Disparities in Student Discipline by Race and Family Income

Nathan Barrett
Tulane University

Andrew McEachin
RAND Corporation

Jonathan Mills
University of Arkansas

Jon Valant
Brookings Institution

This study explores student discipline disparities by race (black/white) and family income. First, we decompose gaps across districts, across schools in the same district, and within schools. Second, we assess disparities using regression models. Third, we examine punishments for fights between black and white or poor and non-poor students. We find that black and poor students are disciplined more often and harshly than their peers, with disparities arising across districts, across schools, and within schools. Moreover, black and poor students tend to receive slightly longer suspensions after getting into fights with white and non-poor peers, suggesting at least some degree of discriminatory punishment.

VERSION: May 2019

Disparities in Student Discipline by Race and Family Income

Nathan Barrett
Tulane University

Andrew McEachin
RAND Corporation

Jonathan Mills
University of Arkansas

Jon Valant
Brookings Institution

Abstract

This study explores student discipline disparities by race (black/white) and family income. First, we decompose gaps across districts, across schools in the same district, and within schools. Second, we assess disparities using regression models. Third, we examine punishments for fights between black and white or poor and non-poor students. We find that black and poor students are disciplined more often and harshly than their peers, with disparities arising across districts, across schools, and within schools. Moreover, black and poor students tend to receive slightly longer suspensions after getting into fights with white and non-poor peers, suggesting at least some degree of discriminatory punishment.

I. Introduction

In the United States, students of color are suspended and expelled from school at higher rates than white students. The U.S. Department of Education's Office for Civil Rights (2016) reports that, compared to white children, black children are 3.6 times more likely to receive an out-of-school suspension in preschool, 3.8 times more likely to receive an out-of-school suspension in grades K-12, and 2.2 times more likely to be referred to law enforcement or subject to a school-related arrest. Although the Office for Civil Rights does not release similar comparisons for poor and non-poor students nationwide, researchers have observed higher suspension rates for Arkansas students from low-income families than their peers (Anderson & Ritter, 2017) and found that black students who attend high-poverty schools are suspended at higher rates than black students who attend other schools (Loveless, 2017).

These gaps are among the most discussed and disputed topics in education policy. In 2014, the Obama administration issued a Dear Colleague Letter (DCL) that outlined federal laws prohibiting discrimination in school discipline and explained how "intentional discrimination" and "disparate impact" constitute violations of federal law (U.S. Department of Justice & U.S. Department of Education, 2014). Partly in response to the DCL, 22 states and Washington, DC passed laws to deemphasize exclusionary discipline (suspension and expulsion), while 23 of the largest 100 school districts required non-punitive discipline strategies or limited the use of suspension (Steinberg & Lacoe, 2017). In 2018, the Trump administration rescinded the DCL, claiming that it advanced "policy preferences and positions not required or contemplated" by federal law (U.S. Department of Justice & U.S. Department of Education, 2018).

At issue is neither the existence nor size of today's discipline disparities. Rather, much of the debate focuses on the causes of these disparities and, in particular, the possible role of

discriminatory practices by educators. Discrimination is evident in many aspects of American life, including employee hiring (Lavergne & Mullainathan, 2004), criminal justice (Anwar & Fang, 2015; Arnold, Dobbie, & Yang, 2018; Knowles, Persico, & Todd, 2001; Park, 2017), mortgage lending (Munnell et al., 1996), peer-to-peer lending (Pope & Sydnor, 2011), and medical care (Schulman et al., 1999). However, few studies have assessed discrimination in student discipline, in part because of the challenges of identifying it in administrative data that do not enable researchers to observe students' behaviors (Kinsler, 2011).

This study uses rich, student-level data from Louisiana from the 2000-01 through 2013-14 school years to examine the origins and causes of discipline disparities between black and white, and poor and non-poor, students. In doing so, it explicitly tests for a type of intentional discrimination as defined in the DCL—that students of different backgrounds receive different punishments for similar behaviors.

To begin, we decompose black/white and poor/non-poor suspension gaps into their within-school, across-school-within-district, and across-district components. This decomposition illuminates the organizational levels where gaps arise—a key step to identifying their causes and remedies. Next, we estimate the length of students' suspensions by race and poverty status while controlling for infraction type and student background characteristics, along with various combinations of school, grade, and year fixed effects. Finally, we examine our data for a type of intentional discrimination described in the DCL (and explained in the Background section): the punishments that result when a black and white student, or poor and non-poor student, get into a fight with one another.[1] We include models that control for student-level factors that might lead educators to differentiate punishments for nondiscriminatory reasons (e.g., showing more leniency for a student's first infraction).

Our results show strong relationships between race, poverty status, and rates of exclusionary discipline. For example, black students are suspended at approximately twice the rate of white students—25% and 12% per year, respectively—when we assess overall infraction rates (descriptively, without covariates). Poor students, defined as those eligible for free or reduced-price lunch (FRPL) in every year observed, are suspended at approximately twice the rate of non-poor students who never qualify for FRPL (21% and 9% per year, respectively).[2] These gaps arise from a combination of within-school, across-school-within-district, and across-district sources, with the majority of black/white and poor/non-poor gaps in suspension rates arising within schools. Both within and across schools, being black and poor is associated with longer suspensions, even conditional on a rich set of covariates. Finally, when we hone in on the punishments resulting from fights between black and white, and poor and non-poor, students, we find that black and poor students receive longer punishments—a finding robust to numerous specifications and particularly evident in middle and high school grades. Under certain assumptions that we describe in detail, and believe to be reasonable, these gaps constitute evidence of intentional discrimination in student discipline.

The paper proceeds with a review of research on student discipline disparities and discussion of the potential sources and causes of these disparities. We then describe our data, methods, and results, before concluding with our assessment of the study's implications for diagnosing and addressing race- and poverty-based discipline gaps.

II. Background on Discipline Disparities

Concerns about discipline disparities relate to both their causes (i.e., what produces the disparities) and effects (i.e., how high rates of exclusionary discipline might harm students of color and students in poverty). Regarding the effects, school suspensions are correlated with

negative outcomes such as low academic achievement (Arcia, 2006; Beck & Muschkin, 2012; Raffaele Mendez, Knoff, & Ferron, 2002; Raffaele Mendez, 2003; Skiba & Rausch, 2004), low probability of on-time graduation (Ekstrom, Goertz, Pollack, & Rock, 1986; Raffaele Mendez, 2003; Suh, Suh, & Houston, 2007; Wehlage & Rutter, 1986), and increased contact with the juvenile and criminal justice systems (Morgan, Salomon, Plotkin, & Cohen, 2014; Fabelo et al., 2011; Nicholson-Crotty, Birchmeier, & Valentine, 2009; Skiba, Arredondo, & Williams, 2014). These relationships are correlational, leaving unanswered questions about whether outcomes would have differed if students were disciplined differently (Steinberg & Lacoe, 2017). However, causal evidence of negative effects is emerging (Lacoe & Steinberg, 2018), and there is an intuitive argument that students learn more when they are present in school.[3]

This study focuses on the causes, not the effects, of discipline disparities. While the existence of discipline gaps has been documented for decades (e.g., Children's Defense Fund, 1975; Losen, Hodson, Keith II, Morrison, & Belway, 2015; U.S. Department of Education, 2016), what produces them remains a topic of considerable uncertainty and inquiry. These disparities surely arise from a complex mix of causes, some with origins inside of education systems and others with origins outside of them.

One empirical approach taken to illuminate the causes of, and potential remedies for, discipline disparities has been to assess whether gaps arise within or across schools. Gaps arising within schools could provide suggestive, though certainly not definitive, evidence that educators systematically discipline their black and white students differently for similar behaviors (a type of "intentional discrimination" that we discuss in detail). Gaps arising across schools might reflect more punitive practices in schools that serve high proportions of black students—a possibility consistent with the "racial threat" theory that authorities use more aggressive forms of

control in predominantly African-American settings (Blalock, 1967) and evident in school administrators' survey responses (Welch and Payne, 2010). Gaps also might arise across schools if students of different races and socioeconomic status behave differently in school, perhaps because of their heightened exposure to poverty, discrimination, and the related challenges outside of school (see Gregory, Skiba, & Noguera, 2010, for a review).

Anderson and Ritter (2017) used longitudinal, student-level infraction data from Arkansas to assess the extent to which gaps arise within and across schools. Conditioning on the type of infraction for which students were referred to the principal's office, they found that black and FRPL-eligible students were much more likely than their peers to receive exclusionary discipline. However, the gaps were markedly smaller in models with school fixed effects, leading the authors to conclude that the key drivers of racial discipline gaps likely relate primarily to differences in the types of schools that black and white students attend.[4] Kinsler (2011) and Skiba et al. (2014) similarly found evidence suggesting that overall discipline gaps arise primarily across schools, not from gaps within the same school.

Decomposing the within-school and across-school origins of discipline disparities is useful for illuminating the causes of, and potential remedies for, these disparities, but the existing literature is limited in what it reveals. One limitation is methodological. Comparisons of regression models with and without school fixed effects can provide misleading impressions of the relative balance of within-school and across-school sources of the gaps. The addition of school fixed effects can lead to students in relatively (or fully) segregated schools receiving little (or no) weight in regression estimates. A second limitation is that knowing whether disparities arise within or across schools still leaves considerable ambiguity about the specific causes of the disparities—and whether they reflect discriminatory behaviors by educators.[5]

Understanding discipline disparities requires attention to student behaviors, infractions, and punishments, and the relationships among them. We refer to *behaviors* as what students do in school, *infractions* as behaviors that schools document as misconduct, and *punishments* as the penalties associated with those infractions. Behaviors are generally not observable to researchers. However, student infractions and punishments are observable in some administrative datasets, which enables researchers to examine group differences in (a) the rates at which students receive infractions and (b) the severity of the punishments associated with those infractions.

Disparities in infractions or punishments can arise from a combination of three sources. The first source is true differences in behavior across groups. If, for example, black students use profanity in class more often than white students, then, all else equal, we would expect disparities in profanity infractions. Similarly, if, when students use profanity, black students' language tends to be more explicit, we would expect disparities in punishments for the same recorded infractions. These gaps could arise from educators responding evenhandedly to the behaviors they observe. While various types of discrimination might contribute to such gaps— perhaps, for example, black students act out in response to teachers having lower expectations for them (Gershenson, Holt, & Papageorge, 2016)—these causes are distinct from administrators unevenly punishing students of different backgrounds for similar behaviors.

The second possible source is what we call intentional discrimination in student discipline. We adopt both the term and its definition from the DCL, which describes intentional discrimination as school staff disciplining students differently based on their race ("different treatment"). This could take several forms. A policy that is discriminatory on its face (e.g., explicitly says that students of different races will be disciplined differently) or a policy created to target students of a certain race (e.g., based on hairstyles or clothing) would, in the absence of

a legitimate educational purpose, constitute intentional discrimination. Another form of intentional discrimination—and the focus of this study—involves school officials disciplining students of different backgrounds differently for the same offense. The DCL gives an example of a fight between a group of Asian-American students and a group of Native-American students, all without prior discipline histories, that leads to more severe punishments for the Asian-American students despite no indication that their behaviors warranted harsher punishment. The letter notes that this type of discrimination could arise in decisions about whether to administer discipline (i.e., infractions) or how harshly to do so (i.e., punishments). Intentional discrimination could, but need not, involve racial animus.[6] Rather, its defining characteristic is that students' race (or other background characteristic) affects how they are disciplined.[7]

A third possible source of disparities involves circumstances that, while correlated with students' background characteristics, are not fundamentally about student background. For instance, school officials might, in a racially nondiscriminatory way, treat students more leniently for their first offense. The DCL alludes to this in describing the fight between Asian-American and Native-American students. If the Asian-American students received harsher punishments for that fight because they have longer discipline histories—and, therefore, had previously used up the warnings given a first offense—then group differences in punishments might not constitute intentional discrimination.

This paper is concerned with the possibility that school officials punish black students or students in poverty more harshly than other students, for similar behaviors, without strong educational or circumstantial reasons for doing so. A few studies provide evidence suggestive of this type discrimination. Gilliam, Maupin, Reyes, Accavitti, and Shic (2016) asked 132 early childhood teachers to observe videos of four young children—a black boy, black girl, white boy,

and white girl—to watch for "challenging behavior in the classroom." The researchers selected videos with, in their view, no signs of challenging behavior, and then tracked participants' eye movements to see whether teachers tended to monitor certain students. They found that teachers focused a disproportionate share of their time watching black children, and especially black boys.[8] Okonofua and Eberhardt (2015) randomly assigned K-12 teachers to read vignettes of misbehaving students with either stereotypically black or white student names. They found that teachers felt more troubled by black students' misbehaviors and were more inclined to regard the black students as troublemakers. A second experiment, designed similarly to the first, showed that teachers were more likely to envision suspending the black students in the future.

Other studies have used student-level administrative data to examine whether black and white students and disciplined differently. Anderson and Ritter (2017) estimated models with infraction fixed effects, finding, using Arkansas data, that black students received longer punishments than white students. As described above, they attributed these gaps primarily to differences in the types of schools attended by black and white students. Using data from North Carolina, Lindsay and Hart (2017) found that black students who had more black teachers were less likely to be suspended or expelled. They observed that the types of behaviors that leave discretion to teachers about whether and how to punish—such as "willful disobedience"— contributed heavily to these patterns. Kinsler (2011) found that North Carolina teachers referred black students to administrators for behavioral infractions at higher rates than white students, but conditional on the type of infraction assessed by administrators, black and white students received similar punishments. One exception was for fights, as black students' suspensions for fighting were about 20% longer than white students' suspensions for fighting (in models with school fixed effects). He did not find significant disparities when restricting his sample to cases

of multiple students being suspended for fighting on the same day in the same school, although this sample could have included instances such as one fight between white students and another (more or less severe) fight between black students.

This study of student discipline in Louisiana contributes to the literature in several ways. First, and most importantly, we use rich, student-level data to isolate infractions for which black and white, and poor and non-poor, students seemingly should receive equal punishments if not for discriminatory discipline practices within schools: fights between two students of different backgrounds with similar (or no) prior discipline records. This provides an "actions-based" test of intentional discrimination in student discipline.[9] Within this fight sample, we explore several related questions of interest, including where discrimination is evident and how it interacts with other student characteristics such as gender and academic achievement. Second, we assess the relative contribution of within-school and across-school disparities using a decomposition approach well suited to that task. Our findings indicate that a substantially larger share of discipline disparities arises within schools than the prior literature suggests. And third, by focusing on gaps in the southern state of Louisiana, we illuminate the race and class dynamics of student discipline in a state with large populations of black and poor students and a long history of problems where race, class, and schools collide (Egalite, Mills, & Wolf, 2017).

III. Data

This study uses data provided by the Louisiana Department of Education (LDOE) for the 2000-01 through 2013-14 school years. LDOE provided records for students in grades K-12 in Louisiana public schools, which includes both traditional public and charter schools. Taken together, these data contain nearly 10 million student-year observations, with approximately 1.8 million unique observations spread over 14 school years.

The LDOE data contain variables commonly found in state administrative data, including students' basic demographic characteristics (e.g., race, FRPL eligibility, gender, and special education status), grade levels, and test scores from Louisiana's state assessments, which we standardized by test type, subject, year, and grade. Table 1 provides descriptive statistics. Our race variable contains a non-missing value for nearly all student observations. We coded that variable into three categories: black (46% of the sample), white (48%), and other race (5%), with the latter encompassing several racial and ethnic groups. The sample is split almost evenly between male (51%) and female (49%) students, and approximately 12% of students have a disability and the corresponding special education (SPED) status. SPED status is an important covariate for studying discipline gaps by race and poverty status, since black and poor students comprise a disproportionate share of the population of disabled students and students with disabilities are suspended at higher rates than their peers (Losen et al., 2015). This is true despite federal rules protecting students from punishments for behaviors caused by manifestations of their disabilities.

Our free/reduced-price lunch (FRPL) variable contains a non-missing value for each student in each year, and a student's FRPL status could change from year to year. Based on the full sample of student-year observations, we find 55% eligible for free lunch and 7% eligible for reduced-price lunch. Since a central focus of this study is to compare outcomes for poor and non-poor students, we constructed variables that represent the overall picture—across years—of a student's family income. Following Michelmore and Dynarski's (2017) observation that persistent FRPL eligibility is a better indicator of poverty than eligibility in any particular year, we construct an "always FRPL" variable to identify students who qualified for FRPL in every year observed (48% of the student-year observations) and a "never FRPL" variable to identify

students who never qualified for FRPL in our data (27% of the student-year observations). This study's comparisons of poor and non-poor students reflect outcomes for always-FRPL and never-FRPL students, respectively. We omit sometimes-FRPL students from these comparisons in order to draw a relatively sharp contrast between poor and non-poor students. Students in the sometimes-FRPL group (24% of the student-year observations) might have experienced changes in family income over the period observed—or might simply have administrative data that do not reflect their true financial situations in some years (e.g., because of data or paperwork errors). However, we retain this sometimes-FRPL group when using family income as a covariate to keep from dropping observations.[10]

LDOE also provided rich data on student infractions and their corresponding punishments. These data include infractions that resulted in either an in-school or out-of-school suspension but do not include data on incidents that did not yield suspensions. The infraction variable distinguishes between 49 types of offenses that resulted in suspensions. We coded these infractions as "violent" if they seemed to have a relatively high probability of inflicting serious physical or emotional harm and as "nonviolent" otherwise.[11]

Table 1 shows large raw differences in suspension rates by race and poverty status. Twenty-five percent of black students were suspended in a given year, with 13% suspended for a violent infraction and 19% for a nonviolent infraction (some students were suspended for both a violent and nonviolent infraction in the same year). Only 12% of white students were suspended in a given year, with 5% suspended for a violent infraction and 9% for a nonviolent infraction. Twenty-one percent of poor students were suspended in a given year (11% for a violent infraction and 15% for a nonviolent infraction), compared to 9% of non-poor students (4% for a violent infraction and 7% for a nonviolent infraction).

Table 2 displays the total counts and percentages of these infractions statewide, as well as disaggregated for black, white, poor, and non-poor students. The nine most common infraction types account for 92% of the recorded infractions. We grouped all other infraction codes in an "Other" category. The most common infractions, in order, are willful disobedience (23% of all infractions), fights in school (14%), habitually violates a rule (13%), and disrespects authority (13%). In general, the distributions of infractions for black, white, poor, and non-poor students are similar, although relatively large proportions of black and poor students' suspensions resulted from fights in school (16% for each, compared to 10% for white students and 9% for non-poor students). Approximately 29% of black students' suspensions and 31% of poor students' suspensions resulted from violent infractions, compared to 28% of white students' suspensions and 26% of non-poor students' suspensions.

Our punishment data also include the length, in days, of each suspension, which we use as a measure of the punishment's severity. The data show where the suspension was administered (i.e., in school, out of school, or, in a few cases, at another site). Figure 1 shows statewide trends in the use of in-school suspensions, disaggregated for black and white students. Figure 2 shows the same for poor and non-poor students. These figures indicate that both in-school and out-of-school suspensions are common. They also indicate differences across groups in whether students' suspensions are handled in school or out of school—a topic we return to later. Unless otherwise noted, we do not distinguish between in-school and out-of-school suspensions, as each is a form of punishment that removes students from the classroom.

Finally, our data contain the date on which each infraction occurred. We use these dates to analyze fights between black and white, and poor and non-poor, students. To identify these incidents, we flagged fights that occurred on the same day in the same school. Our data do not

explicitly link the students who fought one another, so we limited our fight sample to days in which exactly two students were disciplined for fighting: a black student and white student (for the race analyses) or a poor student and non-poor student (for the poverty analyses). We cannot rule out the possibility that some observations in our fight sample result from two separate incidents on the same day in the same school, one with only a black/poor student suspended and the other with only a white/non-poor student suspended (each coded as a fight for just one student rather than an assault or other infraction). However, since this would be an unusual set of circumstances that seems unlikely to comprise a substantial portion of our fight sample, we refer to these observations as fights between black/poor and white/non-poor students.[12]

IV. Methods

We examine discipline disparities in three ways to assess the nature of these disparities and the discriminatory practices they might, or might not, reflect. First, we decompose raw discipline gaps across districts, across schools within the same district, and within schools. This reveals where disparities arise. Second, we explore whether students of different backgrounds receive similar punishments for observably similar infractions. Disparities in, for example, how schools punish black and white students for infractions coded as willful disobedience might suggest discriminatory practices. However, we believe our third set of analyses—the fight analyses—provide a more credible glimpse of a potential form of intentional discrimination in student punishment.

A. *Decomposing Gaps into Across-District, Across-School, and Within-School Components*

We begin by breaking overall, raw black/white and poor/non-poor discipline gaps into across-district, across-school-within-district, and within-school components. We do so for two discipline outcomes: whether students were suspended and for how many days they were

suspended. Specifically, we start by defining the raw average discipline rate ($\bar{R}_{isd}$) for a given group of students in a given grade weighted across students, schools, and districts. We define $\bar{R}_{isd}$ as:

(1)
$$\bar{R}_{isd} = \frac{\sum_i \sum_s \sum_d Group_{isd} Y_{isd}}{\sum_i \sum_s \sum_d Group_{isd}}$$

where $Group_{isd}$ takes a "1" for a black (or FRPL) student, and a "0" for a white (or non-FRPL) student, and Y_{isd} indicates a suspension outcome for student i in school s and district d. We define $\bar{\bar{R}}_{isd}$ as the discipline rate for the historically non-disadvantaged group (white or non-poor). The overall gap in suspension rates between two groups of students is simply $\bar{R}_{isd} - \bar{\bar{R}}_{isd}$.

We then take an approach similar to others who have decomposed gaps related to students' exposure to teachers of varying experience (Clotfelter, Ladd, & Vigdor, 2005) and performance (Goldhaber, Lavery, & Theobald, 2015) into across-district, across-school, and within-school components. To do this, we define additional sets of suspensions rates. The first, $\bar{R}_d$ and $\bar{\bar{R}}_d$, replaces Y_{isd} with the average suspension outcome in a given district, $\bar{Y}_d$. The second, $\bar{R}_{sd}$ and $\bar{\bar{R}}_{sd}$, replaces Y_{isd} with the average suspension outcome in a given school, $\bar{Y}_{sd}$. With these two rates, we decompose the overall gap into the following components:

(2) $\bar{R}_{isd} - \bar{\bar{R}}_{isd} = \bar{R}_d - \bar{\bar{R}}_d + \left(\left(\bar{R}_{sd} - \bar{\bar{R}}_{sd} \right) - \left(\bar{R}_d - \bar{\bar{R}}_d \right) \right) + \left(\left(\bar{R}_{isd} - \bar{\bar{R}}_{isd} \right) - \left(\bar{R}_{sd} - \bar{\bar{R}}_{sd} \right) \right)$

Given that $\bar{R}_d$ and $\bar{\bar{R}}_d$ are a function of the same variable ($\bar{Y}_d$), any difference between these two rates arises from black/white or poor/non-poor students attending different districts. Similarly, a non-zero value of $\left(\left(\bar{R}_{sd} - \bar{\bar{R}}_{sd} \right) - \left(\bar{R}_d - \bar{\bar{R}}_d \right) \right)$ is due to black/white or poor/non-poor students attending different schools within the same district. Finally, a non-zero value of $\left(\left(\bar{R}_{isd} - \bar{\bar{R}}_{isd} \right) - \left(\bar{R}_{sd} - \bar{\bar{R}}_{sd} \right) \right)$ is due to within-school differences in suspensions for

black/white or poor/non-poor students. We present kernel density plots of these gaps by grade, as well as the raw numerical gaps and the share of gaps explained at each level.[13]

B. Regression Analyses Examining Gaps

Second, we estimate linear probability and OLS regression models to move beyond unconditional gaps and look for gaps within and across schools in models that include assorted covariates. Specifically, these models assess whether white and black, and poor and non-poor, students receive different punishments—in the number of days suspended—for the same infractions as coded in our data. An important goal of this study is to assess whether gaps arise from intentionally discriminatory behaviors by adults working in schools. As such, we estimate models that account for the nonrandom sorting of students to schools and spurious correlations between student characteristics and the propensity to be punished.

Formally, we use models of the following form:

(3) $$Y_{igsnt} = \alpha_0 + \beta_1 Race_{it} + \beta_2 Poverty_i + \beta_3 X_{igst} + \theta_n + \delta_{gst} + \varepsilon_{igsnt}$$

where discipline outcomes for student i in grade g in school s for infraction n in time t are modeled as a linear function of: race, $Race_{it}$, with binary indicators for black and other race (with white students as the reference group); poverty status, $Poverty_i$, with binary indicators for always-FRPL and sometimes-FRPL (with never-FRPL students as the reference group); a vector of observable student characteristics related to special education, gender, math and English language arts (ELA) scores from the prior school year; a set of indicator variables for the type of infraction, θ_n; school-grade-year (SGY) fixed effects, δ_{gst} (or grade and year fixed effects in some models); and an idiosyncratic error term, ε_{igsnt}. We cluster our standard errors by school-grade-years.

The primary coefficients of interest across our statistical models estimate the black/white and poor/non-poor (always-FRPL/never-FRPL) differences in discipline outcomes. Some models contain additional student covariates and others do not. Models with these covariates are useful for comparing discipline outcomes for students who are similar but for their race or poverty status. However, these models could control for characteristics that contribute to how racial or socioeconomic groups are perceived. Since each type of comparison is substantively important, we include models with and without these covariates.

Students in Louisiana are not randomly assigned to schools, and prior research suggests that a large share of the variation in differential patterns of student punishment is explained across, rather than within, schools (Anderson & Ritter, 2017; Kinsler, 2011; Skiba et al., 2014). We include models with grade and year fixed effects and others with SGY fixed effects to look for overall gaps and within-school/grade gaps, respectively. Students within SGY cells likely have more similar in-school and out-of-school experiences, and focusing on gaps within SGYs helps to identify gaps local to a cohort of peers within a particular grade, school, and year.

Models with infraction fixed effects, with days of suspension as their dependent variable, are well suited for identifying gaps in suspension length that arise from similar infractions. These models are not necessarily well suited for identifying evidence of discriminatory school practices. For example, if black students are suspended for more days than white students for getting into fights, it could reflect administrators discriminating against black students. Alternatively (or additionally), it could reflect fights involving black students being systematically more severe than—or otherwise different from—fights involving white students. Thus, we turn to our third set of analyses.

C. Gaps from Fights between Black and White, and Poor and Non-Poor, Students

Our third methodological approach exploits a particular setting in which disparities seem most likely to reflect intentionally discriminatory school discipline practices. We test for differences in the length of suspensions that black and white, and poor and non-poor, students receive when they fight. We do so while, in some models, controlling for students' prior discipline histories and other background characteristics. These covariates account for the possibility that administrators might punish students differently based on their prior discipline records—and that certain groups of students (e.g., those who have fought previously or have lower test scores) might be more likely to instigate or escalate these fights.

The key identifying assumption is that when black and white (or poor and non-poor) students with similar discipline records and background characteristics get into a fight at school, the black and white students' behaviors should warrant equal punishments. We note that these incidents involve two students suspended for fighting, as indicated by the infraction variable. We exclude fights that resulted in different infraction codes—such as one student written up for assault—since they could give rise to legitimately different punishments. Still, this assumption could be violated if, in these very particular circumstances, black and white students exhibit systematically different behaviors that warrant different punishments. On the other hand, if this assumption holds, then gaps likely reflect discriminatory behavior from school leaders who treat similar-behaving students of different backgrounds differently.

We examine two related outcomes for these fight analyses: the number of days for which students are suspended and whether a student receives a longer suspension than the student with whom he or she fights. Formally, our base model for these analyses is:

(4) $$Y_{ist} = \alpha_0 + \beta_1 Group_{it} + \tau_{st} + \varepsilon_{ist}$$

where discipline outcomes Y for student i in school s in time t are modeled as a linear function of race or poverty, $Group_{it}$, with binary indicators for black or poor students (with white and non-poor students, respectively, as the reference group); school-year fixed effects, τ_{st}; and an idiosyncratic error term, ε_{ist}. We cluster standard errors by school-years. For analyses with the number of suspension days as their outcome, we censor the number of days to 20 to limit the influence of outlier suspension lengths—and check the robustness of our findings to alternate ways of censoring.[14]

In addition to the base model, we include specifications that control for a vector of observable student characteristics related to race (for the poverty analyses), poverty status (for the race analyses), special education, gender, and math and ELA test scores from the prior school year. Results are generally similar regardless of whether we include these covariates. Our preferred models exclude them, as we believe these models provide more substantively important comparisons of how black and poor students—with all of the characteristics that contribute to perceptions of those groups—are punished relative to white and non-poor students.

In many of our analyses, we restrict the sample to fights between students with similar, limited histories of fights or other suspensions. We do this to address a possible source of bias in the fight models—that schools might punish students who are repeatedly in trouble more severely than students who are rarely in trouble, and that discipline history could relate to race and poverty status. Along with showing results for all qualifying fights between black and white or poor and non-poor students, we show results after limiting the sample to fights between two students: 1) without a prior fight in that school year; 2) without a prior fight at any point in our data; 3) without a prior suspension of any kind in that school year; and 4) without a prior

suspension of any kind at any point in our data.[15] We also provide an assortment of robustness checks.

We include results from analyses that modify, in various ways, the sample or dependent variable in the fight models. We show punishments between particular subgroups of students, such as black females and white females. We explore heterogeneity across different types of school settings where fights occur. For example, we observe the race of individual school administrators, which enables us to compare disparities that arise from fights in schools with all-white, all-non-white, and partially-white administrative staffs. Administrators typically determine suspension lengths, and prior research indicates relationships between discipline outcomes and principal, teacher, and student races (Kinsler, 2011; Lindsay & Hart, 2017). Our data do not indicate which administrator in a school actually determined a punishment, so we disaggregate by the racial composition of the full administrative staff (see Price and Wolfers, 2010, for an analogous approach). Finally, we assess whether administrators used in-school suspensions and out-of-school suspensions differently for white and black, or poor and non-poor, students who were suspended for fighting each other. While we do not believe that differences in this outcome would constitute clear evidence of discrimination, they could illuminate patterns in how schools punish students of different backgrounds.

V. Results

A. Decomposing Gaps into Across-District, Across-School, and Within-School Components

First, we decompose gaps in two outcomes—the likelihood of getting suspended and the number of days suspended—into across-district, across-schools-within-the-same-district, and within-school components. We present this analysis visually in Figures 3 through 6, with the underlying raw numbers presented in Appendix Tables A1 and A2.

A few notable patterns emerge from the data. The first relates to the changing size and nature of discipline gaps across the age spectrum from kindergarten through grade 12. In Figure 3, for example, the overall gap between black and white students in whether they were suspended (shaded gray density) starts around 3 percentage points in kindergarten, grows to a peak of 21 percentage points in grades 6 and 7, and shrinks to 9 percentage points in grade 12. A similar pattern appears in the poor/non-poor gaps in whether students were suspended (Figure 4) and for the days suspended outcome (Figures 5 and 6). The larger gaps in middle school could reflect higher rates of exclusionary discipline after students leave elementary school, with those rates declining in high school as many struggling students drop out.

The second interesting pattern—different from patterns observed in other studies (e.g., Anderson & Ritter, 2017; Kinsler, 2011)—is that within-school differences account for a large portion of the overall black/white and poor/non-poor gaps, especially in middle and high school.[16] This is particularly the case for outcomes showing whether students were suspended. For both the black/white and poor/non-poor comparisons, within-school differences account for at least 50 percent of the gap in kindergarten and grades 5 through 12. Still, differences across schools constitute an important share of the black/white and poor/non-poor gaps as well. With respect to the suspension length outcome, black/white and poor/non-poor discipline gaps are spread more evenly across districts, across schools, and within schools.

The relatively large within-school differences in suspension rates for black and white, and poor and non-poor, students have important implications. They indicate that many Louisiana students attend schools in which black and poor students are suspended at higher rates than white and non-poor students—and that many Louisiana administrators are suspending their black and poor students at higher rates than they suspend their white and non-poor students. This suggests

that much of the discipline gap is not a product of different practices or resources across schools and districts.

This finding of within-school differences is not, in itself, necessarily evidence of discrimination, as these differences in punishments could reflect systematic differences in behaviors. Our subsequent analyses examine this question more closely.

B. Regression Analyses Examining Gaps

We next examine black/white and poor/non-poor discipline gaps within a regression framework. Tables 3 and 4 show how suspension length (in days) varies for students of different groups, with infraction fixed effects to control for the type of infraction that produced the suspension. The primary coefficients of interest are "Black" and "Poor." However, as we describe below, while differences observed in Tables 3 and 4 might suggest possible discrimination, they do not provide definitive evidence.

Across all models, black (poor) students receive significantly longer suspensions than white (non-poor) students do. The first column of Table 3 shows that black students' suspensions are, on average, 0.43 days longer than white students' suspensions (whose suspensions last an average of 1.74 days). When we include SGY fixed effects to look within school-grade-year cohorts, the difference falls to 0.10 days but remains statistically significant. The third and fourth columns of Table 3 show results for the analogous poor/non-poor comparisons—and suggest similarly sized gaps. Without SGY fixed effects, we estimate poor students' suspensions to be 0.41 days longer than non-poor students' suspensions (which last 1.75 days); with SGY fixed effects, the difference is 0.09 days (and again significant). The next three columns of this table include both race and poverty variables in the same model, with Column 7 also controlling for whether students were suspended in the prior year. Controlling for prior suspensions could help

to make these black/white and poor/non-poor groups more comparable, although it could be a problematic covariate if the prior year's suspensions reflect racial or socioeconomic bias. In Table 3, including both race and poverty variables results in the black and poor coefficients dropping somewhat but retaining their significance. The suspended-in-the-prior-year indicator is significant itself but its inclusion only modestly affects the other variables estimated. The last three columns in Table 3 show similar models with a full set of student covariates included. The most saturated model—which controls for SGY fixed effects, student background characteristics, and prior year suspensions (Column 10)—indicates that black students' suspensions are 0.04 days longer than white students' suspensions ($p<.01$) and poor students' suspensions are 0.07 days longer than non-poor students' suspensions ($p<.01$).

Table 4 replicates the models from Table 3 after limiting the sample to only students' first suspensions of the year. Perhaps schools are more lenient in punishing students' first infractions, and since black and poor students in our data are more likely to have multiple infractions in the same year, the gaps observed in Table 3 could result from punishments intensifying as students repeatedly get in trouble. Even with this restriction, black/white and poor/non-poor gaps appear. Without SGY fixed effects, black/white gaps range from 0.28 to 0.41 days and poor/non-poor gaps range from 0.16 to 0.36 days. With SGY fixed effects, black/white gaps range from 0.02 to 0.05 days and poor/non-poor gaps range from 0.03 to 0.05 days. All of these differences are statistically significant. The black/white and poor/non-poor differences evident in Tables 3 and 4 are robust to further limiting the suspensions to only out-of-school suspensions (see Appendix Tables A3 and A4).

The results in this section have important connections to the analyses that precede them and the analyses that follow. First, we note that it might be tempting to compare coefficients

from models with and without SGY fixed effects in Tables 3 and 4 in order to draw conclusions about the relative contributions of within-school and across-school differences to overall discipline gaps. However, these comparisons can mislead. A model that regresses suspension outcomes on race and includes SGY fixed effects will assign the greatest weight to students in SGYs with a relatively even balance of white and black students. It will give no weight at all to students in fully segregated SGYs, since there is no within-SGY variation in student race, and relatively little weight to students in mostly segregated SGYs. As a result, differences in gaps across models with and without SGY fixed effects could reflect patterns in where gaps arise— but also could reflect differences in the populations represented after this implicit weighting. It may be, for example, that segregated schools with particularly severe (or modest) problems with discipline gaps receive minimal weight in the models with SGY fixed effects. For this reason, we prefer our decomposition method for assessing the relative impact of within-school, across-school (within district), and across-district factors.

Second, while the results from this section, more than the decomposition section, suggest the possibility of discriminatory practices within schools, too much remains unobserved for this evidence to be conclusive. It remains possible that behaviors of black and poor students systematically differ from behaviors of white and non-poor students even when they yield the same infraction code. Perhaps, for example, black students' "willful disobedience," as it is recorded, is generally more severe than white students' willful disobedience. While we have no reason to believe this is the case, our next set of analyses—involving fights between black and white students or poor and non-poor students—help us focus even more narrowly on punishments for the same infraction types that arise from very similar circumstances.

C. Gaps from Fights between Black and White, and Poor and Non-Poor, Students

The remaining tables and figures examine punishment gaps arising from fights between black and white students and between poor and non-poor students. In focusing on these fights, we isolate incidents between two students on the same day that resulted in each student getting a fighting-related suspension. We examine these gaps in a variety of ways, including looking for heterogeneity across certain groups of students and schools.

We begin by providing a descriptive examination of the differences in punishment length for fights between black/white and poor/non-poor students. Figure 7 depicts the fights that yielded different punishments for the black and white, or poor and non-poor, students. The first bar in the first figure of Panel A, for example, depicts the number of fights in which the black student was suspended for one day longer than the white student. We observe that the vast majority of black/white and poor/non-poor fights with different punishment lengths have gaps in punishments of under five days. At each day-gap interval below five days, black and poor students account for a larger share of students receiving more days of suspension than white and non-poor students, respectively.

Our primary estimates of the fight disparities appear in Tables 5 and 6. Table 5 shows gaps in the days suspended for black/white fights (Panel A) and poor/non-poor fights (Panel B). It shows results with and without student-level covariates for each of the five ways in which we define the sample of black/white and poor/non-poor fights. Estimates from the first two columns show that, whether or not we include student covariates, black students' suspensions are 0.05 days longer than white students' suspensions. We find very stable estimates of black-white disparities regardless of how we attempt to control for prior discipline history. In each case, black students receive suspensions that are 0.04 to 0.06 days longer than white students, with all

of these differences statistically significant ($p<.05$). Differences in the poor/non-poor samples are less consistent across specifications. Several specifications that control for student characteristics are not statistically significant, and estimates of the size of the poor/non-poor disparities vary more than estimates of the black/white disparities. However, most specifications indicate that poor students receive longer suspensions than their non-poor counterparts.

Complete results for the full sample models in this table appear in columns 1 and 3 of Appendix Tables A5 (for race comparisons) and A6 (for poverty comparisons). To check robustness, we also limit the sample to only out-of-school suspensions. These results appear in Table A7 and are consistent with the full sample though generally slightly larger in magnitude.

Table 6 shows a parallel set of results with a slightly different outcome—whether the black or poor student received a longer suspension than the white or non-poor student. These results could be less sensitive to outliers in suspension length.[17] Here, we find consistent evidence that black students are more likely to receive a longer suspension than the white students with whom they fight. Panel A shows that black students are 1.2 to 2.1 percentage points more likely to receive a longer suspension. For example, the results from Column 7 indicate that when a fight occurs between a black student and white student who have not been suspended in that year, the white student receives the longer suspension about 7.5 percent of the time (displayed as the constant) while the black student receives the longer suspension about 9.1 percent of the time. In most cases, students are suspended for the same number of days. Panel B of this table shows results for fights between poor and non-poor students. Here, we see consistent evidence that poor students are more likely to receive longer suspension than the non-poor students they fight than vice versa. Across all models, poor students are 1.3 to 2.8 percentage points more likely than non-poor students to receive a longer suspension. Here, too, most fights

result in the same number of days of suspension for the two students. Complete results for the

full sample models in this table appear in columns 5 and 7 of Appendix Tables A5 (for race

comparisons) and A6 (for poverty comparisons). Appendix Tables A8 and A9 show that these

findings are robust to classifying only out-of-school suspensions as suspensions (as opposed to

both out-of-school and in-school suspensions, as we do in our preferred models).

These results establish gaps in how severely schools punish black and white, and poor

and non-poor, students for getting into fights. Next, we look for heterogeneity in these results by

student and school characteristics.

Table 7 shows the suspensions resulting from fights between various subgroups of

students, disaggregated by several combinations of race, poverty status, gender, and academic

performance. The first row in this table indicates that of the 20,142 fights in our data between a

black student and white student, the black student received a longer suspension 11.2 percent of

the time, the white student receive a longer suspension 9.2 percent of the time, and the students

received the same length of suspension 79.6 percent of the time. This difference for black and

white students is statistically significant ($p<.01$).

Several comparisons in this table are notable. Most importantly, black students are

consistently more likely to receive longer suspensions than white students and poor students are

consistently more likely to receive longer suspensions than non-poor students, across many

combinations of student characteristics. (When race and poverty collide such that a poor white

student fights a non-poor black student, the differences are not statistically significant.) Perhaps

most strikingly, after an interracial fight, black boys more often receive longer suspensions than

white boys (1.8 percentage-point difference; $p<.01$), black girls more often receive longer

suspensions than white girls (2.3 percentage points; $p<.01$), black boys more often receive longer

suspensions than white girls (3.4 percentage points; $p<.05$), and black girls more often receive

longer suspensions than white boys (2.6 percentage points; $p<.1$). Similar patterns appear in the

interactions between poverty and gender, although some of these differences are not statistically

significant. Interestingly, when boys and girls fight (irrespective of race), we do not observe

punishment gaps. Students with low test scores are more likely to receive the longer suspension

when they get into a fight with a student with high test scores. Finally, we note that Appendix

Table A10 shows the same table after restricting to fights in which it results in each student's

first suspension of the year. Although the samples are smaller, the patterns in this table are

similar to those in Table 7.

Table 8 displays fight punishment disparities between black and white students

disaggregated by various school-level characteristics. It shows that disparities for fights between

a black student and white student are larger at the middle school and high school levels than the

elementary school level. Patterns by the racial demographics of administrators and the full

student body are less clear. Disaggregating by school location, we see some indication that these

gaps are relatively large in urban schools. Table A11 shows the same analyses for fights between

poor and non-poor students. Here, the gaps appear larger in low-poverty schools and schools

outside of urban areas.

The final table, Table 9, looks again at the punishments resulting from black/white and

poor/non-poor fights but replaces the length of punishment with another aspect of the

punishment within the school's control: whether the student serves the suspension in school or

out of school. We do not believe that differences by race or poverty status would be plainly

discriminatory. Neither in-school nor out-of-school suspension is an inherently more severe

punishment than the other, and school administrators could have valid reasons for making

different decisions for different students. For example, if poor students are less likely to have someplace to go during the school day—if, say, low-income parents have less flexible work schedules—then we might expect poor students to receive disproportionately more in-school suspensions. Interestingly, the results in Table 9 point in the opposite direction. When a fight between a poor and non-poor student yields one out-of-school suspension and one in-school suspension, the poor student is (slightly but significantly) more likely to receive the out-of-school suspension. The same is true for black students relative to white students, and these patterns appear in both the full sample of fights and the subset of fights that yielded each student's first suspension of the year (Table A12). While not clear evidence of discrimination, these differences could reflect school leaders' beliefs about how best to handle different groups of students.

Looking across the fight analyses, we see clear, consistent differences in how schools punish black or poor students versus white or non-poor students. The magnitudes of the differences are not all large, but the differences appear across a variety of analyses and are robust to numerous analytical decisions.

VI. Discussion

Questions about why and students of color and students in poverty are suspended at higher rates than their peers—and what to do about it—have emerged among the most pressing and controversial issues facing education policymakers. At this point, there is little dispute that black and poor students are suspended and expelled at higher rates. However, addressing inequities in exclusionary discipline requires not only establishing that gaps exist but also explaining their origins. Gaps in exclusionary discipline could arise from true differences in students' behaviors, differences in how schools translate those behaviors to infractions, and

differences in how schools punish students for the same infractions. The reality that gaps could arise within schools, across schools within districts, or across districts adds complexity, while the lack of available data on the true behaviors of large numbers of students imposes constraints on how researchers can assess these gaps.

This study uses rich administrative data from the state of Louisiana to explore the causes of black/white and poor/non-poor gaps in exclusionary discipline. Louisiana is an appropriate setting for this study due to its large (and relatively even) populations of black and white students and its historical challenges related to race, class, and schools. We observe large black/white and poor/non-poor differences in student discipline, with these gaps evident in a variety of contexts.

A fundamental—and much debated—question about discipline gaps is whether they arise from school leaders' discriminatory treatment of minority or poor students. Discrimination of this type is extremely difficult to identify in large-scale administrative data. This is largely because researchers typically do not observe students' true behaviors—only reflections of those behaviors as perceived and addressed by educators. Some forms of discrimination are virtually impossible to observe in administrative data. For example, if teachers tend to be more forgiving of certain students' use of profanity, opting not to refer those students for punishment, then researchers likely will never know that the students used profanity in the first place. This places severe limitations on quantitative researchers' abilities to identify the true degree and types of discrimination that occurs in schools. Most of what happens between students and educators is simply unobservable to the researcher.

In this context, we looked for evidence of a very particular type of discipline disparity—one that would allow us to credibly rule out other explanations for why schools punish black or poor students more harshly than other groups. We compared the punishments that arise after

black and white, or poor and non-poor, students get into fights, in some cases controlling for the

students' prior discipline histories and background characteristics. Even in these narrowly

defined cases, we find that schools punish black students more severely than they punish white

students. The difference amounts to about 0.05 days across black-white fights—with black

students one to two percentage points more likely to receive a longer suspension. We observed

similar differences, although somewhat less consistent, in comparisons of poor and non-poor

students' punishments.

These models still cannot provide conclusive evidence of racial bias, since we must rely

on some unverifiable assumptions, including that black students do not systematically behave

differently than white students in these interracial fights (after accounting for students'

background characteristics). Still, with our findings robust to numerous alternate specifications,

this study provides evidence that suggests at least some degree of intentional discrimination in

student discipline.

Moreover, although these particular differences are small in magnitude, there is reason to

believe that disparities could be larger in circumstances less amenable to this type of analysis.

We examine black/white and poor/non-poor fights because we believe they provide the most

credible glimpse in our data at whether schools punish students differently for similar behaviors.

However, we should note that administrators who issue different punishments after interracial

fights know that they are doing so—and know that others will see the different punishments

administered. This could temper the resulting disparities. If so, one might expect larger

disparities when comparing the punishments after two black students fight to the punishments

after two white students fight. This, too, could constitute a type of "intentional discrimination" as

defined by the federal government, but we focus on interracial (and poor/non-poor) fights in

hopes of being as cautious and careful as possible about what goes unobserved. It is important to note that in no way does this study aim to quantify the full extent of discrimination in student discipline. Rather, it aims to test for the presence of intentional discrimination where we can most credibly rule out alternate explanations.

Finally, while we have examined within-school (within-fight) gaps for evidence of discrimination, we should note that across-school gaps also can reflect various forms of discrimination. If schools that enroll high percentages of students of color and students in poverty employ harsher discipline practices than other schools, then poor and minority students could accrue discipline records that non-poor and white students would not accrue for similar behaviors. Broader economic and societal patterns of discrimination also could produce varying behaviors from students of different races and socioeconomic classes. These represent different types of problems than within-school gaps—and would require solutions tailored to those problems—but still can reflect discrimination in student discipline. As this study shows, discipline gaps arise from multiple sources and likely require more than one type of response.

---

[1] Our analysis focuses exclusively on situations in which two—and only two—students in the same school and grade were recorded as getting into a fight on the same day. Although our data do not explicitly state that these students fought one another, we infer it from the data, as described in the Data section.

[2] Based on guidelines from the National School Lunch Program, students whose household income is at or below 130 percent of the poverty line are eligible for free lunch and students whose household income is at or below 185 percent of the poverty line are eligible for reduced-price lunch.

[3] The decision of whether to remove a student from school could also have implications for that student's classmates. These externalities have not received as much attention from researchers as the direct effects on the suspended students (see Kinsler, 2013, for analysis that considers the externalities of suspending students along with the deterrent and direct effects on suspended students).

[4] Anderson and Ritter (2017) use a dependent variable that takes a value of 1 if a student receives an out-of-school suspension and a 0 if the student receives an in-school suspension, a non-suspension punishment (e.g., detention), or no punishment.

---

[5] Finding that discipline disparities arise across schools rather than within them does not rule out the possibility of discriminatory or inequitable causes of the disparities. For example, various forms of discrimination could lead black and white students to behave differently or attend different types of schools. This is analogous in some ways to the expansive research on wage gaps by race and gender (e.g., Cotton, 1988; Groshen, 1991; Reimers, 1983; Weichselbaumer & Winter-Ebmer, 2005). This research tends to show that controlling for variables such as occupation, education, and experience yield smaller estimates of wage gaps than simple raw comparisons, although race and gender differences on these covariates could themselves result from various forms and sources of discrimination.

[6] Much of the economic research on discrimination attempts to disentangle racial animus from statistical discrimination (e.g., Anwar & Fang, 2015; Arnold, Dobbie, & Yang, 2018; Fang & Moro, 2010). Each could exist in student discipline. For example, school officials could be personally prejudiced or hostile toward students of color (animus), or, in the absence of knowing how to apportion blame for an incident they did not witness, they could rely on their prior observations of which groups of students tend to be blameworthy (a type of statistical discrimination). This would be difficult to tease apart empirically in the context of student discipline, and that is not our intent with this study. Rather, we seek to identify discrimination regardless of whether that discrimination is rooted in racial animus.

[7] The DCL identifies "different treatment" (intentional discrimination) and "disparate impact" as the two distinct types of unlawful discrimination. While different treatment refers to discipline policies that are either explicitly discriminatory or facially neutral but administered unevenly across races, disparate impact refers to discipline policies that are facially neutral and evenhandedly administered but have a "disproportionate and unjustified effect on students of a particular race" (U.S. Department of Justice & U.S. Department of Education, 2014).

[8] In a second study, they did not find a relationship between children's races and whether teachers thought the children should be suspended or expelled.

[9] The economic literature contrasts actions-based tests of discrimination (which use the ways in which different groups are treated to assess discrimination) to outcomes-based tests (which use groups' subsequent behaviors to draw inferences about whether some were treated discriminatorily) (e.g. Anwar & Fang, 2006, 2015; Arnold, Dobbie, & Yang, 2018). These studies argue that outcomes-based measures are, in some contexts, better suited to identifying discrimination. However, we pursue actions-based tests in this study for several reasons. First, while a strength of some outcomes-based tests is their ability to disentangle racial animus from statistical discrimination and potential omitted variables bias, we are less concerned about the distinction between racial animus and statistical discrimination and have data well suited to minimizing omitted variables bias. Second, an outcomes-based test is not feasible in our context, since—unlike with a setting like parole decisions—we do not observe subsequent outcomes that are measured objectively and independent of the influence of the officials who administered the initial punishments. Rather, the teachers and administrators who determined students' punishments often have a say in their subsequent discipline records. Third, while parole decisions are fundamentally forward-looking and outcomes-oriented—shaped by Parole Boards' assessments of how prisoners might act if released—school officials' decisions about how to punish students are perhaps more fundamentally based on what seems like a fair punishment for the student's earlier behavior.

[10] Figures A1 through A4 depict the racial and socioeconomic demographics in Louisiana districts and schools. These figures show that schools tend to be more segregated than districts. More students are enrolled in schools with less than 10%, or more than 90%, of students being black or free lunch-eligible than are enrolled in districts with those demographics. However, with respect to both schools and districts, many students are enrolled in relatively diverse settings by race and poverty status.

[11] The following infractions were coded as violent (as labeled in LDOE data): immoral or vicious practices; habits injurious to his/her associates; weapon (Sec 921 of Title 18 of the U.S. Code); weapon (not prohibited by federal law); throws missiles liable to injure others; fights while under school supervision; commits any other serious offense; murder; assault and/or battery; rape and/or sexual battery; kidnapping; arson; misappropriate with violence; use weapon prohibited by federal law; possess blade with length less than 2.5 in.; serious bodily injury; bullying; cyber bullying; and sexual harassment.

[12] Moreover, even if the two students did not get in fights with one another, it remains true that school administrators would be assessing punishments for the same infraction on the same day. We would not expect, *a priori*, to observe variation in punishments to be systematically related to either race or poverty, barring differences in student behavior. The existence of differences would instead suggest some systematic factor at play.

[13] Note that a negative gap would imply that white/non-poor students are more likely to be suspended or have longer suspensions than black/poor students.

[14] We also estimated models with specific-fight fixed effects to represent each individual pair of fighters. Results from these models were virtually identical to the results presented in this paper.

[15] We focus on first infractions—rather than controlling for prior discipline records—because students' prior discipline records could make for problematic controls if black or poor students previously received suspensions for behaviors that would not have yielded suspensions for white or non-poor students. Still, we are cautious about restricting to students' first suspensions or fights for similar reasons. If, for example, a fight occurs between a black student and white student with no prior suspensions—but the white student previously did something that would have gotten the black student suspended had he done it—then these sample restrictions may not serve the purpose we intend them to serve.

[16] Kinsler (2011) did find evidence of within-school differences in the likelihood of being referred to the principals' office for a behavioral offense. However, conditional on being referred to the principal's office and controlling for infraction, the within-school differences in the likelihood or length of suspensions were not statistically significant.

[17] Our preferred models censor at 20 days to limit the influence of these outliers. We assessed the robustness of our findings to various numbers of days at which to censor. Appendix Table A9 shows parallel sets of results depending on whether we censor the overall suspension length at 5 days, 10 days, 20 days, or not at all. This table also shows the consequences of censoring the gap in suspension days given to the two fighters. We display results after censoring these gaps at 3 days, 5 days, or (from our preferred models) not at all. As Figure 7 demonstrates, very few fights results in gaps greater than 5 days.

References

Anderson, K. P., & Ritter, G. W. (2017). Disparate use of exclusionary discipline: Evidence on inequities in school discipline from a U.S. state. *Education Policy Analysis Archives, 25*(49), 1-36.

Anwar, S., & Fang, H. (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review, 96*(1), 127-151.

Anwar, S., & Fang, H. (2015). Testing for racial prejudice in the parole board release process: Theory and evidence. *Journal of Legal Studies, 44*(1), 1-37.

Arcia, E. (2006). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district. *Education and Urban Society, 38*(3), 359-369.

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. *Quarterly Journal of Economics, 133*(4), 1885-1932.

Beck, A. N., & Muschkin, C. G. (2012). The enduring impact of race: Understanding disparities in student disciplinary infractions and achievement. *Sociological Perspectives, 55*(4), 637-662.

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology, 66*(1), 5-20.

Blalock, H. M. (1967). *Toward a theory of minority-group relations*. New York, NY: Wiley and Sons.

Children's Defense Fund. (1974). *Children out of school in America*. Cambridge, MA: Children's Defense Fund.

Children's Defense Fund. (1975). *School suspensions: Are they helping children?* Cambridge, MA: Children's Defense Fund.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, *24*, 377–392.

Cotton, J. (1988). On the decomposition of wage differentials. *Review of Economics and Statistics, 70*(2), 236-243.

Egalite, A. J., Mills, J. N., & Wolf, P. J. (2017). The impact of targeted school vouchers on racial stratification in Louisiana schools. *Education and Urban Society, 49*(3), 271-296.

Ekstrom, R. B., Goertz, M. E., Pollack, J. M., & Rock, D. A. (1986). Who drops out of high school and why? Findings from a national study. *Teachers College Record, 87*(3), 356-73.

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks III, M. P., Booth, E. A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement.* New York, NY: Council of State Governments Justice Center.

Fang, H, & Moro, A. (2010). *Theories of statistical discrimination and affirmative action: A survey* (Working Paper 15860). Cambridge, MA: National Bureau of Economic Research.

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

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). *Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions?* New Haven, CT: Yale Child Study Center.

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, *44*(5), 293–307.

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

Groshen, E. L. (1991). The structure of the female/male wage differential: Is it who you are, what you do, or where you work?. *Journal of Human Resources, 26*(3), 457-472.

Haynes, N. M., Emmons, C., & Ben-Avie, M. (1997). School climate as a factor in student adjustment and achievement. *Journal of Educational and Psychological Consultation, 8*(3), 321-329.

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

Kinsler, J. (2013). School discipline: a source or salve for the racial achievement gap? *International Economic Review, 54*(1), 355-383.

Knowles, J., Persico, N., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy, 109*(1), 203-229.

Lavergne, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*(4), 991-1013.

Lindsay, C. A., & Hart, C. M. D. (2017). Exposure to same-race teachers and student disciplinary outcomes for black students in North Carolina. *Educational Evaluation and Policy Analysis, 39*(3), 485-510.

Lacoe, J., & Steinberg, M. P. (2018). Do suspensions affect student outcomes? Educational Evaluation and Policy Analysis, 41(1), 34–62.

Losen, D., Hodson, C., Keith II, M. A., Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap?* Los Angeles, CA: The Center for Civil Rights Remedies, University of California.

Loveless, T. (2017). *The 2017 Brown Center report on American education: How well are American students learning?* Washington, DC: Brookings Institution.

Michelmore, K., & Dynarski, S. (2017). The gap within the gap: Using longitudinal data to understand income differences in educational outcomes. *AERA Open, 3*(1), 1-18.

Morgan, E., Salomon, N., Plotkin, M., & Cohen, R. (2014). *The school discipline consensus report: Strategies from the field to keep students engaged in school and out of the juvenile justice system.* New York, NY: Council of State Governments Justice Center.

Munnell, A. H., Tootell, G. M., Browne, L. E., & McEneaney, J. (1996). Mortgage lending in Boston: Interpreting HMDA data. *American Economic Review, 86*(1), 25-53.

Nicholson-Crotty, S., Birchmeier, Z., & Valentine, D. (2009). Exploring the impact of school discipline on racial disproportion in the juvenile justice system. *Social Science Quarterly, 90*(4), 1003-1018.

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes race and the disciplining of young students. *Psychological Science, 26*(5), 617-624.

Park, K. H. (2017). The impact of judicial elections in the sentencing of black crime. *Journal of Human Resources, 52*(4), 998-1031.

Pope, D. G., & Sydnor, J. R. (2011). What's in a Picture? Evidence of Discrimination from Prosper. com. *Journal of Human Resources, 46*(1), 53-92.

Price, J., & Wolfers, J. (2010). Racial discrimination among NBA referees. *Quarterly Journal of Economics, 125*(4), 1859-1887.

Raffaele Mendez, L. M. (2003). Predictors of suspension and negative outcomes: A longitudinal investigation. *New Directions for Youth Development, 99*, 17-33.

Raffaele Mendez, L. M., Knoff, H. M., & Ferron, J. M. (2002). School demographic variables and out-of-school suspension rates: A quantitative and qualitative analysis of a large, ethnically diverse school district. *Psychology in the Schools, 39*(3), 259-277.

Reimers, C. W. (1983). Labor market discrimination against Hispanic and black men. *Review of Economics and Statistics, 65*(4), 570-579.

Schulman, K. A., Berlin, J. A., Harless, W., Kerner, J. F., Sistrunk, S., Gersh, B. J., Dubé, R., Taleghani, C. K., Burke, J. E., Williams, S., Eisenberg, J. M., Ayers, W., & Escarce, J. J. (1999). The effect of race and sex on physicians' recommendations for cardiac catheterization. *New England Journal of Medicine, 340*(8), 618-626.

Skiba, R. J., Arredondo, M. I., & Williams, N. T. (2014). More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity and Excellence in Education*, *47*(4), 546-564.

Skiba, R. J., Chung, C. G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). 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.

Skiba, R., & Rausch, M. K. (2004). *The relationship between achievement, discipline, and race: An analysis of actors predicting ISTEP scores. Children Left Behind Policy Briefs,*

Supplementary analysis 2-D. Bloomington, IN: Center for Evaluation and Education Policy, Indiana University. Retrieved from http://files.eric.ed.gov/fulltext/ED488899.pdf

Skiba, R. J., & Williams, N. T. (2014). *Are black kids worse? Myths and facts about racial differences in behavior: A summary of the literature.* Bloomington, IN: The Equity Project at Indiana University. Retrieved from http://www.indiana.edu/~atlantic/wp-content/uploads/2014/03/African-American-Differential-Behavior_031214.pdf

Steinberg, M. P., & Lacoe, J. (2017). What do we know about school discipline reform?: Assessing the alternatives to suspensions and expulsions. *Education Next, 17*(1), 44-52.

Suh, S., Suh, J., & Houston, I. (2007). Predictors of categorical at-risk high school dropouts. *Journal of Counseling and Development, 85*(2), 196-203.

U. S. Department of Education. (2016). *School climate and discipline: Know the data.* Retrieved from http://www2.ed.gov/policy/gen/guid/school-discipline/index.html#featured

U. S. Department of Education Office for Civil Rights. (2016). *2013-14 Civil Rights Data Collection: A first look.* Retrieved from https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf

U. S. Department of Justice, & U.S. Department of Education. (2014). *Dear Colleague Letter on the nondiscriminatory administration of school discipline.* Retrieved from https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.html

U. S. Department of Justice, & U.S. Department of Education. (2018). *Dear Colleague Letter.* Retrieved from https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201812.pdf

Wehlage, G. G., & Rutter, R. A. (1986). Dropping out: How much do schools contribute to the problem?. *Teachers College Record*, *87*(3), 374-92.

Weichselbaumer, D., & Winter-Ebmer, R. (2005). A meta-analysis of the international gender wage gap. *Journal of Economic Surveys, 19*(3), 479-511.

Welch, K., & Payne, A. A. (2010). Racial threat and punitive school discipline. *Social Problems*, *57*(1), 25-48.

Table 1
Descriptive Statistics

	All students		Race comparison		Poverty comparison	
			Black	White	Poor	Non-Poor
	N	%	%	%	%	%
Total	9,999,240					
Suspended	1,802,382	18%	25%	12%	21%	9%
Suspended, violent infraction	870,791	9%	13%	5%	11%	4%
Suspended, nonviolent infraction	1,370,761	14%	19%	9%	15%	7%
Race/Ethnicity	9,991,085					
Black	4,630,883	46%	100%	0%	66%	14%
White	4,810,988	48%	0%	100%	29%	81%
Other	549,214	5%	0%	0%	6%	5%
Poverty status	9,999,240					
Free lunch	5,456,642	55%	77%	33%	90%	0%
Reduced-price lunch	738,126	7%	6%	9%	10%	0%
Full-price lunch	3,804,472	38%	17%	58%	0%	100%
Always FRPL ("Poor")	4,840,970	48%	69%	29%	100%	0%
Sometimes FRPL	2,417,077	24%	23%	25%	0%	0%
Never FRPL ("Non-poor")	2,741,193	27%	8%	46%	0%	100%
Special education status	8,993,890					
SPED	1,057,802	12%	13%	11%	14%	7%
Non-SPED	7,936,088	88%	87%	89%	86%	93%
Gender	9,999,240					
Male	5,126,563	51%	51%	52%	51%	52%
Female	4,872,677	49%	49%	48%	49%	48%
Standardized state test score (t-1)						
English language arts	3,622,000	0.12	-0.17	0.37	-0.14	0.60
Math	3,625,553	0.11	-0.25	0.42	-0.16	0.63
Science	3,512,728	0.10	-0.30	0.45	-0.19	0.63
Social studies	3,511,937	0.10	-0.24	0.39	-0.17	0.60

Notes. The unit of observation is the student-year, meaning that students observed in multiple years account for multiple observations. In total, the data contain 9,999,240 student-year observations from 1,778,128 students. The notes in Table 2 list the violent infractions. "Always FRPL," "Sometimes FRPL," and "Never FRPL" describe the student's free or reduced-price lunch status across all years observed. "Poor" refers to the always-FRPL group and "non-poor" refers to the never-FRPL group.

Table 2

Number of Infractions by Infraction Type and Student Subgroup

	All students		Race Comparison				Poverty Comparison			
			Black		White		Poor		Non-Poor	
	N	%	N	%	N	%	N	%	N	%
Total number of suspensions	4,258,559		2,915,863		1,223,363		2,457,157		500,119	
Willful disobedience	999,339	23%	699,943	24%	273,162	22%	580,604	24%	109,643	22%
Fights in school	604,719	14%	467,074	16%	125,606	10%	393,086	16%	46,512	9%
Habitually violates a rule	559,983	13%	393,453	13%	151,917	12%	333,812	14%	63,334	13%
Disrespects authority	536,668	13%	393,442	13%	131,529	11%	323,438	13%	51,823	10%
Any other serious offense	315,827	7%	186,856	6%	118,142	10%	162,527	7%	52,851	11%
Profane	255,728	6%	164,830	6%	83,912	7%	140,249	6%	32,680	7%
Leaves school	256,553	6%	157,183	5%	88,563	7%	118,476	5%	44,061	9%
Habitually tardy	203,312	5%	133,782	5%	61,372	5%	98,032	4%	35,373	7%
Injurious habits	183,594	4%	118,794	4%	58,913	5%	119,299	5%	19,258	4%
Other	342,836	8%	200,506	7%	130,247	11%	187,634	8%	44,584	9%
Violent infractions	1,232,478	29%	856,312	29%	343,206	28%	753,695	31%	132,025	26%
Nonviolent infractions	3,026,081	71%	2,059,551	71%	880,157	72%	1,703,462	69%	368,094	74%

Notes. The unit of observation is the infraction, so some students have multiple observations within the same year while students who did not commit an infraction are not represented. The table lists the nine most common infractions and aggregates all other infractions as "Other." Columns with percentages show the percentage of infractions recorded for that group of students that were of the infraction type listed. The following infractions were coded as violent (as labeled in LDOE data): immoral or vicious practices; habits injurious to his/her associates; weapon (Sec 921 of Title 18 of the U.S. Code); weapon (not prohibited by federal law); throws missiles liable to injure others; fights while under school supervision; commits any other serious offense; murder; assault and/or battery; rape and/or sexual battery; kidnapping; arson; misappropriate with violence; use weapon prohibited by federal law; possess blade with length less than 2.5 in.; serious bodily injury; bullying; cyber bullying; and sexual harassment.

Table 3
Predictors of Length of Suspension (in Days)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black	0.427***	0.099***			0.378***	0.088***	0.066***	0.287***	0.062***	0.043***
	(0.006)	(0.003)			(0.006)	(0.003)	(0.004)	(0.008)	(0.005)	(0.005)
Other race	0.085***	-0.018**			0.057***	-0.024***	-0.022***	0.035**	-0.033***	-0.028**
	(0.010)	(0.007)			(0.010)	(0.007)	(0.008)	(0.014)	(0.011)	(0.011)
Poor (always FRPL)			0.414***	0.092***	0.205***	0.056***	0.051***	0.226***	0.085***	0.071***
			(0.007)	(0.004)	(0.007)	(0.004)	(0.005)	(0.009)	(0.006)	(0.007)
Sometimes FRPL			0.339***	0.092***	0.199***	0.068***	0.064***	0.184***	0.080***	0.067***
			(0.007)	(0.004)	(0.006)	(0.004)	(0.005)	(0.008)	(0.006)	(0.007)
SPED								-0.103***	-0.064***	-0.075***
								(0.008)	(0.006)	(0.006)
Male								0.032***	0.035***	0.015***
								(0.005)	(0.004)	(0.004)
Math score (t-1)								-0.055***	-0.017***	-0.013***
								(0.004)	(0.003)	(0.003)
ELA score (t-1)								-0.062***	-0.033***	-0.027***
								(0.004)	(0.003)	(0.003)
Suspended in prior year							0.187***			0.179***
							(0.003)			(0.004)
Constant	1.739***	2.074***	1.751***	2.060***	1.635***	2.028***	1.906***	1.867***	1.963***	1.890***
	(0.030)	(0.004)	(0.030)	(0.005)	(0.030)	(0.005)	(0.006)	(0.298)	(0.008)	(0.009)
Observations	4,253,425	4,253,425	4,256,323	4,256,323	4,253,425	4,253,425	3,806,497	1,922,514	1,922,514	1,884,608
R-squared	0.048	0.183	0.045	0.183	0.049	0.183	0.181	0.046	0.180	0.180
Year FE	Yes	No	Yes	No	Yes	No	No	Yes	No	No
Grade FE	Yes	No	Yes	No	Yes	No	No	Yes	No	No
SGY FEs	No	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes

Notes. The unit of observation is the infraction, and the sample is restricted to students who were suspended. All models contain infraction fixed effects. Omitted reference groups are white students, non-poor students (students never eligible for free or reduced-price lunch), and females. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within school-grade-year. "SGY FEs" refers to school-grade-year fixed effects.
*** $p<.01$, ** $p<.05$, * $p<.1$

Table 4

Predictors of Length of Suspension for First Suspension of Year (in Days)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black	0.408***	0.047***			0.366***	0.040***	0.033***	0.282***	0.023***	0.019***
	(0.006)	(0.004)			(0.006)	(0.004)	(0.005)	(0.008)	(0.006)	(0.007)
Other race	0.086***	-0.023**			0.062***	-0.027***	-0.027***	0.050***	-0.014	-0.014
	(0.010)	(0.009)			(0.010)	(0.009)	(0.010)	(0.015)	(0.014)	(0.014)
Poor (always FRPL)			0.361***	0.048***	0.161***	0.033***	0.037***	0.174***	0.041***	0.040***
			(0.007)	(0.005)	(0.007)	(0.005)	(0.006)	(0.010)	(0.008)	(0.008)
Sometimes FRPL			0.285***	0.052***	0.156***	0.042***	0.045***	0.132***	0.035***	0.033***
			(0.007)	(0.005)	(0.006)	(0.005)	(0.006)	(0.009)	(0.008)	(0.008)
SPED								-0.065***	-0.027***	-0.031***
								(0.008)	(0.008)	(0.008)
Male								-0.018***	-0.025***	-0.030***
								(0.005)	(0.005)	(0.005)
Math score (*t*-1)								-0.039***	-0.004	-0.002
								(0.004)	(0.004)	(0.004)
ELA score (*t*-1)								-0.044***	-0.014***	-0.013***
								(0.004)	(0.004)	(0.004)
Suspended in prior year							0.048***			0.041***
							(0.004)			(0.005)
Constant	1.597***	1.904***	1.628***	1.891***	1.519***	1.878***	1.834***	1.638***	1.877***	1.861***
	(0.028)	(0.005)	(0.029)	(0.006)	(0.028)	(0.006)	(0.007)	(0.471)	(0.010)	(0.010)
Observations	1,801,105	1,801,105	1,802,382	1,802,382	1,801,105	1,801,105	1,593,672	802,597	802,597	784,941
R-squared	0.067	0.247	0.062	0.247	0.068	0.247	0.244	0.064	0.245	0.244
Year FE	Yes	No	Yes	No	Yes	No	No	Yes	No	No
Grade FE	Yes	No	Yes	No	Yes	No	No	Yes	No	No
SGY FEs	No	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes

Notes. The unit of observation is the infraction, and the sample is restricted to students who were suspended. All models contain infraction fixed effects. Omitted reference groups are white students, non-poor students (students never eligible for free or reduced-price lunch), and females. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within school-grade-year. "SGY FEs" refers to school-grade-year fixed effects.

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 5

Gaps in Days Suspended for Black/White and Poor/Non-Poor Fights

	Full sample		First fight of year		First fight ever		First suspension of year		First suspension ever	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Race comparison										
Black student	0.054***	0.048***	0.047***	0.044***	0.049***	0.046***	0.041***	0.047***	0.054***	0.059**
	(0.010)	(0.013)	(0.010)	(0.014)	(0.013)	(0.018)	(0.011)	(0.016)	(0.018)	(0.026)
Constant	2.892***	2.992***	2.998***	3.111***	2.984***	3.069***	2.887***	2.989***	2.944***	3.016***
	(0.027)	(0.043)	(0.005)	(0.038)	(0.049)	(0.078)	(0.005)	(0.043)	(0.061)	(0.102)
Observations	40,284	40,284	29,824	29,824	17,232	17,232	21,492	21,492	10,506	10,506
R-squared	0.008	0.014	0.008	0.008	0.007	0.007	0.009	0.008	0.007	0.007
Panel B: Poverty comparison										
Poor student	0.071***	0.049***	0.041***	0.022	0.042**	0.022	0.072***	0.033	0.098***	0.074***
	(0.013)	(0.018)	(0.014)	(0.019)	(0.017)	(0.022)	(0.015)	(0.021)	(0.021)	(0.027)
Constant	3.075***	3.093***	3.178***	3.198***	3.246***	3.299***	3.046***	3.022***	3.173***	3.149***
	(0.034)	(0.055)	(0.007)	(0.056)	(0.039)	(0.085)	(0.007)	(0.060)	(0.050)	(0.105)
Observations	20,380	20,380	15,826	15,826	11,946	11,946	11,860	11,860	7,826	7,826
R-squared	0.007	0.011	0.007	0.007	0.006	0.006	0.008	0.012	0.008	0.011
Controls for student characteristics	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls for number of years in data	No	No	No	No	Yes	Yes	No	No	Yes	Yes

Notes. Table shows the difference in the number of days for which the black (or poor) student was suspended minus the number of days for which the white (or non-poor) student was suspended. The unit of observation is the infraction, and the sample is restricted to students who were suspended for fighting with a student of a different race (Panel A) or poverty status (Panel B). The "first fight" and "first suspension" sample restrictions apply to both students involved in the fight. For example, the "first fight of year" columns restrict the sample to fights between two students who had not been suspended for a fight earlier in that school year. All models contain school-year fixed effects. Student controls consist of FRPL status (Panel A only), black and other race (Panel B only), special education status, gender, and math and ELA scores from the prior year. The reference group for "Black student" is white students. The reference group for "Poor student" is non-poor students. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within schools.

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 6

Who Received Longer Suspensions for Black/White and Poor/Non-Poor Fights

	Full sample		First fight of year		First fight ever		First suspension of year		First suspension ever	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Race comparison										
Black student	0.017***	0.013***	0.014***	0.012***	0.017***	0.014***	0.016***	0.013***	0.021***	0.019***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.006)
Constant	0.046***	0.058***	0.079***	0.088***	0.074***	0.086***	0.075***	0.087***	0.061***	0.071***
	(0.004)	(0.006)	(0.002)	(0.005)	(0.004)	(0.008)	(0.002)	(0.005)	(0.005)	(0.011)
Observations	40,284	40,284	29,824	29,824	17,232	17,232	21,492	21,492	10,506	10,506
R-squared	0.018	0.021	0.007	0.01	0.007	0.01	0.008	0.011	0.01	0.011
Panel B: Poverty comparison										
Poor student	0.023***	0.021***	0.018***	0.016***	0.016***	0.013**	0.026***	0.022***	0.028***	0.026***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)
Constant	0.032***	0.021***	0.071***	0.061***	0.071***	0.059***	0.066***	0.056***	0.054***	0.046***
	(0.006)	(0.007)	(0.002)	(0.006)	(0.004)	(0.009)	(0.003)	(0.007)	(0.005)	(0.012)
Observations	20,380	20,380	15,826	15,826	11,946	11,946	11,860	11,860	7,826	7,826
R-squared	0.009	0.011	0.008	0.008	0.007	0.007	0.008	0.01	0.008	0.01
Controls for student characteristics	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls for number of years in data	No	No	No	No	Yes	Yes	No	No	Yes	Yes

Notes. Table shows the difference in the probability that black (or poor) student was suspended longer than white (or non-poor) student minus the probability that the white (or non-poor) student was suspended longer than the black (or poor) student). The unit of observation is the infraction, and the sample is restricted to students who were suspended for fighting with a student of a different race (Panel A) or poverty status (Panel B). The "first fight" and "first suspension" sample restrictions apply to both students involved in the fight. For example, the "first fight of year" columns restrict the sample to fights between two students who had not been suspended for a fight earlier in that school year. All models contain school-year fixed effects (and not fight occurrence fixed effects). Student controls consist of FRPL status (Panel A only), black and other race (Panel B only), special education status, gender, and math and ELA scores from the prior year. The reference group for "Black student" is white students. The reference group for "FRPL student" is non-FRPL students. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within schools.

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 7

Who Received Longer Suspensions in Fights Between Students of Various Subgroups

Student 1	Student 2	Total number of fights	Same suspension length	Student 1 suspended longer	Student 2 suspended longer	Z-score	p-value
Subgroups							
Black	White	20,142	79.6%	11.2%	9.2%	6.541	0.000
Black	Black	81,719	77.1%	--	--	--	--
White	White	20,144	83.1%	--	--	--	--
Poor	Non-poor	10,190	81.1%	10.7%	8.1%	6.449	0.000
Poor	Poor	52,446	78.3%	--	--	--	--
Non-poor	Non-poor	3,167	85.8%	--	--	--	--
Male	Female	17,142	69.1%	15.5%	15.4%	0.239	0.811
Male	Male	75,900	79.7%	--	--	--	--
Female	Female	34,306	80.8%	--	--	--	--
Low scorer	High scorer	3,227	81.5%	10.7%	7.8%	3.996	0.000
Low scorer	Low scorer	9,738	78.4%	--	--	--	--
High scorer	High scorer	1,466	85.9%	--	--	--	--
Subgroup interactions with race							
Black, Poor	White, Poor	5,930	79.5%	10.9%	9.6%	2.209	0.027
Black, NP	White, NP	459	84.3%	8.9%	6.8%	1.228	0.219
Black, Poor	White, NP	2,797	81.6%	11.4%	7.0%	5.596	0.000
Black, NP	White, Poor	352	81.5%	8.2%	10.2%	-0.911	0.362
Black, Male	White, Male	14,899	81.2%	10.3%	8.5%	5.140	0.000
Black, Female	White, Female	2,889	81.7%	10.3%	8.0%	3.148	0.002
Black, Male	White, Female	996	68.9%	17.3%	13.9%	2.102	0.036
Black, Female	White, Male	1,358	66.2%	18.2%	15.6%	1.792	0.073
Black, High scorer	White, High scorer	302	88.4%	7.0%	4.6%	1.219	0.223
Black, Low scorer	White, Low scorer	974	79.5%	11.7%	8.8%	2.090	0.037
Black, High scorer	White, Low scorer	127	82.7%	7.9%	9.4%	-0.446	0.656
Black, Low scorer	White, High scorer	690	81.6%	11.2%	7.2%	2.514	0.012
Subgroup interactions with poverty							
Poor, Male	NP, Male	6,991	82.3%	10.0%	7.6%	5.005	0.000
Poor, Female	NP, Female	2,108	82.6%	10.4%	7.0%	3.988	0.000
Poor, Male	NP, Female	480	70.6%	16.5%	12.9%	1.550	0.121
Poor, Female	NP, Male	611	71.0%	15.4%	13.6%	0.894	0.371
Poor, High scorer	NP, High scorer	214	86.9%	9.8%	3.3%	2.737	0.006
Poor, Low scorer	NP, Low scorer	317	79.8%	11.0%	9.1%	0.791	0.429
Poor, High scorer	NP, Low scorer	86	82.6%	11.6%	5.8%	1.351	0.177
Poor, Low scorer	NP, High scorer	245	84.5%	11.4%	4.1%	3.040	0.002

Notes. The unit of observation is the infraction, and samples are restricted to fights between a student from the first group listed and a student from the second group listed. "Poor" indicates that the student appeared eligible for free or reduced-price lunch in all years of our data. "NP" indicates that the student never appeared eligible for free or reduced-price lunch in our data. "High scorer" is defined as a student in the top quintile of average performance, where average performance is calculated as the mean score across all tested subjects. "Low scorer" is defined as a student in the bottom quintile of performance across all tested subjects. Significance tests assess whether the probability that Group 1 received a longer punishment differs from the probability that Group 2 received a longer punishment. Observation numbers indicate count of observed fights. No covariates were included in these tests. The number of days suspended is censored to 20 for suspensions that exceeded 20 days.
*** $p<.01$, ** $p<.05$, * $p<.1$

Table 8

School-Level Heterogeneity in Discipline Gaps in Fights Between Black and White Students

Outcome variable	Difference in days suspended (black minus white)		Black student received longer suspension	
	Full sample	First suspension of year	Full sample	First suspension of year
Grade levels				
Elementary (K-5)	0.012	-0.006	0.010**	0.008
	(0.016)	(0.017)	(0.005)	(0.006)
Middle (6-8)	0.079***	0.048***	0.018***	0.013**
	(0.015)	(0.017)	(0.004)	(0.006)
High (9-12)	0.064**	0.085***	0.023***	0.024**
	(0.029)	(0.029)	(0.008)	(0.011)
Administrators' races				
All white	0.058***	0.029**	0.017***	0.014***
	(0.014)	(0.014)	(0.004)	(0.005)
All non-white	0.038**	0.042*	0.016***	0.010
	(0.018)	(0.022)	(0.005)	(0.007)
Both white and non-white	0.056	0.036	0.021*	0.015
	(0.036)	(0.032)	(0.012)	(0.013)
Students' races				
More than 75% white	0.053***	0.033	0.010	0.012*
	(0.019)	(0.020)	(0.006)	(0.007)
25% to 75% white	0.053***	0.042*	0.020***	0.010
	(0.013)	(0.022)	(0.004)	(0.007)
Less than 25% white	0.053*	0.036	0.018**	0.015
	(0.030)	(0.032)	(0.009)	(0.013)
Urbanicity				
Urban	0.078***	0.059**	0.027***	0.013
	(0.025)	(0.028)	(0.007)	(0.009)
Rural	0.036*	0.027	0.013**	0.015**
	(0.019)	(0.020)	(0.006)	(0.008)
Other	0.056***	0.030**	0.015***	0.011**
	(0.014)	(0.015)	(0.004)	(0.005)
Student controls	No	No	No	No
School-Year FEs	Yes	Yes	Yes	Yes

Notes. Table shows results from models in Table 5, Panel A, Columns 1 (full sample) and 7 (first suspension of year) for difference in days suspended comparison and Table 6, Panel A, Columns 1 and 7 for which student received a longer suspension. The unit of observation is the infraction, and the sample is restricted to students who were suspended for fighting with a student of a different race. The "first suspension" sample restrictions apply to both students involved in the fight. These models do not contain covariates but do contain school-year fixed effects. Standard errors appear in parentheses and account for the clustering of students within schools. The number of days suspended is censored to 20 for suspensions that exceeded 20 days.

*** *p*<.01, ** *p*<.05, * *p*<.1

Table 9

Assignment of In-School and Out-of-School Suspensions for Fighting

Student 1	Student 2	Total number of fights	Both received in-school suspension	Both received out-of-school suspension	Student 1 out-of-school, Student 2 in-school	Student 1 in-school, Student 2 out-of-school	Z-score	p-value
Panel A: Full sample								
Black	White	20,142	27.8%	66.6%	3.1%	2.5%	3.553	0.000
Black	Black	81,719	29.1%	66.3%	--	--	--	--
White	White	20,144	24.2%	72.0%	--	--	--	--
Poor	Non-poor	10,190	28.4%	67.0%	2.5%	2.1%	1.777	0.076
Poor	Poor	52,446	30.2%	65.0%	--	--	--	--
Non-poor	Non-poor	3,167	28.7%	67.3%	--	--	--	--

Notes. The unit of observation is the infraction, and samples are restricted to fights between a student from the first group listed and a student from the second group listed. Significance tests assess whether the probability that only Group 1 was suspended out of school differs from the probability that only Group 2 was suspended out of school. Observation numbers indicate count of observed fights. No covariates were included in these tests. The number of days suspended is censored to 20 for suspensions that exceeded 20 days.

*** $p<.01$, ** $p<.05$, * $p<.1$

Figure 1. In-School Suspensions as Proportion of All Suspensions, by Race and Infraction Type
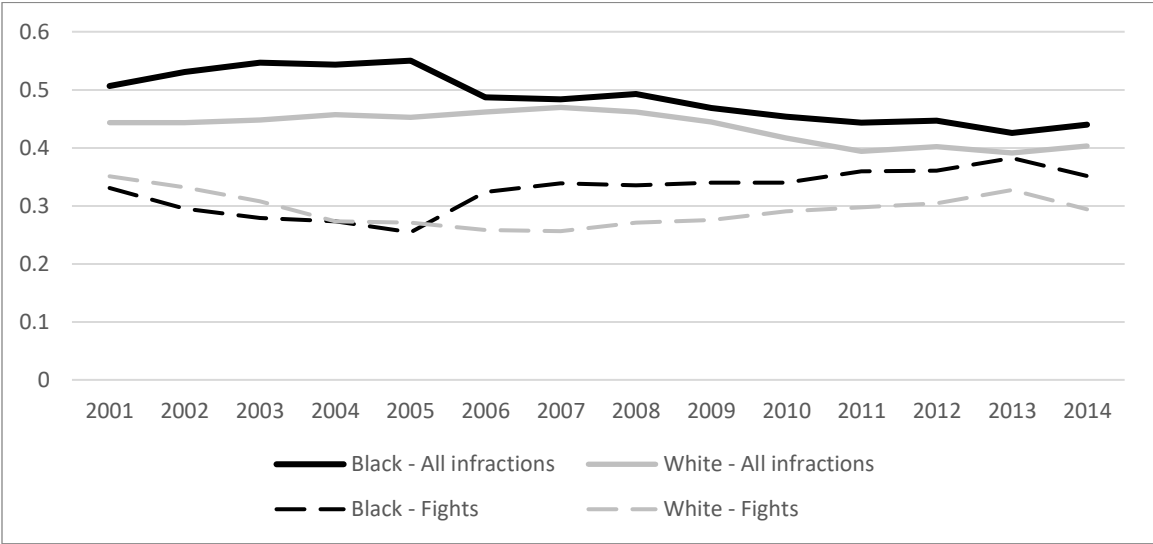
Figure 2. In-School Suspensions as Proportion of All Suspensions, by Poverty Status and Infraction Type
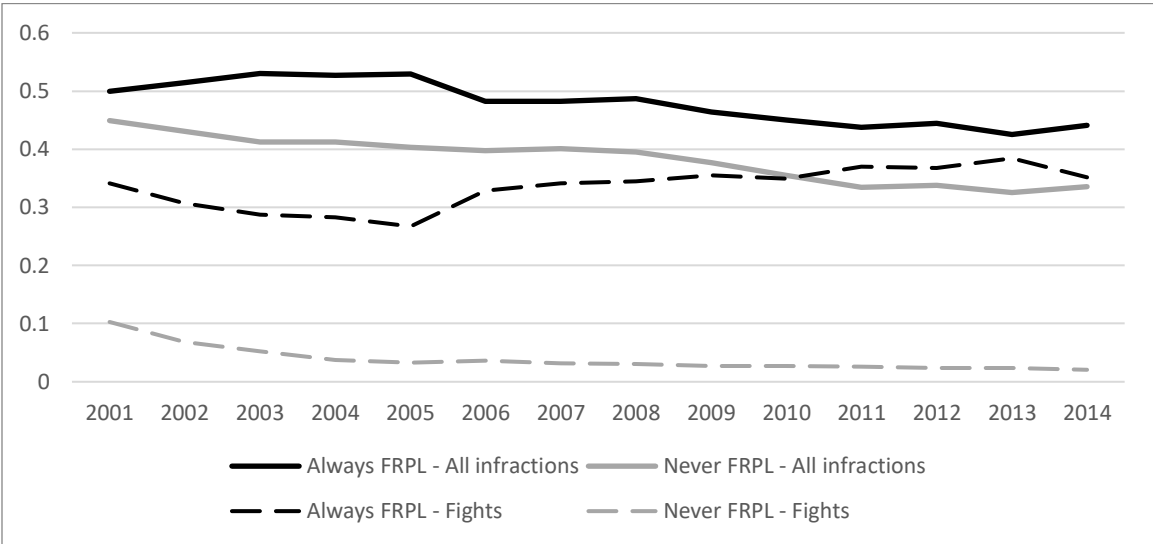
Figure 3. Decomposition of Black/White Gaps—Whether Suspended

Figure 4. Decomposition of Poor/Non-Poor Gaps—Whether Suspended
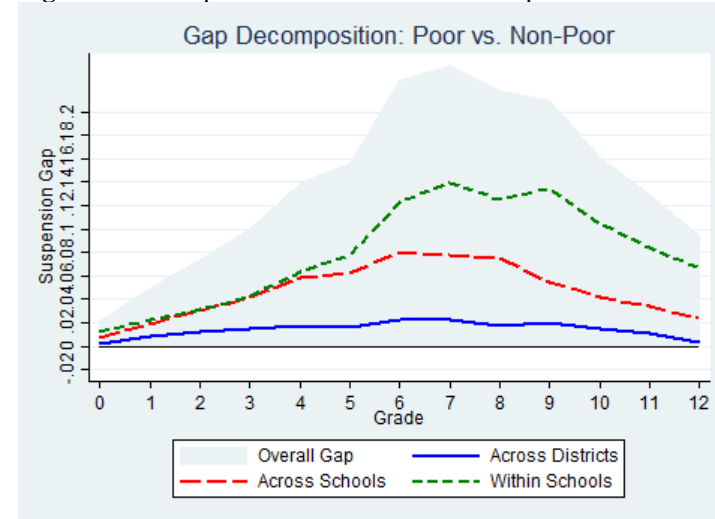
Figure 5. Decomposition of Black/White Gaps—Days Suspended
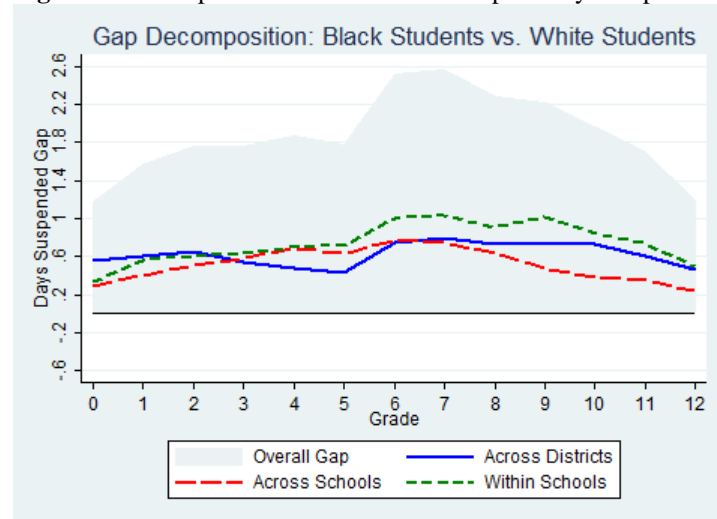
Figure 6. Decomposition of Poor/Non-Poor Gaps—Days Suspended
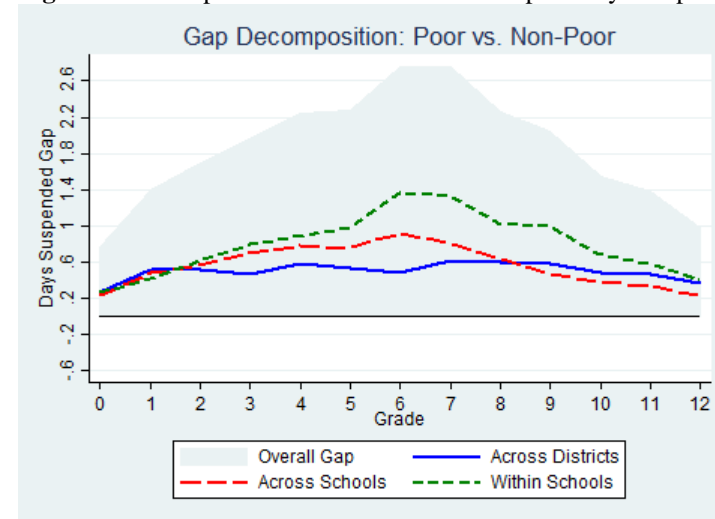
Figure 7. Frequency of Suspension Length Differences Between Black/White and Poor/Non-Poor Students in Fights

Panel A: Black/White Fights

Panel B: Poor/Non-Poor Fights
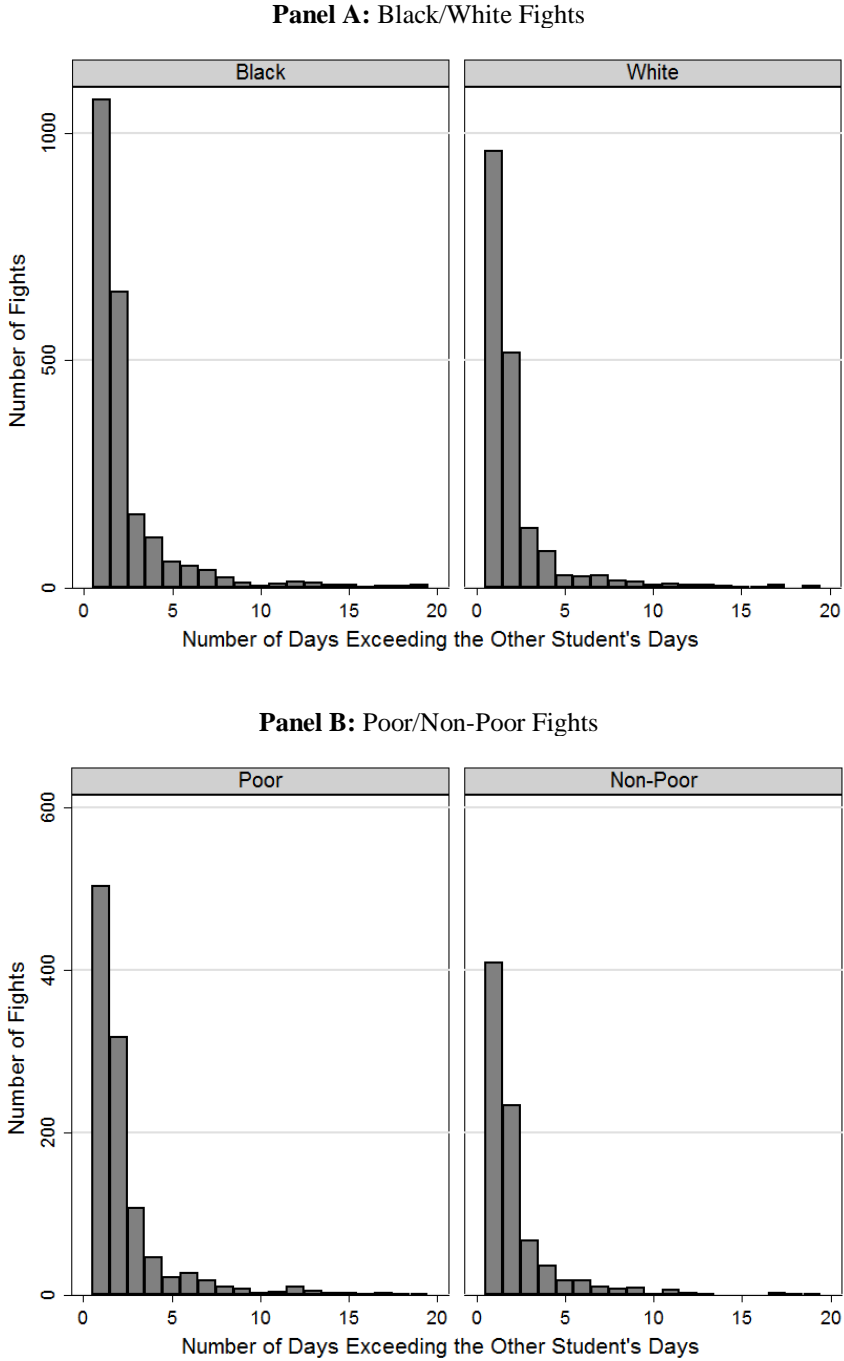
Appendix Table A1

Raw Gaps in Whether Suspended Across Districts, Across Schools, and Within Schools

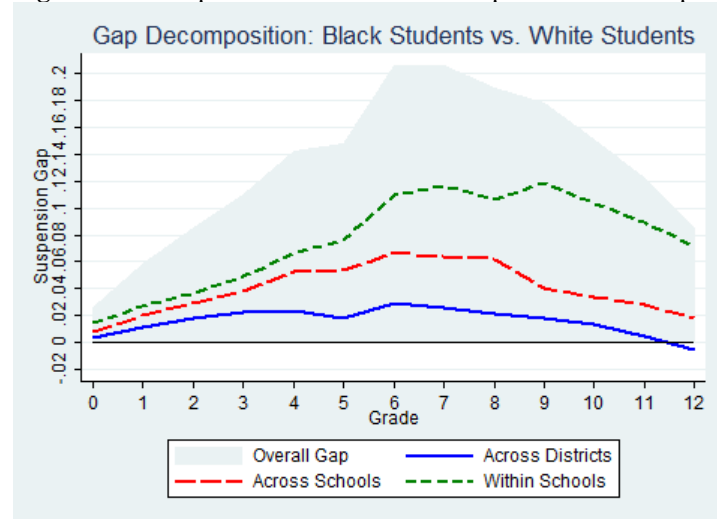| Grade	Source of gap	Black/White		Poor/Non-poor	
K	Across districts	0.004	13.5%	0.002	11.0%
	Across schools	0.008	29.9%	0.007	32.0%
	Within school	0.015	56.7%	0.012	57.1%
	Overall	0.026		0.022	
Grade 1	Across districts	0.011	19.4%	0.008	16.1%
	Across schools	0.020	34.1%	0.019	38.8%
	Within school	0.027	46.5%	0.022	45.1%
	Overall	0.058		0.050	
Grade 2	Across districts	0.019	22.1%	0.012	16.2%
	Across schools	0.029	34.6%	0.030	41.6%
	Within school	0.037	43.3%	0.031	42.3%
	Overall	0.085		0.073	
Grade 3	Across districts	0.022	20.2%	0.015	15.0%
	Across schools	0.038	35.0%	0.042	41.8%
	Within school	0.049	44.8%	0.043	43.2%
	Overall	0.109		0.100	
Grade 4	Across districts	0.023	16.4%	0.018	13.0%
	Across schools	0.052	37.0%	0.058	41.6%
	Within school	0.066	46.6%	0.063	45.4%
	Overall	0.142		0.139	
Grade 5	Across districts	0.018	12.2%	0.016	10.0%
	Across schools	0.053	36.1%	0.063	40.1%
	Within school	0.076	51.7%	0.078	50.0%
	Overall	0.148		0.156	
Grade 6	Across districts	0.029	14.1%	0.023	10.2%
	Across schools	0.067	32.6%	0.080	35.4%
	Within school	0.109	53.2%	0.123	54.4%
	Overall	0.205		0.226	
Grade 7	Across districts	0.026	12.5%	0.022	9.3%
	Across schools	0.063	30.8%	0.078	32.5%
	Within school	0.117	56.7%	0.140	58.2%
	Overall	0.206		0.240	
Grade 8	Across districts	0.021	11.2%	0.018	8.1%
	Across schools	0.062	32.8%	0.075	34.6%
	Within school	0.106	56.0%	0.125	57.4%
	Overall	0.189		0.218	
Grade 9	Across districts	0.018	10.3%	0.020	9.6%
	Across schools	0.040	22.8%	0.055	26.1%
	Within school	0.119	66.9%	0.135	64.3%
	Overall	0.177		0.209	
Grade 10	Across districts	0.014	9.3%	0.014	9.0%
	Across schools	0.034	22.3%	0.042	25.9%
	Within school	0.103	68.4%	0.105	65.1%
	Overall	0.151		0.161	
Grade 11	Across districts	0.005	3.7%	0.011	8.2%
	Across schools	0.028	23.3%	0.034	26.7%
	Within school	0.089	73.0%	0.084	65.1%
	Overall	0.122		0.129	
Grade 12	Across districts	-0.005	-6.1%	0.004	4.2%
	Across schools	0.018	21.5%	0.023	25.0%
	Within school	0.072	84.6%	0.066	70.8%
	Overall	0.085		0.094	

Appendix Table A2
Raw Gaps in Number of Days Suspended Across Districts, Across Schools, and Within Schools

Grade	Source of gap	Black/White		Poor/Non-poor	
K	Across districts	0.554	47.0%	0.263	34.7%
	Across schools	0.294	25.0%	0.231	30.5%
	Within school	0.329	28.0%	0.264	34.8%
	Overall	1.177		0.758	
Grade 1	Across districts	0.605	38.5%	0.509	36.4%
	Across schools	0.403	25.7%	0.477	34.1%
	Within school	0.562	35.8%	0.412	29.5%
	Overall	1.569		1.398	
Grade 2	Across districts	0.644	36.8%	0.521	30.6%
	Across schools	0.503	28.8%	0.563	33.1%
	Within school	0.603	34.4%	0.617	36.3%
	Overall	1.750		1.701	
Grade 3	Across districts	0.544	30.9%	0.469	24.0%
	Across schools	0.583	33.1%	0.696	35.5%
	Within school	0.632	35.9%	0.793	40.5%
	Overall	1.758		1.958	
Grade 4	Across districts	0.471	25.3%	0.580	25.8%
	Across schools	0.690	37.0%	0.778	34.6%
	Within school	0.702	37.7%	0.888	39.5%
	Overall	1.862		2.246	
Grade 5	Across districts	0.428	24.2%	0.539	23.6%
	Across schools	0.631	35.6%	0.755	33.1%
	Within school	0.713	40.3%	0.985	43.2%
	Overall	1.772		2.279	
Grade 6	Across districts	0.739	29.4%	0.478	17.3%
	Across schools	0.768	30.6%	0.919	33.2%
	Within school	1.005	40.0%	1.367	49.5%
	Overall	2.512		2.764	
Grade 7	Across districts	0.786	30.7%	0.615	22.4%
	Across schools	0.749	29.2%	0.808	29.4%
	Within school	1.029	40.1%	1.327	48.3%
	Overall	2.564		2.750	
Grade 8	Across districts	0.734	32.3%	0.593	26.3%
	Across schools	0.631	27.8%	0.644	28.6%
	Within school	0.905	39.9%	1.017	45.1%
	Overall	2.271		2.255	
Grade 9	Across districts	0.728	32.9%	0.590	28.8%
	Across schools	0.469	21.2%	0.462	22.6%
	Within school	1.017	45.9%	0.994	48.6%
	Overall	2.214		2.046	
Grade 10	Across districts	0.733	37.4%	0.480	31.1%
	Across schools	0.385	19.7%	0.381	24.7%
	Within school	0.842	43.0%	0.680	44.1%
	Overall	1.961		1.541	
Grade 11	Across districts	0.601	35.6%	0.466	33.7%
	Across schools	0.354	21.0%	0.332	24.0%
	Within school	0.734	43.5%	0.585	42.3%
	Overall	1.689		1.383	
Grade 12	Across districts	0.468	39.3%	0.362	36.7%
	Across schools	0.234	19.6%	0.227	23.0%
	Within school	0.491	41.1%	0.397	40.3%
	Overall	1.193		0.986	

Appendix Table A3

Predictors of Length of Suspension (in Days)—Out-of-School Suspensions Only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black	0.434***	0.108***			0.409***	0.102***	0.086***	0.308***	0.075***	0.061***
	(0.008)	(0.005)			(0.008)	(0.006)	(0.006)	(0.010)	(0.008)	(0.008)
Other race	0.201***	0.018			0.186***	0.015	0.011	0.129***	-0.015	-0.011
	(0.015)	(0.013)			(0.015)	(0.013)	(0.014)	(0.022)	(0.019)	(0.019)
Poor (always FRPL)			0.361***	0.078***	0.151***	0.040***	0.048***	0.188***	0.089***	0.078***
			(0.009)	(0.007)	(0.009)	(0.007)	(0.008)	(0.013)	(0.011)	(0.012)
Sometimes FRPL			0.360***	0.084***	0.211***	0.059***	0.066***	0.197***	0.087***	0.076***
			(0.009)	(0.007)	(0.009)	(0.007)	(0.008)	(0.013)	(0.011)	(0.012)
SPED								-0.198***	-0.172***	-0.182***
								(0.011)	(0.009)	(0.010)
Male								0.013*	0.017***	0.001
								(0.007)	(0.007)	(0.007)
Standardized math (t-1)								-0.051***	-0.008*	-0.006
								(0.005)	(0.005)	(0.005)
Standardized ELA (t-1)								-0.058***	-0.029***	-0.024***
								(0.005)	(0.004)	(0.004)
Suspended in prior year							0.155***			0.153***
							(0.005)			(0.006)
Constant	1.903***	2.319***	1.981***	2.324***	1.837***	2.280***	2.166***	1.756***	2.251***	2.183***
	(0.031)	(0.007)	(0.032)	(0.008)	(0.032)	(0.008)	(0.010)	(0.349)	(0.014)	(0.015)
Observations	2,001,819	2,001,819	2,003,209	2,003,209	2,001,819	2,001,819	1,781,445	861,056	861,056	844,312
R-squared	0.037	0.180	0.033	0.180	0.038	0.180	0.179	0.032	0.177	0.177
Year FE	Yes	No	Yes	No	Yes	No	No	Yes	No	No
Grade FE	Yes	No	Yes	No	Yes	No	No	Yes	No	No
SGY FEs	No	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes

Notes. Table is constructed identically to Table 3 but only codes out-of-school suspensions as suspensions (and disregards in-school suspensions). The unit of observation is the infraction, and the sample is restricted to students who were suspended. All models contain infraction fixed effects. Omitted reference groups are white students, non-poor students (students never eligible for free or reduced-price lunch), and females. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within school-grade-year. "SGY FEs" refers to school-grade-year fixed effects.

*** $p<.01$, ** $p<.05$, * $p<.1$

Appendix Table A4

Predictors of Length of Suspension for First Offense of Year (in Days)—Out-of-School Suspensions Only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black	0.413***	0.056***			0.391***	0.052***	0.048***	0.282***	0.032***	0.029**
	(0.008)	(0.007)			(0.009)	(0.008)	(0.008)	(0.011)	(0.012)	(0.012)
Other race	0.221***	0.011			0.208***	0.009	0.007	0.167***	0.016	0.015
	(0.018)	(0.017)			(0.018)	(0.017)	(0.019)	(0.027)	(0.027)	(0.027)
Poor (always FRPL)			0.320***	0.043***	0.120***	0.025***	0.039***	0.135***	0.044***	0.046***
			(0.010)	(0.009)	(0.010)	(0.009)	(0.010)	(0.016)	(0.015)	(0.016)
Sometimes FRPL			0.324***	0.051***	0.187***	0.039***	0.052***	0.152***	0.043***	0.044***
			(0.011)	(0.009)	(0.010)	(0.009)	(0.010)	(0.016)	(0.015)	(0.016)
SPED								-0.092***	-0.072***	-0.076***
								(0.012)	(0.012)	(0.013)
Male								-0.037***	-0.036***	-0.040***
								(0.009)	(0.009)	(0.009)
Standardized math (*t*-1)								-0.041***	-0.002	-0.001
								(0.006)	(0.006)	(0.006)
Standardized ELA (*t*-1)								-0.042***	-0.013**	-0.012*
								(0.006)	(0.006)	(0.006)
Suspended in prior Year							0.031***			0.028***
							(0.006)			(0.009)
Constant	1.678***	2.103***	1.769***	2.100***	1.630***	2.079***	2.033***	1.710***	2.086***	2.073***
	(0.033)	(0.009)	(0.034)	(0.010)	(0.034)	(0.011)	(0.013)	(0.470)	(0.019)	(0.020)
Observations	833,129	833,129	833,714	833,714	833,129	833,129	730,433	355,204	355,204	347,220
R-squared	0.050	0.262	0.045	0.262	0.050	0.262	0.261	0.042	0.265	0.264
Year FE	Yes	No	Yes	No	Yes	No	No	Yes	No	No
Grade FE	Yes	No	Yes	No	Yes	No	No	Yes	No	No
School-Grade-Year FE	No	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes

Notes. Table is constructed identically to Table 4 but only codes out-of-school suspensions as suspensions (and disregards in-school suspensions). The unit of observation is the infraction, and the sample is restricted to students who were suspended. All models contain infraction fixed effects. Omitted reference groups are white students, non-poor students (students never eligible for free or reduced-price lunch), and females. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within school-grade-year. "SGY FEs" refers to school-grade-year fixed effects.

*** $p<.01$, ** $p<.05$, * $p<.1$

Appendix Table A5

Discipline Gaps in Fights Between Black and White Students

	Number of days suspended				Whether received longer suspension than peer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black student	0.054***	0.052***	0.048***	0.041***	0.017***	0.016***	0.013***	0.011***
	(0.010)	(0.011)	(0.013)	(0.012)	(0.003)	(0.003)	(0.003)	(0.004)
Student characteristics								
Number of prior fights	0.100***	0.129***	0.105***	0.129***	0.034***	0.040***	0.034***	0.040***
	(0.020)	(0.013)	(0.020)	(0.013)	(0.003)	(0.004)	(0.003)	(0.004)
Poor			-0.039	-0.033*			-0.005	-0.005
			(0.028)	(0.019)			(0.003)	(0.005)
Non-poor			-0.092**	-0.105***			-0.025***	-0.035***
			(0.041)	(0.025)			(0.005)	(0.007)
SPED			-0.125***	-0.069***			-0.009**	-0.026***
			(0.034)	(0.023)			(0.004)	(0.007)
Male			-0.103***	-0.02			-0.012***	0.003
			(0.039)	(0.039)			(0.004)	(0.012)
Standardized math score (*t*-1)			0.026	-0.007			-0.003	-0.007
			(0.024)	(0.015)			(0.003)	(0.005)
Standardized ELA score (*t*-1)			-0.064***	-0.017			-0.003	0
			(0.024)	(0.015)			(0.003)	(0.005)
Constant	2.892***	2.853***	2.992***	2.890***	0.046***	0.038***	0.058***	0.045***
	(0.027)	(0.018)	(0.043)	(0.040)	(0.004)	(0.005)	(0.006)	(0.011)
Observations	40,284	40,284	40,284	40,284	40,284	40,284	40,284	40,284
R-squared	0.008	0.009	0.014	0.016	0.018	0.019	0.021	0.021
School-Year FEs	Yes	No	Yes	No	Yes	No	Yes	No
Fight occurrence FEs	No	Yes	No	Yes	No	Yes	No	Yes

Notes. The unit of observation is the infraction, and the sample is restricted to students who were suspended for fighting (with a student of a different race). The number of prior fights refers to the number of fights for which the student had been suspended earlier in the same school year. Omitted reference groups are white students, sometimes-poor students, students without disabilities, and females. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within schools.

*** *p*<.01, ** *p*<.05, * *p*<.1

Appendix Table A6

Discipline Gaps in Fights Between Poor and Non-Poor Students

	Number of days suspended				Whether received longer suspension than peer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Poor student	0.071***	0.067***	0.049***	0.059***	0.023***	0.023***	0.021***	0.020***
	(0.013)	(0.013)	(0.018)	(0.015)	(0.004)	(0.004)	(0.005)	(0.005)
Student characteristics								
Number of prior fights	0.066***	0.121***	0.069***	0.126***	0.038***	0.043***	0.037***	0.044***
	(0.025)	(0.020)	(0.026)	(0.019)	(0.004)	(0.006)	(0.004)	(0.006)
Black			0.091**	0.065***			0.011**	0.020**
			(0.044)	(0.025)			(0.005)	(0.008)
Other race			0.099	0.019			0.021	0.015
			(0.118)	(0.060)			(0.013)	(0.020)
SPED			-0.087	-0.135***			-0.004	-0.023**
			(0.055)	(0.034)			(0.006)	(0.010)
Male			-0.118**	-0.021			0.003	0.005
			(0.052)	(0.049)			(0.005)	(0.016)
Standardized math score (t-1)			-0.053	-0.003			-0.007	-0.01
			(0.035)	(0.020)			(0.005)	(0.007)
Standardized ELA score (t-1)			0.018	-0.004			0.001	0.005
			(0.035)	(0.018)			(0.005)	(0.007)
Constant	3.075***	3.004***	3.093***	2.975***	0.032***	0.025***	0.021***	0.006
	(0.034)	(0.027)	(0.055)	(0.048)	(0.006)	(0.008)	(0.007)	(0.015)
Observations	20,380	20,380	20,380	20,380	20,380	20,380	20,380	20,380
R-squared	0.007	0.007	0.011	0.011	0.009	0.009	0.011	0.011
School-Year FEs	Yes	No	Yes	No	Yes	No	Yes	No
Fight occurrence FEs	No	Yes	No	Yes	No	Yes	No	Yes

Notes. The unit of observation is the infraction, and the sample is restricted to students who were suspended for fighting (with a student of a different poverty status). The number of prior fights refers to the number of fights for which the student had been suspended earlier in the same school year. Omitted reference groups are white students, students without disabilities, and females. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within schools.

*** $p<.01$, ** $p<.05$, * $p<.1$

Appendix Table A7

Gaps in Days Suspended for Black/White and Poor/Non-Poor Fights--Out-of-School Suspensions Only

	Full sample		First fight of year		First fight ever		First suspension of year		First suspension ever	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Race comparison										
Black student	0.061***	0.059***	0.049***	0.057***	0.049***	0.040*	0.051***	0.065***	0.066***	0.059**
	(0.011)	(0.015)	(0.011)	(0.015)	(0.015)	(0.021)	(0.012)	(0.019)	(0.020)	(0.029)
Constant	2.954***	3.059***	3.097***	3.220***	3.076***	3.153***	2.967***	3.060***	2.968***	2.974***
	(0.032)	(0.052)	(0.005)	(0.045)	(0.062)	(0.095)	(0.006)	(0.053)	(0.075)	(0.125)
Observations	28,268	28,268	21,104	21,104	11,640	11,640	14,928	14,928	7,120	7,120
R-squared	0.008	0.014	0.008	0.008	0.007	0.007	0.009	0.008	0.007	0.007
Panel B: Poverty comparison										
Poor student	0.082***	0.058**	0.039**	0.014	0.030	0.008	0.077***	0.024	0.105***	0.086**
	(0.021)	(0.027)	(0.019)	(0.027)	(0.027)	(0.033)	(0.024)	(0.031)	(0.029)	(0.036)
Constant	3.213***	3.262***	3.284***	3.318***	3.335***	3.449***	3.142***	3.155***	3.299***	3.368***
	(0.065)	(0.109)	(0.010)	(0.096)	(0.049)	(0.120)	(0.012)	(0.118)	(0.067)	(0.166)
Observations	15,002	15,002	11,706	11,706	8,564	8,564	8,530	8,530	5,562	5,562
R-squared	0.007	0.011	0.007	0.007	0.006	0.006	0.008	0.012	0.008	0.011
Controls for student characteristics	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls for number of years in data	No	No	No	No	Yes	Yes	No	No	Yes	Yes

Notes. Table is constructed identically to Table 5 but only codes out-of-school suspensions as suspensions (and disregards in-school suspensions). The unit of observation is the infraction, and the sample is restricted to students who were suspended for fighting with a student of a different race (Panel A) or poverty status (Panel B). The "first fight" and "first suspension" sample restrictions apply to both students involved in the fight. For example, the "first fight of year" columns restrict the sample to fights between two students who had not been suspended for a fight earlier in that school year. All models contain school-year fixed effects. Student controls consist of FRPL status (Panel A only), black and other race (Panel B only), special education status, gender, and math and ELA scores from the prior year. The reference group for "Black student" is white students. The reference group for "Poor student" is non-poor students. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within schools.

*** $p<.01$, ** $p<.05$, * $p<.1$

Appendix Table A8

Who Received Longer Suspensions for Black/White and Poor/Non-Poor Fights--Out-of-School Suspensions Only

	Full sample		First fight of year		First fight ever		First suspension of year		First suspension ever	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Race comparison										
Black student	0.019***	0.015***	0.016***	0.015***	0.017***	0.015***	0.020***	0.017***	0.026***	0.024***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)	(0.004)	(0.005)	(0.006)	(0.007)
Constant	0.037***	0.049***	0.072***	0.083***	0.071***	0.088***	0.070***	0.081***	0.056***	0.063***
	(0.005)	(0.006)	(0.002)	(0.005)	(0.005)	(0.010)	(0.002)	(0.006)	(0.007)	(0.014)
Observations	28,268	28,268	21,104	21,104	11,640	11,640	14,928	14,928	7,120	7,120
R-squared	0.018	0.021	0.007	0.01	0.007	0.01	0.008	0.011	0.01	0.011
Panel B: Poverty comparison										
Poor student	0.024***	0.023***	0.018***	0.016***	0.014**	0.013*	0.029***	0.026***	0.033***	0.031***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Constant	0.036***	0.020**	0.077***	0.062***	0.078***	0.056***	0.072***	0.052***	0.062***	0.048***
	(0.007)	(0.009)	(0.003)	(0.007)	(0.005)	(0.011)	(0.003)	(0.009)	(0.007)	(0.014)
Observations	15,002	15,002	11,706	11,706	8,564	8,564	8,530	8,530	5,562	5,562
R-squared	0.009	0.011	0.008	0.008	0.007	0.007	0.008	0.010	0.008	0.010
Controls for student characteristics	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls for number of years in data	No	No	No	No	Yes	Yes	No	No	Yes	Yes

Notes. Table is constructed identically to Table 6 but only codes out-of-school suspensions as suspensions (and disregards in-school suspensions). The unit of observation is the infraction, and the sample is restricted to students who were suspended for fighting with a student of a different race (Panel A) or poverty status (Panel B). The "first fight" and "first suspension" sample restrictions apply to both students involved in the fight. For example, the "first fight of year" columns restrict the sample to fights between two students who had not been suspended for a fight earlier in that school year. All models contain school-year fixed effects (and not fight occurrence fixed effects). Student controls consist of FRPL status (Panel A only), black and other race (Panel B only), special education status, gender, and math and ELA scores from the prior year. The reference group for "Black student" is white students. The reference group for "Poor student" is non-poor students. The number of days suspended is censored to 20 for suspensions that exceeded 20 days. Standard errors appear in parentheses and account for the clustering of students within schools.

*** $p<.01$, ** $p<.05$, * $p<.1$

Appendix Table A9

Robustness of Fight Suspension Disparities to Alternate Censoring of Length of Individual Suspensions and Gaps

	Max. suspension length: 5 days		Max. suspension length: 10 days		Max. suspension length: 20 days	
	Full sample	First suspension of year	Full sample	First suspension of year	Full sample	First suspension of year
Panel A: Race comparison (black minus white)						
Max. gap in suspension length: 3 days	0.020***	0.018***	0.020***	0.017***	0.017***	0.016***
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
Max. gap in suspension length: 5 days	0.027***	0.023***	0.029***	0.026***	0.026***	0.027***
	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)
Max. gap in suspension length: No max.	0.027***	0.023***	0.046***	0.032***	0.054***	0.041***
	(0.005)	(0.006)	(0.007)	(0.008)	(0.010)	(0.011)
Panel B: Poverty comparison (poor minus non-poor)						
Max. gap in suspension length: 3 days	0.030***	0.024***	0.028***	0.025***	0.026***	0.024***
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
Max. gap in suspension length: 5 days	0.042***	0.036***	0.042***	0.039***	0.040***	0.038***
	(0.007)	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)
Max. gap in suspension length: No max.	0.043***	0.037***	0.056***	0.050***	0.071***	0.072***
	(0.007)	(0.009)	(0.011)	(0.013)	(0.013)	(0.015)

Notes. Table shows sensitivity of results from Table 5. Panel A, Column 1 (full sample) and Column 7 (first suspension of year) to censoring at different levels. "Max. suspension length" indicates the number of days to which we censored each student's suspension length if it exceeded 20 days. "Max. gap in suspension length" indicates the number of days to which we censored the difference between the fighting students' suspensions lengths. Each estimate comes from a distinct regression. The unit of observation is the infraction, and the sample is restricted to students who were suspended for fighting with a student of a different race (Panel A) or poverty status (Panel B). The "first suspension" sample restrictions apply to both students involved in the fight. These models do not contain covariates but do contain school-year fixed effects. Standard errors appear in parentheses and account for the clustering of students within schools.
*** $p<.01$, ** $p<.05$, * $p<.1$

Appendix Table A10

Who Received Longer Suspensions in Fights Between Students of Various Subgroups--First Suspensions of Year

Student 1	Student 2	Total number of fights	Same suspension length	Student 1 suspended longer	Student 2 suspended longer	Z-score	p-value
Subgroups							
Black	White	10,746	83.4%	9.1%	7.5%	4.253	0.000
Black	Black	43,967	80.4%	--	--	--	--
White	White	12,448	85.5%	--	--	--	--
Poor	NP	5,930	84.0%	9.3%	6.7%	5.115	0.000
Poor	Poor	28,784	81.8%	--	--	--	--
NP	NP	2,014	87.7%	--	--	--	--
Male	Female	9,022	73.5%	13.3%	13.3%	-0.066	0.947
Male	Male	40,906	82.9%	--	--	--	--
Female	Female	20,272	83.6%	--	--	--	--
Low scorer	High scorer	1,863	85.7%	7.7%	6.5%	1.400	0.162
Low scorer	Low scorer	5,204	81.8%	--	--	--	--
High scorer	High scorer	1,032	87.5%	--	--	--	--
Subgroup interactions with race							
Black, Poor	White, Poor	3,153	84.0%	8.5%	7.4%	1.627	0.104
Black, NP	White, NP	277	87.0%	7.2%	5.8%	0.689	0.491
Black, Poor	White, NP	1,564	83.5%	10.2%	6.3%	4.030	0.000
Black, NP	White, Poor	185	85.9%	6.5%	7.6%	-0.407	0.684
Black, Male	White, Male	7,891	84.9%	8.2%	6.9%	3.135	0.002
Black, Female	White, Female	1,713	84.2%	8.8%	7.0%	1.902	0.057
Black, Male	White, Female	486	72.0%	16.5%	11.5%	2.219	0.026
Black, Female	White, Male	656	71.3%	15.4%	13.3%	1.103	0.270
Black, High scorer	White, High scorer	215	89.8%	5.1%	5.1%	0.000	1.000
Black, Low scorer	White, Low scorer	477	83.2%	10.7%	6.1%	2.570	0.010
Black, High scorer	White, Low scorer	67	86.6%	10.4%	3.0%	1.726	0.084
Black, Low scorer	White, High scorer	349	86.5%	8.9%	4.6%	2.266	0.023
Subgroup interactions with poverty							
Poor, Male	NP, Male	4,024	84.8%	8.8%	6.4%	4.166	0.000
Poor, Female	NP, Female	1,309	85.2%	9.1%	5.7%	3.283	0.001
Poor, Male	NP, Female	276	76.1%	13.4%	10.5%	1.049	0.294
Poor, Female	NP, Male	321	76.3%	11.8%	11.8%	0.000	1.000
Poor, High scorer	NP, High scorer	150	88.0%	8.7%	3.3%	1.945	0.052
Poor, Low scorer	NP, Low scorer	169	82.2%	8.3%	9.5%	-0.383	0.702
Poor, High scorer	NP, Low scorer	54	87.0%	11.1%	1.9%	1.954	0.051
Poor, Low scorer	NP, High scorer	125	84.8%	11.2%	4.0%	2.148	0.032

Notes. Table is constructed identically to Table 7 but restricts sample to fights between students who had not previously been suspended that year. The unit of observation is the infraction, and samples are restricted to fights between a student from the first group listed and a student from the second group listed. "Poor" indicates that the student appeared eligible for free or reduced-price lunch in all years of our data. "NP" indicates that the student never appeared eligible for free or reduced-price lunch in our data. "High scorer" is defined as a student in the top quintile of average performance, where average performance is calculated as the mean score across all tested subjects. "Low scorer" is defined as a student in the bottom quintile of performance across all tested subjects. Significance tests assess whether the probability that Group 1 received a longer punishment differs from the probability that Group 2 received a longer punishment. Observation numbers indicate count of observed fights. No covariates were included in these tests. The number of days suspended is censored to 20 for suspensions that exceeded 20 days.

*** $p<.01$, ** $p<.05$, * $p<.1$

Table A11
School-Level Heterogeneity in Discipline Gaps in Fights Between Poor and Non-Poor Students

Outcome variable	Difference in days suspended (poor minus non-poor)		Poor student received longer suspension	
	Full sample	First suspension of year	Full sample	First suspension of year
Grade levels				
Elementary (K-5)	0.023	0.027	0.022**	0.011
	(0.021)	(0.024)	(0.009)	(0.010)
Middle (6-8)	0.882	0.061	0.032***	0.026***
	(0.797)	(0.063)	(0.006)	(0.007)
High (9-12)	0.068	0.041	0.022**	0.025**
	(0.042)	(0.040)	(0.009)	(0.012)
Administrators' races				
All white	0.588	0.016	0.020***	0.017***
	(0.539)	(0.043)	(0.005)	(0.006)
All non-white	0.101***	0.148***	0.033***	0.040***
	(0.036)	(0.047)	(0.009)	(0.012)
Both white and non-white	0.136**	0.051	0.039*	0.017
	(0.057)	(0.059)	(0.021)	(0.024)
Students' poverty status				
More than 75% FRPL	0.041	0.043	0.006	0.011
	(0.057)	(0.072)	(0.020)	(0.023)
25% to 75% FRPL	0.048	0.062*	0.023**	0.013
	(0.038)	(0.035)	(0.011)	(0.013)
Less than 25% FRPL	0.121***	0.077**	0.036***	0.031***
	(0.024)	(0.032)	(0.007)	(0.008)
Urbanicity				
Urban	0.001	-0.003	0.017*	0.010
	(0.083)	(0.127)	(0.009)	(0.012)
Rural	0.082***	0.069***	0.028***	0.033***
	(0.022)	(0.022)	(0.009)	(0.011)
Other	0.104***	0.063**	0.027***	0.021***
	(0.026)	(0.025)	(0.006)	(0.007)
Student controls	No	No	No	No
School-Year FEs	Yes	Yes	Yes	Yes

Notes. Table shows results from models in Table 5, Panel B, Columns 1 (full sample) and 7 (first suspension of year) for difference in days suspended comparison and Table 6, Panel B, Columns 1 and 7 for which student received a longer suspension. The unit of observation is the infraction, and the sample is restricted to students who were suspended for fighting with a student of a different race. The "first suspension" sample restrictions apply to both students involved in the fight. These models do not contain covariates but do contain school-year fixed effects. Standard errors appear in parentheses and account for the clustering of students within schools. The number of days suspended is censored to 20 for suspensions that exceeded 20 days.
*** $p<.01$, ** $p<.05$, * $p<.1$

Appendix Table A12

Assignment of In-School and Out-of-School Suspensions for Fighting--First Suspensions of Year

Student 1	Student 2	Total number of fights	Both received in-school suspension	Both received out-of-school suspension	Student 1 out-of-school, Student 2 in-school	Student 1 in-school, Student 2 out-of-school	Z-score	p-value
Black	White	10,746	29.9%	65.9%	2.4%	1.9%	2.406	0.016
Black	Black	43,967	31.1%	65.4%	--	--	--	--
White	White	12,448	25.5%	71.4%	--	--	--	--
Poor	Non-poor	5,930	30.4%	66.1%	2.1%	1.4%	2.952	0.003
Poor	Poor	28784	31.7%	64.7%	--	--	--	--
Non-poor	Non-poor	2014	31.2%	65.6%	--	--	--	--

Notes. The unit of observation is the infraction, and samples are restricted to fights between a student from the first group listed and a student from the second group listed. Significance tests assess whether the probability that only Group 1 was suspended out of school differs from the probability that only Group 2 was suspended out of school. Observation numbers indicate count of observed fights. No covariates were included in these tests. The number of days suspended is censored to 20 for suspensions that exceeded 20 days.

*** $p<.01$, ** $p<.05$, * $p<.1$

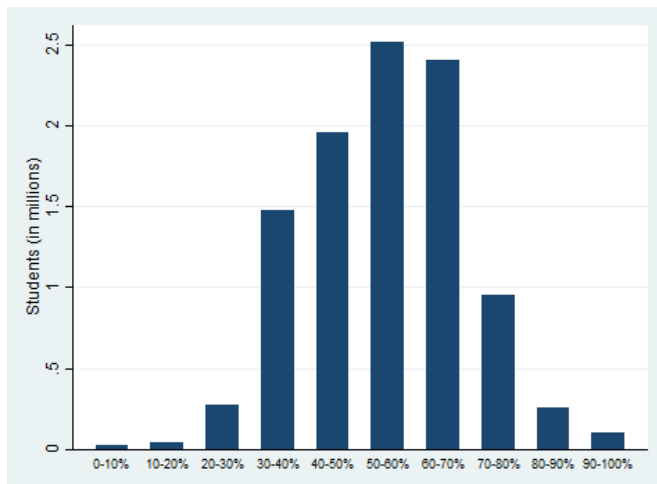**Figure A1**. Student Enrollment by District's Free Lunch Eligibility

Figure A2. Student Enrollment by District's Black Student Population
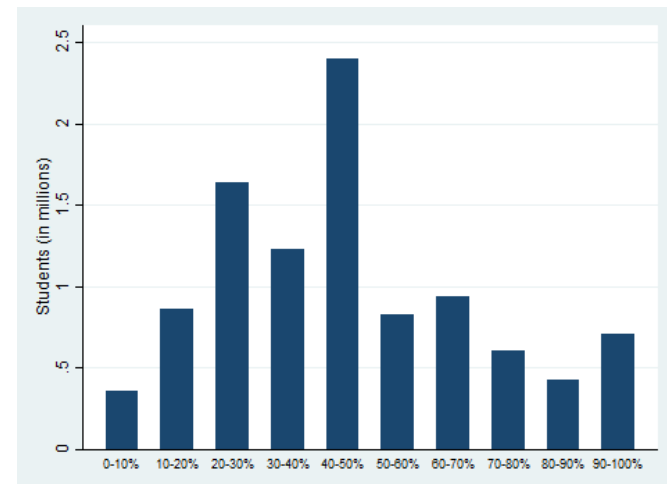

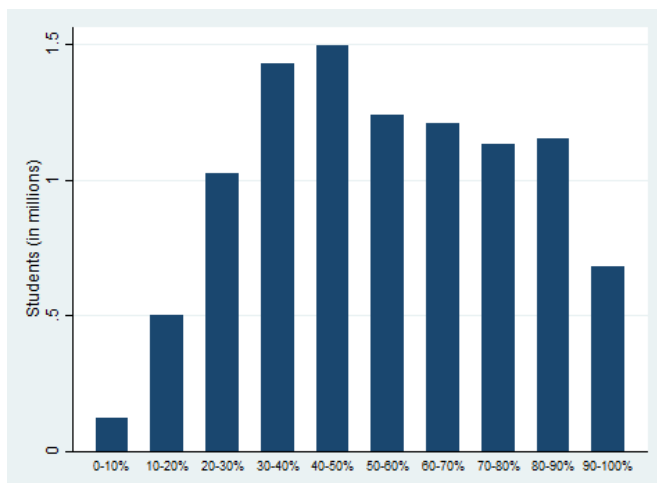**Figure A3**. Student Enrollment by School's Free Lunch Eligibility

Figure A4. Student Enrollment by School's Black Student Population