparate impact at the institution- or system-level.

[10]As a robustness check, we have also run specifications augmenting specifications in Tables 5 and 6 with time-varying covariates on *peers'* economic disadvantage, students with disabilities, limited English proficiency, previous referrals, and lagged test scores in math and reading. Findings using these specifications are similar to our main results.

4.3 Heterogeneity by Incident Type

Finally, we explore heterogeneity in the magnitude of racial differences across different infraction types. The extent of punishment severity (including recommended minimum and maximum levels) is specified for infraction categories in Codes of Student Conduct at the school district level in North Carolina. In addition, state legislative guidelines permit principals and local boards some discretion over the administration of exclusionary discipline.[11] As a result, the range of possible suspension outcomes is broader for some incident types compared to others. We examine whether the magnitude of racial gaps in exclusionary discipline vary for infractions in which the determination of disciplinary outcomes is more subjective compared to those in which the determination of outcomes is more objective. Subjectivity in this context is defined by the extent of administrator discretion in the determination of disciplinary outcomes. For our analysis, we consulted Codes of Student Conduct for multiple districts to help guide the classification of infractions. In particular, we use the guidelines for the appropriate range of disciplinary severity for each infraction, as defined by minimum and maximum disciplinary response levels. Objective infractions are infractions with a narrower range of recommended disciplinary outcomes, while subjective infractions are those that come with wider ranges. We are able to classify 85% of observations in the sample as objective vs. subjective using these guidelines. Table A4 in the appendix provides a breakdown of the specific infractions included in each category.

Table 7 assesses whether racial differences vary across subjective and objective infractions. Findings indicate racial gaps in discipline are driven by subjective infractions: For objective infractions involving one Black and one White student, column (1) shows 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 0.06 for Black students and also statistically insignificant. In contrast, results in column (2) indicate Black students are 0.5 percentage points more likely to be suspended than White peers involved in the same incident and suspended for 0.055 days longer on average in subjective infractions. These results

[11] Principal and school board discretion is codified in Article 27 of the North Carolina General Statutes.

suggest that situations in which more individual discretion in judgment is involved may invite more bias.

Tables A5 and A6 in the appendix analyze results separately for objective and subjective incidents, respectively. These tables include racial gaps using both incidents restricted to interracial pairs, as well as the full sample of incidents. In Table A5, results indicate no significant racial difference in either suspension outcome for the full sample or among interracial pairs for objective incidents. These findings provide further support to our identifying assumption that unobservable characteristics correlated with race are not driving results. In contrast to the objective incidents, there are significant differences in disciplinary outcomes by race for both the full sample and interracial pair sample for subjective incidents, as shown in Table A6.

5 Conclusion

Disparities in exclusionary discipline are well-established empirically, but scholarship is still lacking on their origins. This paper uses uniquely rich statewide administrative data to provide evidence of racial bias in school discipline decisions. We use two complementary identification strategies to show the role of discrimination in the racial suspension gap.

The first identification strategy leverages within-incident variation in suspension outcomes for students from different racial groups. Black students in the same incident as White peers are 0.4 percentage points more likely to be suspended, and receive suspensions that average 0.05 days longer. Black students are also suspended 0.03 days longer than Hispanic students in the same incident. 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 0.03 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 racial disparities are not merely capturing behavioral differences.

While we cannot definitively rule out the possibility that measured racial differences in disciplinary outcomes are influenced by factors other than race, we believe that our complementary identification strategies provide compelling evidence of the presence of racial bias in disciplinary decisions in our setting. Even within incidents and controlling for prior disciplinary history and student characteristics, there is a possibility that there exist unobserved components of behavior correlated with race that influence suspension outcomes. To the extent that these differences are driven by fixed differences across students, our student fixed effects approach should adequately control for these factors. However, it is possible that there may be time-varying factors driving unobserved student behaviors that are correlated with the race of peers as well as disciplinary outcomes.

While we provide strong support 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, in which race is used to make inferences about individuals in a limited information environment (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 notable finding is that infraction categories permitting more school administrator discretion may drive racial disparities in suspension outcomes. Future research on the nature of these administrator-student interactions and the influence of local student codes of conduct on the distribution of disciplinary outcomes can inform 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.

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.

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

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*.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187-204.

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., 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
Was Suspended	0.12	0.08	0.21	0.09
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
Econonomic 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				
Was Suspended	0.58	0.53	0.62	0.57
	(0.49)	(0.50)	(0.49)	(0.49)
Total Days Suspended	1.52	1.31	1.71	1.43
	(2.30)	(2.14)	(2.43)	(2.22)
N	8,216,993	2,979,092	3,915,644	792,645

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. Total days suspended are censored at 20 for suspensions exceeding 20 days.

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)
Black/White Sample						
Black	0.075***	0.014***	0.004***	0.328***	0.060***	0.047***
	(0.004)	(0.001)	(0.001)	(0.017)	(0.003)	(0.006)
N	5,856,343	5,855,751	98,466	5,856,343	5,855,751	98,466
Baseline	0.582	0.582	0.663	1.493	1.493	1.990
Hispanic/White Sample						
Hispanic	0.037***	0.009***	0.001	0.095***	0.036***	-0.006
	(0.004)	(0.001)	(0.002)	(0.016)	(0.004)	(0.010)
N	3,165,767	3,164,710	34,736	3,165,767	3,164,710	34,736
Baseline	0.545	0.545	0.634	1.312	1.312	1.789
Black/Hispanic Sample						
Black	0.037***	0.001	-0.000	0.224***	0.013***	0.030***
	(0.004)	(0.001)	(0.002)	(0.013)	(0.004)	(0.010)
N	3,956,716	3,955,557	46,734	3,956,716	3,955,557	46,734
Baseline	0.608	0.608	0.696	1.596	1.596	2.153
School-year FE		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, and include student-incident observations across all infraction types. The full sample includes incidents involving up to two individuals belonging to the specified racial and ethnic groups, and excludes incidents with three or more students. Columns 3 and 6 restrict the sample to only incidents involving two individuals of specified racial and ethnic groups. Standard errors are clustered at the school level.

Table 3: Racial Differences in Suspension Rates, Adjusted for Student Covariates

	Dependent Variable: Was Suspended				
	(1)	(2)	(3)	(4)	(5)
Black/White Sample					
Black	0.004***	0.003***	0.002**	0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
N	92,368	92,368	92,368	92,368	13,264
Baseline=0.663					
Hispanic/White Sample					
Hispanic	0.001	0.001	0.002	-0.001	-0.005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
N	30,150	30,150	30,150	30,150	5,356
Baseline=0.634					
Black/Hispanic Sample					
Black	-0.001	-0.001	-0.002	-0.001	0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
N	40,842	40,842	40,842	40,842	5,934
Baseline=0.696					
School-year FE	Y	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y	Y
Previous referrals		Y			
Previous referrals this year			Y		
Previous referrals before this year			Y		
Student characteristics				Y	Y
No previous referrals					Y

*** p<0.01, ** p<0.05, * p<0.1. Sample includes student-incident observations in grades K-12 during 2008-2018 with non-missing data on referral history and student characteristics. Previous referrals is the total number of referrals a student accumulates at the school before a given disciplinary incident, entered linearly. Previous referrals this year is the cumulative number of referrals in a given academic year at the school, while previous referrals before this year are all remaining referrals that took place before the current academic year. Student characteristics include gender, and indicators for birth year and month, economic disadvantage, students with disabilities, and limited English proficiency. Standard errors are clustered at the school level.

Table 4: Racial Differences in Suspension Length, Adjusted for Student Covariates

	Dependent Variable: Total Days Suspended				
	(1)	(2)	(3)	(4)	(5)
Black/White Sample					
Black	0.046***	0.037***	0.033***	0.033***	0.040***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.015)
N	92,368	92,368	92,368	92,368	13,264
Baseline=1.990					
Hispanic/White Sample					
Hispanic	-0.001	-0.001	0.000	-0.008	-0.014
	(0.011)	(0.011)	(0.011)	(0.013)	(0.026)
N	30,150	30,150	30,150	30,150	5,354
Baseline=1.789					
Black/Hispanic Sample					
Black	0.031***	0.024**	0.019*	0.037***	0.023
	(0.010)	(0.010)	(0.010)	(0.012)	(0.024)
N	40,842	40,842	40,842	40,842	5,934
Baseline=2.153					
School-year FE	Y	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y	Y
Previous referrals		Y			
Previous referrals this year			Y		
Previous referrals before this year			Y		
Student characteristics				Y	Y
No previous referrals					Y

*** p<0.01, ** p<0.05, * p<0.1. Sample includes student-incident observations in grades K-12 during 2008-2018 with non-missing data on referral history and student characteristics. Previous referrals is the total number of referrals a student accumulates at the school before a given disciplinary incident, entered linearly. Previous referrals this year is the cumulative number of referrals in a given academic year at the school, while previous referrals before this year are all remaining referrals that took place before the current academic year. Student characteristics include gender, and indicators for birth year and month, economic disadvantage, students with disabilities, and limited English proficiency. Standard errors are clustered at the school level.

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

	Dep. Var.: Student Suspended, But Not Peer			
	(1)	(2)	(3)	(4)
Black/White Sample: Black-White or Black-Black Incident				
Black Student in Interracial Incident	0.005***	0.005***	0.005***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)
N	233,170	233,169	233,170	233,170
Baseline=0.0263				
Hispanic/White Sample: Hispanic-White or Hispanic-Hispanic Incident				
Hispanic Student in Interracial Incident	0.011*	0.011	0.011*	0.011*
	(0.007)	(0.007)	(0.007)	(0.007)
N	19,345	19,345	19,345	19,345
Baseline=0.0258				
Black/Hispanic Sample: Black-Hispanic or Black-Black Incident				
Black Student in Interracial Incident	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
N	211,546	211,545	211,546	211,546
Baseline=0.0257				
Student FE	Y	Y	Y	Y
School-year FE	Y	Y	Y	Y
Previous referrals		Y		
Previous referrals this year			Y	
Previous referrals before this year			Y	
Student characteristics				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 takes on a value of 1 if the focal student is suspended while the peer in the same incident is not. Coefficients therefore capture differences in the probability of suspension when the student is involved in an interracial incident, relative to differences in suspension outcomes when the student is involved in an incident with a same-race peer. Student characteristics include economic disadvantage status, special education status, and limited English proficiency status. Standard errors are clustered at the school level.

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

	Dependent Variable: Diff. in Days Susp.			
	(1)	(2)	(3)	(4)
Black/White Sample: Black-White or Black-Black Incident				
Black Student in Interracial Incident	0.069***	0.069***	0.069***	0.069***
	(0.013)	(0.013)	(0.013)	(0.013)
N	233,170	233,169	233,170	233,170
Baseline=0.0029				
Hispanic/White Sample: Hispanic-White or Hispanic-Hispanic Incident				
Hispanic Student in Interracial Incident	-0.036	-0.036	-0.036	-0.035
	(0.036)	(0.036)	(0.036)	(0.036)
N	19,345	19,345	19,345	19,345
Baseline=-0.0029				
Black/Hispanic Sample: Black-Hispanic or Black-Black Incident				
Black Student in Interracial Incident	0.032*	0.033*	0.032*	0.032*
	(0.018)	(0.018)	(0.018)	(0.018)
N	211,546	211,545	211,546	211,546
Baseline=-0.0017				
Student FE	Y	Y	Y	Y
School-year FE	Y	Y	Y	Y
Previous referrals		Y		
Previous referrals this year			Y	
Previous referrals before this year			Y	
Student characteristics				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. Coefficients therefore capture any differential suspension length when the student is involved in an interracial incident, relative to differences in suspension outcomes when the student is involved in an incident with a same-race peer. Student characteristics include economic disadvantage status, special education status, and limited English proficiency status. Standard errors are clustered at the school level.

Table 7: Heterogeneity in Racial Differences: 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	0.006	0.055***	0.006
	(0.003)	(0.001)	(0.003)	(0.011)	(0.007)	(0.011)
Black × Subjective			0.004			0.049***
			(0.003)			(0.013)
N	13,102	75,432	88,534	13,102	75,432	88,534
Baseline	0.518	0.728	0.663	1.188	2.206	1.540
Hispanic/White Sample						
Hispanic	0.001	0.003	0.001	-0.014	0.004	-0.014
	(0.004)	(0.002)	(0.004)	(0.014)	(0.013)	(0.014)
Hispanic × Subjective			0.002			0.017
			(0.004)			(0.019)
N	7,312	23,336	3,0648	7,312	23,336	30,648
Baseline	0.305	0.656	0.483	0.582	1.815	1.276
Black/Hispanic Sample						
Black	-0.004	0.001	-0.004	-0.009	0.040***	-0.009
	(0.005)	(0.002)	(0.005)	(0.017)	(0.011)	(0.017)
Black × Subjective			0.005			0.049**
			(0.005)			(0.020)
N	6,342	36,126	42,468	6,342	36,126	42,468
Baseline	0.588	0.752	0.696	1.208	2.401	2.153
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 A4 provides a breakdown of objective and subjective classifications by infraction type. Columns 1 and 4 use only the sample of objective infractions, columns 2 and 5 use only the sample of subjective infractions, and columns 3 and 6 use the full sample. Standard errors are clustered at the school level.

APPENDIX

Table A1: Unadjusted Racial Differences in Disciplinary Outcomes

	Was Suspended		Total Days Susp.		Female		Econ. Disadv.		SWD		LEP	
	N=1	N=2	N=1	N=2	N=1	N=2	N=1	N=2	N=1	N=2	N=1	N=2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Black/White Sample												
White	0.54	0.66	1.29	1.97	0.25	0.26	0.54	0.56	0.23	0.22	0.04	0.04
N	2,489,463	49,233	2,489,463	49,233	2,489,463	49,233	2,310,184	47,220	2,489,463	49,233	2,307,871	47,133
Black	0.61	0.66	1.63	2.01	0.31	0.24	0.84	0.80	0.24	0.24	0.07	0.06
N	3,268,414	49,233	3,268,414	49,233	3,268,414	49,233	3,026,164	47,499	3,268,414	49,233	3,024,613	47,460
Hispanic/White Sample												
White	0.54	0.63	1.29	1.79	0.25	0.26	0.54	0.56	0.23	0.20	0.04	0.06
N	2,489,463	17,368	2,489,463	17,368	2,489,463	17,368	2,310,184	16,767	2,489,463	17,368	2,307,871	16,708
Hispanic	0.57	0.63	1.38	1.78	0.25	0.25	0.82	0.81	0.18	0.17	0.45	0.39
N	641,568	17,368	641,568	17,368	641,568	17,368	598,776	16,788	641,568	17,368	549,267	15,424
Black/Hispanic Sample												
Hispanic	0.57	0.70	1.38	2.14	0.25	0.28	0.82	0.83	0.18	0.19	0.45	0.44
N	641,568	23,367	641,568	23,367	641,568	23,367	598,776	22,549	641,568	23,367	549,267	20,798
Black	0.61	0.70	1.63	2.17	0.31	0.28	0.84	0.82	0.24	0.23	0.07	0.09
N	3,268,414	23,367	3,268,414	23,367	3,268,414	23,367	3,026,164	22,599	3,268,414	23,367	3,024,613	22,559

Observations denote student-incident observations for all students in grades K-12, 2008-2018. N=2 for Black/White sample refers to incidents with two students, one of whom is White and the other is Black.

Table A2: Racial Differences in Disciplinary Outcomes, Coefficients on Student Covariates

	Dependent Var.: Was Suspended						Dependent Var.: Total Days Susp.					
	Sample: Black/White		Hispanic/White		Black/Hispanic		Black/White		Hispanic/White		Black/Hispanic	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Black	0.002	0.002			-0.001	0.004	0.033***	0.040***			0.037***	0.023
	(0.001)	(0.003)			(0.002)	(0.004)	(0.007)	(0.015)			(0.012)	(0.024)
Hispanic			-0.001	-0.005					-0.008	-0.014		
			(0.002)	(0.004)					(0.013)	(0.026)		
Female	-0.000	0.004	-0.008	0.024**	-0.004	0.004	-0.026**	-0.041	-0.029	0.061	-0.027	0.009
	(0.003)	(0.005)	(0.005)	(0.010)	(0.004)	(0.009)	(0.012)	(0.035)	(0.020)	(0.040)	(0.021)	(0.053)
Econ. disadv.	0.010***	0.003	0.004	0.012**	0.005	0.007	0.054***	0.004	0.040**	0.025	0.042**	0.051
	(0.002)	(0.004)	(0.003)	(0.005)	(0.003)	(0.006)	(0.011)	(0.024)	(0.019)	(0.035)	(0.020)	(0.036)
SWD	0.000	0.006	0.002	-0.005	-0.001	0.009	-0.015	0.073**	-0.004	0.065	-0.008	0.035
	(0.002)	(0.005)	(0.004)	(0.007)	(0.003)	(0.007)	(0.013)	(0.033)	(0.023)	(0.041)	(0.021)	(0.046)
LEP	-0.009	-0.003	0.004	0.003	-0.002	-0.001	0.029	-0.019	-0.016	0.008	0.013	0.001
	(0.005)	(0.010)	(0.004)	(0.006)	(0.003)	(0.006)	(0.031)	(0.058)	(0.022)	(0.042)	(0.019)	(0.041)
N	92,368	13,264	30,150	5,354	40,842	5,934	92,368	13,264	30,150	5,354	40,842	5,934
School-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Student characteristics	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
No previous referrals		Y		Y		Y		Y		Y		Y

*** p<0.01, ** p<0.05, * p<0.1. Sample includes student-incident observations in grades K-12 during 2008-2018. In addition to controls for student disadvantage, students with disabilities, and limited English proficiency, student characteristics also include indicators for birth year and month. Standard errors are clustered at the school level.

Table A3: Racial Differences in Disciplinary Outcomes, Adjusted for Lagged Achievement

	Dependent Var.: Was Suspended			Dependent Var.: Total Days Susp.		
	(1)	(2)	(3)	(4)	(5)	(6)
Black/White Sample						
Black	0.004***	0.001	0.000	0.042***	0.015*	0.042**
	(0.001)	(0.002)	(0.003)	(0.008)	(0.009)	(0.019)
N	51,228	51,228	8,460	51,228	51,228	8,460
Hispanic/White Sample						
Hispanic	0.001	-0.001	-0.003	-0.011	-0.026	-0.006
	(0.002)	(0.003)	(0.005)	(0.013)	(0.016)	(0.030)
N	16,674	16,674	3,328	16,674	16,674	3,328
Black/Hispanic Sample						
Black	0.001	0.001	0.004	0.033***	0.033**	0.013
	(0.002)	(0.003)	(0.005)	(0.013)	(0.015)	(0.030)
N	22,422	22,422	3,640	22,422	22,422	3,640
School-year FE	Y	Y	Y	Y	Y	Y
Incident FE	Y	Y	Y	Y	Y	Y
Student characteristics		Y	Y		Y	Y
Lagged math and reading scores		Y	Y		Y	Y
No previous referrals			Y			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, students with disabilities, and limited English proficient. The second and third specifications also control for lagged math and reading achievement. Standard errors are clustered at the school level.

Table A4: Categorization of Infractions: Objective vs. Subjective Discipline Guidelines

Objective	Leaving school without permission, skipping school, possession of a firearm or powerful explosive, dress code violation, skipping class, falsification of information, bomb threat, excessive tardiness, truancy, cutting class, late to class, cell phone use, leaving class without permission, excessive display of affection, gambling, possession of tobacco, mutual sexual contact between two students, use of tobacco
Subjective	Insubordination, sale of marijuana, fighting, rape, possession of chemical or drug paraphernalia, sale of controlled substances (other), sale of cocaine, sexual offense, sale of Ritalin, inappropriate items on school property, robbery with a dangerous weapon, disorderly conduct, inappropriate language/disrespect, honor code violation, repeat offender, bullying, verbal harassment, disruptive behavior, aggressive behavior, being in an unauthorized area, assault involving the use of a weapon, assault on student, violent assault not resulting in serious injury, robbery without a dangerous weapon, assault on school personnel not resulting in injury, assault resulting in serious injury, property damage, extortion, false fire alarm, other school defined offense, gang activity, sexual assault not involving rape or sexual offense, sexual harassment, misuse of school technology, use of counterfeit items, possession of a weapon (excluding firearms/explosives), unlawfully setting a fire, burning of a school building, communicating threats

Table A5: Racial Differences in Disciplinary Outcomes: Objective Incidents

	Dep Var.: Was Suspended		Dep Var.: Total Days Susp.	
	Full Sample	Interracial Pairs	Full Sample	Interracial Pairs
	(1)	(2)	(3)	(4)
Black/White Sample				
Black	0.004	0.001	0.009	0.006
	(0.002)	(0.003)	(0.006)	(0.011)
N	51,604	13,102	51,604	13,102
Hispanic/White Sample				
Hispanic	0.001	0.001	0.002	-0.014
	(0.002)	(0.004)	(0.007)	(0.014)
N	28,698	7,312	28,698	7,312
Black/Hispanic Sample				
Black	-0.003	-0.004	0.004	-0.009
	(0.003)	(0.005)	(0.010)	(0.017)
N	20,083	6,342	20,083	6,342
Incident FE		Y		Y

*** p<0.01, ** p<0.05, * p<0.1. All samples span grades K-12, 2008-2018, and include student-incident observations for objective incidents only. The full sample includes incidents involving up to two individuals belonging to the specified racial and ethnic groups, and excludes incidents with three or more students. Columns 2 and 4 restrict the sample to only incidents involving two individuals of specified racial and ethnic groups. Standard errors are clustered at the school level.

Table A6: Racial Differences in Disciplinary Outcomes: Subjective Incidents

	Dep Var.: Was Suspended		Dep Var.: Total Days Susp.	
	Full Sample	Interracial Pairs	Full Sample	Interracial Pairs
	(1)	(2)	(3)	(4)
Black/White Sample				
Black	0.004***	0.005***	0.048***	0.055***
	(0.001)	(0.001)	(0.006)	(0.007)
N	118,383	75,432	118,383	75,432
Hispanic/White Sample				
Hispanic	0.003*	0.003	0.009	0.004
	(0.002)	(0.002)	(0.010)	(0.013)
N	35,923	23,336	35,923	23,336
Black/Hispanic Sample				
Black	-0.001	0.001	0.018*	0.040***
	(0.001)	(0.002)	(0.010)	(0.011)
N	58,653	36,126	58,653	36,126
Incident FE		Y		Y

*** p<0.01, ** p<0.05, * p<0.1. All samples span grades K-12, 2008-2018, and include student-incident observations for subjective incidents only. The full sample includes incidents involving up to two individuals belonging to the specified racial and ethnic groups, and excludes incidents with three or more students. Columns 2 and 4 restrict the sample to only incidents involving two individuals of specified racial and ethnic groups. Standard errors are clustered at the school level.