

# Understanding the Black-White School Discipline Gap

Josh Kinsler

February 3, 2009

## Abstract

At state and national levels, black students are more likely to be suspended from school, and conditional on misbehavior, receive stiffer penalties when compared with white students. Racial bias is often cited as a primary contributor to these gaps. Using infraction data from North Carolina, I investigate gaps in punishment within and across schools, and explore how student-teacher and student-principal race interactions affect discipline. I find a significant statewide gap in discipline that is largely generated by cross-school variation in punishment. In addition, there is little evidence that black students are treated differentially according to teacher or principal race.

## 1 Introduction

Out of school suspension is a severe, yet common form of student discipline that prohibits students from entering school grounds or participating in any school-related activities for a school-specified number of days. During the 2000 school year, black students comprised 17% of the U.S. student population but accounted for 34% of out of school suspensions.<sup>1</sup> The proportion of black students receiving at least one day of suspension increased by 120%

---

<sup>1</sup>U.S. Department of Education's Office for Civil Rights Elementary and Secondary School Survey. The data is available online at [www.ed.gov/ocr/](http://www.ed.gov/ocr/)

from 1972 to 2000, while for white students, the same measure increased only 64%.<sup>2</sup> The overrepresentation of black students in school suspension rolls is certainly a concern, but equally as troubling is the fact that conditional on a discipline referral, black students tend to receive stiffer penalties than white students.<sup>3</sup> While varying rates of misbehavior across black and white students may explain some of the racial gap in the incidence of suspension, there is no obvious explanation for the gap in discipline conditional on student behavior.

Public concern regarding racial inequality in school discipline stems from the perceived relationship between suspension and a host of undesirable youth outcomes. Numerous studies from the education literature document a strong negative correlation between suspension and student achievement.<sup>4</sup> While there is little evidence of a direct causal effect, it is widely believed that diminished learning opportunities and a weakening of the student-school bond negatively impact student performance. Dropout rates are also consistently higher for suspended students and some research indicates that schools actually use suspensions to push troublesome students out of school.<sup>5</sup> Not only are students excluded from classroom learning while suspended, they may be unsupervised at home and thus more likely to get in trouble in the community.<sup>6</sup> This provides another channel through which suspension can lead to dropout, as recent evidence suggests that incarceration has a negative effect on high school

---

<sup>2</sup>The year 2000 data come from the same source cited in note(1). The 1972 data come from the Children's Defense Fund (1975).

<sup>3</sup>Braden and Shaw (1990) and Gregory (1995) illustrate this pattern using the application of corporal punishment while McFadden and Marsh (1992) show that black students are more likely to be suspended conditional on committing any offense. This pattern of discipline is similar to evidence from the law and economics literature that indicates that black defendants receive significantly longer sentences and have higher bail amounts set than white defendants who commit the same crime. Examples include Steffensmeier and Demuth (2000), Mustard (2001), Everett and Wojtkiewicz (2002), Schanzenbach (2005), and Ayres and Waldfogel (1994).

<sup>4</sup>See Costenbader and Markson (1998) and Mendez (2003) as examples.

<sup>5</sup>Mendez et al. (2002) shows a strong negative correlation between suspensions in middle school and on-time graduation from high school and Bowditch (1993) provides qualitative evidence that schools use suspensions to filter students.

<sup>6</sup>Jacob and Lefgren (2003) find that when students are out of school property crime rates are significantly higher while violent crime rates are significantly lower. Given the higher prevalence of property crimes among juveniles, being out of school is likely to lead to increased arrest rates. Using survey data, Costenbader and Markson (1998) find that students who have been suspended are five times as likely to have been arrested when compared with non-suspended students.

completion.<sup>7</sup> The harmful consequences associated with suspension: reduced achievement, dropout, and delinquency have all been linked to future economic struggles, particularly in the labor market.<sup>8</sup>

Even if the relationships between suspension and future outcomes cited above are not entirely causal, racial bias in the application of discipline within public schools remains important. First, critics of school discipline often cite racial inequality as one of the primary reasons why out of school suspensions should be eliminated or at least curtailed.<sup>9</sup> While this paper does not take a stand on whether there should be more or less discipline, removing race from the equation may help clarify the issue. Second, racial bias in discipline may be important even if there are no direct effects of discipline on student outcomes. Evidence of racial bias by schools in one dimension (school discipline) can easily lead parents and students to believe that there exists bias in other dimensions (funding, teacher assignments, etc.). Thus, just the perception of bias in discipline can lead to an unraveling of the relationship between communities and schools.

Besides overt racial bias on the part of administrators, a wide variety of explanations for the overrepresentation of black students in suspension data have been offered. The most obvious, though often discounted theory, is higher levels of disruptive behavior among black students.<sup>10</sup> More unsettling theories have also been suggested such as efforts by schools to re-segregate or a lack of understanding of black youth behavior.<sup>11</sup> In order to determine if racial bias is the source of the racial disparity in suspension, as opposed to say varying behavior, detailed micro-data on student conduct and school discipline is needed.

---

<sup>7</sup>Hjalmarsson (2008)

<sup>8</sup>Neal and Johnson (1996) find that most of the observed gap in wages between black and white adults disappear after controlling for 8th grade test scores. Currie and Thomas (1999) show that test scores measured at age 7 are strongly correlated with adult labor market outcomes. In terms of dropout, Lochner and Moretti (2004) illustrate that graduating from high school significantly reduces the likelihood of future incarceration. Waldfogel (1994) and Grogger (1995) provide evidence that a criminal record is detrimental to labor market success.

<sup>9</sup>See Skiba (2000) as an example.

<sup>10</sup>In an analysis of why students get suspended, McCarthy and Hoge (1987) notes that the best predictor of punishment is always self-reports of misbehavior.

<sup>11</sup>See Eitle and Eitle (2004) and Townsend (2000).

Yet, previous empirical research regarding the racial gap in discipline is largely based on national, state, or district level data.<sup>12</sup> While useful for documenting overall disparities in discipline between black and white students, aggregate data is not appropriate for testing theories of racial bias. To be precise, racial bias in this context refers to differential punishment for otherwise identical black and white students who commit the same offense. If two students do not attend the same school, then they are not identical, even if they have common values for all other observed characteristics. The question arises, how much of the aggregate discipline gap can be explained by cross-school variation in punishment?<sup>13</sup>

Using a unique disciplinary data set from North Carolina, I am able to examine punishment decisions at the incident level both within and across schools. I start by considering the probability that a public school student receives at least one out of school suspension during the year, independent of any observed behavior. This is akin to previous work that has documented the overrepresentation of black students in suspension data, except that I estimate the gap within schools. Consistent with the previous literature, I find that black students are significantly more likely to be suspended. Again, it is difficult to conclude anything about racial bias from this analysis since I am not conditioning on actual classroom behavior. However, if racial bias is at the heart of the unconditional suspension disparity, differential patterns of suspension should be evident across various student-teacher race pairings. Using elementary school data I find no evidence that this is the case.

A more direct test of racial bias in discipline is possible by conditioning on observed student behavior. Thus, the focus of this paper is a comparison of discipline outcomes for black and white students conditional on a discipline referral. I consider two discipline margins a principal must decide upon once a student has been referred: whether to suspend a student, and if so, the length of the suspension. For each of the discipline margins, I model

---

<sup>12</sup>Gregory (1995) uses national level data, while McFadden and Marsh (1992) and Braden and Shaw (1990) aggregate discipline files across 10 and 16 schools respectively.

<sup>13</sup>Mendez et al. (2002) show that in a large school district in Florida, schools serving high proportions of black students have the highest out of school suspension rates, suggesting that at least part of the documented racial gap in punishment originates from cross-school variation in discipline. Without data on individual offenses, Mendez et al. (2002) are unable to examine the consistency of discipline within schools.

punishment conditional on the type of offense, the number of infractions the student has already committed in the current year, and a host of other observable student characteristics. Because I am limited to only one year of discipline data, I concentrate on students in 6th and 9th grade who are new to their current schools and have not previously interacted with school administrators.

The findings illustrate a significant statewide racial gap in discipline that is primarily generated by cross-school variation in student punishment. Across North Carolina, black 6th grade students are 79% more likely to be suspended for violating school rules than white 6th grade students who violate school rules. When the same comparison is made controlling for school fixed effects, there is no difference in punishment for black and white students. Similar cross-school and within-school discipline patterns emerge when predicting suspension length conditional on receiving a suspension. Across schools, black 9th grade students receive 22% longer suspensions than white students when committing simple rule violations. This gap disappears entirely when black and white student suspensions are compared within schools. The remaining analysis focuses on the robustness of the basic within-school equity result and identifying the role of principal race in generating the cross-school racial gap in discipline.

A broad outline for the rest of the paper is as follows. Section 2 provides a brief description of the North Carolina student discipline data. Section 3 examines the overrepresentation of black students in suspension rolls, with a focus on the interaction of student and teacher race. The within and across school discipline gaps for 6th and 9th graders is estimated in Section 4. Section 5 checks the robustness of these results and Section 6 concludes.

## **2 North Carolina Student Discipline Data**

During the 2001 academic year, public schools in North Carolina were required by the state to record all student infractions resulting in out of school suspension.<sup>14</sup> For each incident

---

<sup>14</sup>Data for other common disciplinary actions, such as verbal warnings, detention, or suspension of extracurricular activities is not available since schools were not required to record this information.

schools had to capture information about the type of infraction, student involved, date of the incident, and school response. To get a sense for the types of misbehavior students participate in, Table 1 lists the frequency and average suspension length for the ten most common infractions resulting in suspension during the 2001 school year. Typical adolescent behaviors such as rowdiness and fighting are quite common, while more dangerous behaviors like drug or weapon possession are less frequent.

Without additional information, however, the discipline data is not particularly useful for evaluating the role of racial bias in punishment. First, besides race and gender there is little student demographic information available, making it difficult to control for other differences between black and white students. More importantly, incidents are assigned to school districts rather than schools, eliminating any possibility of examining racial disparities within schools. To mitigate these issues I link recorded infractions to student testing data using unique student identifiers.<sup>15</sup> Included in the testing data is the exact school a student attends and more detailed demographic information. A handful of school districts use proprietary identifiers in the discipline data resulting in match rates lower than 80%.<sup>16</sup> I drop these 24 school districts, leaving 93 school districts with an average match rate of 88%. Offenses from the included districts that cannot be linked to the testing data are also dropped from the sample.

The 93 school districts included in the analysis serve approximately 500,000 students in 1,000 schools distributed across three levels of schooling: elementary, middle, and high school. To ensure valid cross-school comparisons of punishment I eliminate charter, magnet, and other vocational schools. In addition, within each level of schooling I keep only those schools with approximately the same grade configuration. For example, only middle schools that span grades 6-8 or 7-8 are included in the sample. Table 2 provides detailed student

---

<sup>15</sup>Testing data is not available for students below third grade since they have not taken any state-wide exams. Thus only incidents occurring in third grade or higher are included. Students below third grade account for only 4% of total incidents.

<sup>16</sup>All of the following analysis was also conducted using cutoffs of 70% and 90% with little impact on the results.

demographic information by schooling level for the entire matched sample.

The first two rows of Table 2 indicate that students are relatively evenly distributed across grades, but that middle and high schools are significantly larger than elementary schools. Black students make up approximately a quarter of the student population across the state. Based on student date of birth and the cutoff date for enrollment, about 20% of students are considered old for their grade. In elementary and middle schools a significant number of students are either physically or learning disabled, and approximately 40% of students are on free or reduced price lunch. Lagged math score is a student's normalized end of grade math score from the previous year, or in the case of high school students, their score in 8th grade. The final row of the table indicates the proportion of students who receive at least one suspension during the year. Overall, approximately 10% of public school students are suspended at least once, with lower rates in elementary school and higher rates in middle and high school. The small decline in suspension rates from middle to high school masks the large drop off in suspension for older high school cohorts (11th and 12th graders), a pattern that likely results from selection into dropout.

In addition to the apparent increase in dropout propensity, how else do problematic students differ from their more well behaved classmates? Table 3 displays the demographic characteristics of students who were suspended at least once during the year. Compared to the entire population, suspended students are significantly more likely to be male, black, and come from low income households. In middle school and high school, the average disorderly student is suspended twice and misses more than a week of school as a result. The final row of Table 3 highlights the strong negative correlation between suspension and achievement. Students who are suspended at least once during the year score close to one-half a standard deviation below their well behaved peers.

The discipline data from North Carolina is largely consistent with previous research regarding student discipline. Suspended students perform significantly worse than their peers and minority students are drastically overrepresented in suspension rolls. As noted in the

introduction, racial bias in the unconditional rate of suspension cannot be separated from racial differences in unobserved behavior. However, a more direct test of bias is possible by comparing black and white discipline outcomes for the same observed behavior. Throughout the paper, I focus primarily on two discipline margins: the probability of suspension conditional on a discipline referral, and suspension length conditional on receiving a suspension. In the text to follow, I refer to the first margin as the extensive discipline margin and the second as the intensive discipline margin.

Before analyzing the raw punishment differences across the extensive and intensive discipline margins, two further caveats are necessary. Because schools were only required to record infractions resulting in suspension, it should not be possible to examine the extensive discipline margin. However, approximately 50% of the schools in the state report infractions resulting in in-school or no suspension. This subset of schools, which I refer to as schools reporting unpunished infractions, are used to model the extensive discipline margin.<sup>17</sup> For the intensive discipline margin I use data from all schools in the matched sample. However, to limit the effect of outliers when analyzing the intensive margin, I include only short-term suspensions, or suspensions for 10 days or fewer.<sup>18</sup> Short-term suspensions account for 99.4% of offenses resulting in an out of school suspension.

To get a sense for whether there is any scope for racial bias in the discipline data, Table 4 highlights racial differences in punishment conditional on behavior across both the extensive and intensive discipline margins. The first row in both the extensive and intensive discipline panels shows punishment conditional on committing any infraction, while the following

---

<sup>17</sup>Along observable dimensions, schools recording unpunished infractions are nearly identical to the full sample of schools. Also, for many school districts I am unable to differentiate between no punishment and in-school suspension. Because of this I do not separately analyze these outcomes.

<sup>18</sup>Short-term suspensions differ from long-term suspensions in two ways besides suspension length. First, the district superintendent must approve a long-term suspension. Second, an appeal of a long-term suspension must be heard by the school board, whereas a short-term suspension appeal may be denied a hearing. In addition, a student on short-term suspension must be provided with, (1) the opportunity to take textbooks home for the duration of the suspension, (2) the right to inquire about homework assignments for the duration of the suspension, and (3) the opportunity to take any quarterly, semester, or grading period examinations missed during the suspension period. Further information regarding due process can be located in Chapter 115C Article 391 of the North Carolina General Statutes for Public School Law.

rows condition on particular types of infraction. When compared to white students, black elementary school students are 10 percentage points more likely to receive an out of school suspension conditional on committing any infraction. For an average high school fighting offense, black students are suspended a full day longer than white students. Table 4 clearly shows that overall black students receive significantly stiffer penalties for the same observed behavior.

The critical question is does racial bias play a significant role in generating these disparities in discipline? Before directly addressing this question, I examine the overrepresentation of black students in suspension rolls and then return to the racial gap in discipline conditional on behavior.

### **3 The Overrepresentation of Black Students in Suspension Data**

While the gaps in discipline conditional on student behavior are statistically and economically significant, the unconditional gap in the likelihood of receiving at least one out of school suspension during the year is staggering. As Tables 2 and 3 illustrate, only 26% of the student population in North Carolina is black, yet 46% of suspended students are black. This disparity in discipline is quite similar to the gap evident at the national level. One concern with these results is the aggregate nature of the data. If schools differ in their tolerance for misbehavior, then much of the gap in suspension rates could be explained by cross-school variation in punishment.

To investigate this possibility, I estimate a linear probability model that predicts the likelihood of receiving at least one suspension conditional on a host of demographic charac-

teristics and school fixed effects.<sup>19</sup> The basic model is given by

$$\text{Suspension}_{igs} = X_i\beta + \alpha_g + \phi_s + \epsilon_{igs} \quad (1)$$

where the dependent variable is binary and equals 1 if student  $i$  in grade  $g$  and school  $s$  received at least one out of school suspension during the year.  $X_i$  are observable student characteristics, such as race, sex, and past achievement, and  $\alpha_g$  are grade dummy variables. I estimate the model separately by school type using all public school students satisfying the sample restrictions noted in the previous section.

Table 5 contains the coefficients of interest for each level of schooling, and it is clear that even within schools there is a significant racial disparity in the likelihood of receiving at least one suspension. A black student is 3.8, 6.6, and 5.7 percentage points more likely to receive a suspension than an otherwise similar white student in elementary, middle and high school respectively. While these numbers seem small, the probability of ever being suspended is also rather small. Thus, for an average male student in elementary, middle, and high school, being black is associated with a 133%, 69%, and 32% increase in the probability of ever being suspended. Interestingly, the disparity in the probability of receiving at least one suspension decreases significantly as students move up the education ladder, even though the overall number of offenders increases dramatically. This could reflect improvement in black student behavior relative to white students, or it could indicate that some of the most troublesome students exit the regular public school system as they get older.

Racial gaps also exist between white students and other racial groups when it comes to the probability of receiving at least one out of school suspension. Asian and Hispanic students are less likely to ever receive a suspension, while all other non-white student groups are more likely to receive at least one suspension. However, none of these gaps is as large in magnitude as the gap between black and white students.

---

<sup>19</sup>Estimating either a logit or probit model with school dummy variables yields essentially identical results.

The sign and significance of the remaining coefficients in Table 5 are as expected. Students who are repeating a grade or have a physical or learning disability are considerably more likely to be suspended during the year. Low income and low achieving students are also more likely to be cited for misbehavior. Estimates of  $\alpha_g$  indicate varying patterns of behavior across schooling levels. In elementary and middle school, students in advanced grades are more likely to be suspended, while the opposite is true in high school. The decreasing likelihood of suspension as students advance through high school is consistent with the notion that problematic students are more likely to dropout.

As noted in the introduction, one common explanation for the black-white disparity in the likelihood of suspension is racial bias. Yet it seems premature to jump to that conclusion based on the evidence in Table 5, since there is no way to control for unobserved behavior differences between students. One indirect test of racial bias that is possible despite the lack of observed classroom behavior is to examine patterns of suspension across different student-teacher race pairings. If racial bias is a primary determinant of the overrepresentation of black students in suspension rolls, then higher rates of suspension should result when a white teacher is paired with a black student. To test for this empirical fact I investigate how the interaction of teacher and student race affects the likelihood of suspension.

The analysis of how student-teacher racial interactions affect suspension is only possible for the elementary school sample. Despite the fact that the discipline data does not indicate where the offense occurred or who reported the offense, elementary school students spend the vast majority of their day in one classroom with one teacher. Thus it is extremely likely that most elementary school incidents are originally reported by a student's teacher. This is not the case for middle and high school students who switch teachers and classrooms throughout the day.

In the testing data, elementary school students are linked to their teacher through a unique identifier. This identifier can be matched to state personnel records that include the race, sex, and tenure of the teacher. Using the sub-sample of students with matched teacher

data, I estimate Equation (1) adding in basic teacher information and interaction terms between the race of the teacher and the race of the student. The sample size is reduced in this case because of missing teacher information and the fact that I limit the sample to only black and white students. Results are shown in Table 6. To keep the table simple, the coefficients on all student demographic variables other than race are omitted. The two columns of results contain estimates with and without school fixed effects.

The coefficients for the interactions between student and teacher race indicate that across schools there is an increased likelihood of suspension for all student-teacher race combinations relative to that of a white teacher with a white student. When only within-school variation in student-teacher race interactions is used, there is no evidence that white students paired with black teachers are more likely to be suspended than white students paired with white teachers. For black students, there is a significant increase in the likelihood of receiving at least one suspension regardless of the teacher's race. The question is whether this increase is identical for black and white teachers. For white teachers the discipline gap between black and white students is equal to 0.0368 and for black teachers the gap is equal to 0.0373. The F-statistic for the null hypothesis that these gaps are equal is 0.01 with a p-value of 0.94. Thus, we fail to reject the null hypothesis that the black/white student suspension gap is identical across teacher race.

The result that the racial gap in the likelihood of receiving at least one suspension does not depend on teacher race certainly doesn't rule out the possibility that there is bias in punishment. A theory proposed by Townsend (2000) that might explain this result is that troublesome black students differ significantly even from their black teachers on dimensions other than race. Thus, the punishment disparity has nothing to do with race per se, but reflects differences in behavioral attitudes across cultures. If this is the case we would expect to see black students being suspended for more minor infractions, since these are more likely to reflect different behavioral perceptions. Yet in elementary school the two most prevalent types of minor offense, rule violations and undisciplined behavior, account for 46% of black

offenses and 44% of white offenses.

Without considering variation in actual student behavior it is difficult to make concrete conclusions about the source of the racial disparity in the likelihood of receiving at least one out of school suspension. The advantage of the North Carolina discipline data is that I am not limited to this discipline margin. Using incident level punishments and variation in punishment for the same types of offense, I can explore how punishment varies for black and white students.

## 4 Punishment Conditional on a Discipline Referral

Students are referred to the principal's office when their behavior at school exceeds an acceptable level. It is then up to the principal to decide the appropriate punishment. The goal of this section is to determine if student race affects a principal's discipline decision along two particular discipline margins.<sup>20</sup> The first is whether to suspend the student from school for any period of time, and the second is conditional on deciding to suspend the student, how long should the suspension last?<sup>21</sup>

While the North Carolina discipline data is quite rich, there is one significant shortcoming that complicates the analysis of race and punishment. After 2001, considerable changes were made to the discipline data that make linking suspended students to their test scores infeasible. As a result I am limited to one year of data. This is problematic for punishment comparisons between black and white students if there is significant variation in past offensive behavior across race. Based on the contemporaneous correlation between race and behavior, omitting all previous misbehavior will likely lead to an upward biased black coefficient. To

---

<sup>20</sup>A principal has many disciplinary tools for dealing with misbehaving students. These include a simple verbal warning, parent teacher conferences, detention, suspension of extracurricular activities, in-school suspension, out-of-school suspension, and expulsion. Data limitations keep me from examining discipline choices along all these margins.

<sup>21</sup>These are decisions that are primarily left to the principal except in very rare circumstances, such as when a student brings a firearm to school. In that case, national and state law mandate removal of the student from the school for a minimum of a year.

minimize the bias induced by unobserved history, I analyze discipline outcomes for 6th and 9th graders who are new to their current school.<sup>22</sup> The basic idea is that student reputation persists less across schools than within schools.

To examine racial variation in discipline conditional on behavior, I estimate extensive and intensive punishment equations separately for 6th and 9th graders. The basic estimating equation is given by

$$\text{Outcome}_{ifs} = X_i\beta + H_i\gamma + \kappa_f + \phi_s + \epsilon_{ifs} \quad (2)$$

where the Outcome for student  $i$  committing infraction type  $f$  in school  $s$  is either a suspension indicator in the case of the extensive margin, or the number of days suspended in the case of the intensive margin.<sup>23</sup>  $X_i$  represents observed demographic variables and  $H_i$  is the current year’s behavioral history.<sup>24</sup>  $\kappa_f$  are dummy variables for the category of offense and  $\phi_s$  are school fixed effects. I estimate the models with and without  $\phi_s$  to determine how much racial variation in discipline can be absorbed by variation in punishment across schools. To accommodate the inclusion of school fixed effects, I estimate a linear model for both the extensive and intensive discipline outcomes.<sup>25</sup> Standard errors are clustered at the student level when estimating Equation (2) and throughout the remainder of the paper since students may appear in the discipline data numerous times. The regression results for the two discipline margins are listed in Table 7.

Before diving in to the results regarding the racial gap in discipline, it is important to note that while not listed, the pattern of infraction category coefficients,  $\kappa_f$ , is as expected.

---

<sup>22</sup>Students repeating 6th or 9th grade are excluded.

<sup>23</sup>Recall that the estimation samples across the two discipline margins differ within grades. The extensive margin sample includes only schools reporting both punished and unpunished infractions, while the intensive margin sample uses all schools but only those infractions that resulted in suspension. Estimation of the intensive margin using only schools reporting both punished and unpunished infractions yields similar results.

<sup>24</sup>Indicators for learning and physically disabled are included as part of the observed demographic variables. According to the NC Handbook of Parent’s Rights, disabled students are treated identically to non-disabled students when it comes to short-term suspensions until the disabled student accumulates more than 10 days out of school across multiple suspensions. As a result, I also estimate the main regressions excluding disabled students. The finding of cross-school disparities and within-school equity in punishment is maintained. Results available upon request.

<sup>25</sup>The extensive model is also estimated using a logit, and the intensive model is also estimated using a log-linear framework. The results are consistent across all specifications.

Committing more serious infractions leads to significant increases in both the probability and length of suspension. For example, within schools a 6th grade fighting infraction increases the probability of suspension by 35% points when compared to a simple rule violation. Similarly, 6th grade suspensions for fighting are 0.77 days longer than suspensions for rule violations. The behavioral history variables included in the model are the number of previous offenses and an indicator for whether the student was written up for multiple offenses on the same day.<sup>26</sup> Not surprisingly, both of these variables are associated with stiffer penalties for misbehavior.

While these results are interesting in their own right, the primary motivation for the analysis is to understand the source of the racial gap in discipline.<sup>27</sup> The results in Table 7 indicate that much, if not all of the statewide discipline gap can be explained by cross-school variation in discipline policy.<sup>28</sup> Across schools the results largely reflect the raw differences in punishment discussed in the data section. Along the extensive margin, a black 6th grade student is 7% points more likely to be suspended conditional on infraction type. For a simple rule violation, this represents a 79% increase in the probability of suspension when compared with an average white student. However, when the probability of suspension for black and white students is compared within schools, this difference disappears entirely. The coefficient on the black indicator is now -0.004 and is rather precisely estimated. The upper bound of the 95% confidence interval for the black coefficient is 0.014, which itself is a drastic decrease from the estimate across schools. A similar pattern of cross-school discipline disparity and within-school discipline equity occurs for the intensive margin in 6th and 9th grade. For the 9th grade extensive margin, the black coefficient is essentially zero both within and across

---

<sup>26</sup>Same day offenses are treated as one incident and are assigned the offense type of the more serious infraction.

<sup>27</sup>Indicators for other racial categories: Asian, Hispanic, and Other non-white are included in the regressions, but excluded from the table for presentation purposes. However, there is no evidence of punishment disparities within schools for these groups when compared with white students. For the remainder of the paper I will focus entirely on the gap in punishment between black and white students.

<sup>28</sup>Differences in punishment across schools could also reflect varying incident severity within particular offense categories rather than any actual policy differences. The conclusions of this paper are not sensitive to the source of the cross-school variation in discipline.

schools.

A closer look at the estimated black coefficient for the 6th and 9th grade intensive margins indicate that neither coefficient is estimated precisely. For 6th grade students, the upper bound of the 95% confidence interval is 0.135. Thus, it is possible that a small gap still exists, but it is certainly significantly smaller than the cross-school result. The intensive discipline gap across schools in 9th grade is quite stark. A simple rule violation results in a 22% longer suspension for a black student. Within schools there is essentially no difference, with the upper bound of the 95% confidence interval of the black coefficient equal to 0.06. Even at this extreme value, the discipline gap between black and white students for a minor offense is reduced to 3%.

On the whole, the results from Table 7 suggest that a significant portion of the racial gap in discipline comes from variation in policy across schools. Thus it appears that the race of the offending student has little impact on individual principal decisions. However, there may be important within-school variation in discipline not captured in the simple model outlined in Equation (2).

## 4.1 Heterogeneity in the Discipline of Black Students

In this section I explore three sources of racial heterogeneity in punishment not discernable in the basic framework previously considered. The three sources of heterogeneity are targeted discipline for particular types of black students, variation in the discipline gap across offense types, and heterogeneity in the within-school discipline gap.

### 4.1.1 Profiling

A common thread in the literature on the application of suspension is not just that black students are disciplined more severely, but that certain categories of black students are targeted.<sup>29</sup> In particular, it is often suggested that low-income, black males suffer the harshest

---

<sup>29</sup>See Gregory (1995), McCarthy and Hoge (1987), and Townsend (2000).

punishments. To determine the extent of punishment variation among black students, I repeat the analysis from the previous section, interacting student race with gender, achievement, and free/reduced lunch status. The coefficients of interest are listed in Table 8. All regressions include school fixed effects.

The only consistent result across all grade and discipline margins is that none of the coefficients is statistically significant. This reflects a combination of small effects and imprecise estimates. There does not seem to be strong evidence to support claims that a certain type of black student is treated more harshly.

#### 4.1.2 Variation by Offense Type

A second source of racial heterogeneity in punishment within schools is variation in punishment by offense type. Studies of the sentencing gap between black and white criminals in U.S. Federal Courts show that the disparity in punishment tends to be greatest for the most serious offenses, such as drug trafficking and bank robbery.<sup>30</sup> To investigate this possibility I re-estimate the baseline specification, interacting race with offense type. Rather than generating an interaction term for each of the 26 types of offense, I interact race with the 7 most common offenses and one category encapsulating the remaining more serious offenses. Only black and white students are included in this analysis, yielding slightly smaller sample sizes.

The results of this exercise for 6th and 9th graders across both discipline margins are shown in Table 9. A non-interacted offense dummy variable is included in each regression, thus the coefficients represent the difference in discipline between black and white students for each offense type within schools. The extensive margin results indicate very little heterogeneity in the discipline gap across offense type. In fact, the only significant differences suggest bias against white students.

In contrast, there does appear to be significant heterogeneity in the discipline gap within

---

<sup>30</sup>See Mustard (2001) or Steffensmeier and Demuth (2000).

schools along the intensive margin. Black students in 6th grade receive significantly longer suspensions for truancy violations, while black students in 9th grade receive slightly shorter suspensions for rule violations. More interesting however is the consistent result across the two schooling levels that black students receive significantly longer suspensions than white students for fighting offenses. Compared to white students, fighting suspensions for black students are 18% and 24% longer in 6th and 9th grade respectively. This variation could reflect heterogeneity in incident severity or bias on the part of administrators.

One way to test for bias in fighting suspensions is to compare discipline outcomes for incidents involving students of different races. However, the North Carolina discipline data does not provide incident identifiers. To circumvent this limitation I assume that fighting infractions occurring in the same school on the same day belong to the same incident. Under this somewhat strong assumption, there are 2,176 fighting incidents in 6th and 9th grade that involve multiple participants, approximately 33% of which involve students of different races. Using this limited sample I find no significant discipline gap between black and white students conditional on either incident or school fixed effects. This is reassuring, though bias in the suspension of black students for one-off incidents of aggressive behavior certainly cannot be ruled out.

#### **4.1.3 Distribution of Bias**

The final source of heterogeneity I consider is variation in the within-school discipline gap. The baseline results showing equity in discipline within schools simply indicates that the average school punishes black and white students equally. However, the same result would be obtained if half the schools in the sample were biased against black students while the other half were equally biased against white students. To estimate the overall distribution of bias, rather than include one fixed effect for each school I include both a black and white school fixed effect. Schools with fewer than 5 black or white incidents are excluded from the analysis to limit the small sample bias in the estimated fixed effects.

Figure 1 plots the distribution of the racial gap for each discipline margin and school level. The x-axis in each graph is simply a school identifier, though these numbers cannot be compared across graphs. The schools are ordered such that school number 1 has the largest white fixed effect relative to its black fixed effect. The point to take away from Figure 1 is that it does not appear that the bias distribution is heavily skewed toward the extremes - bias against both black and white students. As expected, the schools are evenly split above and below the x-axis, and the variation in the bias is rather small. To be a bit more precise, only 11 out of 110 middle schools have black and white fixed effects that are significantly different at a 5% level for the intensive discipline model. Similar ratios result for the other margin and grade combinations. Given that at a 5% level we expect some of the schools to have different fixed effects simply due to chance, it appears that much of the within-school racial variation in discipline is random.

As a final test of racial bias in punishment within schools, I use the estimated black and white fixed effects for the extensive and intensive discipline margins to examine if any schools have significantly larger black fixed effects in both cases. The idea is that a school that purposely discriminates against black students along one margin would also likely discriminate against black students along the other margin. However, none of the 6th or 9th grade schools that have a significantly larger black fixed effect along the extensive margin also have a significantly larger black fixed effect along the intensive margin. This is further evidence that the within-school racial variation in punishment is largely random.

The preponderance of evidence thus far sheds serious doubt on the conclusion that there is racial bias in the application of suspension, particularly when conditioning on observed delinquent behaviors. Most of the statewide racial variation in punishment appears to stem from cross-school variation in discipline policy. However, racial discrimination may be at the heart of this variation. In addition, other sources of discipline bias, such as the misallocation of infractions to offense types, could lead to an incorrect finding of no bias. The next section explores these issues and the robustness of the equity in discipline result.

## 5 Other Sources of Bias and the Black-White Discipline Gap

Conditioning on the type of infraction, the number of previous offenses, and the school a student attends results in a small and economically insignificant discipline gap between black and white students. However, it is possible that any racial bias in punishment is embedded in these conditioning variables, leading to an erroneous finding of equity in discipline. In this section I examine how the estimated discipline gap responds to changes in the included explanatory variables, expansion of the estimation sample, and interactions between principal and student race.

### 5.1 Bias in Offense Type and Previous Offenses

Rather than explicitly punishing black and white students differently for the same offense, bias may exist in the assignment of offense type or in the initial assessment of bad behavior. For example, if schools assign black student infractions to more serious offense categories, either intentionally or unintentionally, then more severe punishments will appear valid. Predicting discipline while conditioning on infraction categories may then understate the extent to which schools discriminate. In addition, if black students accumulate offenses more quickly because of bias in initial reports, then controlling for the number of previous offenses can mask bias in the severity of discipline. Removing these controls from the model should increase the gap in discipline severity if bias exists along these other margins.

Table 10 tests this hypothesis for 6th and 9th graders by excluding from the baseline specification either infraction type, offense history, or both. School fixed effects and other observable individual characteristics are included in all of the regressions. Standard errors are clustered at the student level. For comparison purposes, the coefficients from the baseline specification are included in the first column for both 6th and 9th graders.

Overall the race coefficients increase slightly when infraction type and previous offense

controls are excluded from the model, but in most cases no significant racial gap appears. For 6th graders, the biggest jump in the black coefficient along both discipline margins occurs when the student's behavioral history is not considered. Thus, the accumulation of offenses appears to play an important role in generating punishment differences between black and white students. However, even when both sets of conditioning variables are excluded for the 6th grade sample, there is no significant discipline gap.

For 9th graders, the intensive discipline gap increases slightly from a negative number to essentially zero when both conditioning variables are excluded. Along the extensive margin, the discipline gap expands and becomes significant when either offense type or previous offenses are excluded. When both conditioning variables are excluded the gap is three times as large as in the baseline model and statistically significant. The black coefficient in this case translates into an 8% increase in the probability of suspension for a black student compared to a white student when committing a rule violation. It is important to note that by excluding infraction type and previous offense some valid variation is also lost, thus I would interpret this as an upper bound on the effect of racial bias.

Completely excluding past offenses is an extreme way to limit the effect of bias in the first decision to suspend or discipline a student. Much of the variation in punishment used to identify the coefficients could come from comparing "career" offenders to first time violators. Another way to assess the role of past offenses in generating the racial gap in discipline is to compare punishments across black and white students for a first offense, second offense, and so on. Doing this for 6th and 9th graders across both discipline margins yields no significant differences in discipline between black and white students for a first, second, or third offense.<sup>31</sup> Again, it could be that the threshold for a first black offense is lower than that for a first white offense. However, if this were the case the type of infraction for first black offenses should be skewed towards the less serious infractions as compared to the first white infraction. This is not supported in the data.

---

<sup>31</sup>Results available upon request.

## 5.2 Discipline Gap in All Grades

The lack of a significant black-white discipline gap for 6th and 9th grade students within schools is an encouraging finding. However, does this result extend to students in other grades? Given the results of the previous section, it seems somewhat unlikely. Previous offenses in the current year play an important role in explaining the discipline gap within schools. If previous offenses from past years are also an important determinant of discipline, utilizing discipline data on students at all grade levels will likely lead to a finding of a significant discipline gap. Yet it is important to assess the extent to which equity in discipline permeates throughout the school system.

To examine the discipline gap in all grades, I estimate the baseline model separately for each level of schooling, including all grades in the analysis. For elementary schools the sample includes 3rd, 4th, and 5th graders. The middle school sample includes 6th, 7th, and 8th graders, and the high school sample includes 9th, 10th, 11th, and 12th graders. Table 11 lists the coefficient for the black indicator variable. All regressions include other student demographic variables and infraction type dummy variables. Standard errors are clustered at the student level.

The models are estimated with and without school fixed effects again to illustrate the extent to which cross-school variation in policy can explain the statewide racial gap in discipline. The estimated racial gap in discipline decreases for all schooling levels and discipline margins when school fixed effects are added to the model. In all but one case, the decrease is dramatic. The one time the black coefficient does not change significantly is the extensive margin in high school, for which there was no real gap to begin with. In terms of statistical significance, there remains a significant gap in discipline for high school students along the intensive margin. However, the gap went from almost a half-day difference in suspension to less than a tenth of a day difference. None of the other coefficients are statistically significant at a 5% level.

While most of the coefficients are not statistically significant, we cannot rule out a

positive gap since most of the black coefficients are not precisely zero. For example, along the intensive margin in elementary school, the upper bound of the 95% confidence interval is 0.12. It is safe to rule out a gap larger than that but it remains possible that the true gap is somewhere between zero and about a tenth of a day. Yet even this value is half the size of the gap when estimated without school fixed effects. A similar pattern emerges for all the other imprecise coefficients. The upper bound of the 95% confidence intervals yields minor gaps in discipline that are drastically smaller than the estimated effect across schools.

Despite the potential for upward bias in the black coefficient induced by unobserved behavioral history, the finding of within-school equity in discipline is largely maintained when all grades are included in the analysis. While small and insignificant gaps remain within schools, it is clear that the majority of the statewide racial gap in discipline can be explained by cross-school variation in policy. However, it is still not possible to entirely rule out the potential for racial bias, particularly across schools. Racially motivated principals may choose to work in high minority schools, aware that any strict discipline policy will overwhelmingly affect black students. The next section addresses this issue, examining how principal race affects discipline behavior.

### 5.3 Cross-School Variation in Discipline and Principal Race

Concluding that racial bias is a negligible factor in determining student discipline based strictly on the within-school results is a bit speculative. In fact, the within-school equity result combined with the overall aggregate gap suggests that schools with high proportions of minority students utilize stricter punishments.<sup>32</sup> If racial bias is the source of the cross-school variation in discipline, as opposed to convex costs of misbehavior or heterogeneity in the

---

<sup>32</sup>Replacing the school fixed effects in the basic punishment analysis (Table 7) with observable school characteristics, including minority share of the school, support this assertion for two of the three punishment outcomes for which an aggregate gap exists. The effect of other observable school characteristics is quite mixed across schooling levels and punishment margins. However, the individual racial gap in punishment does not fully dissipate when school characteristics are utilized instead of fixed effects. Results available upon request.

response to punishment, we would expect discipline to vary with principal race conditional on the minority share. The aim of this section is to explore how varying combinations of student and principal race affect discipline outcomes.

The difficulty in examining how the relationship between student and principal race affects discipline using the North Carolina data is that generally no demographic information on principals is available. Principals can be identified in the state payroll files based on their job codes, but because they have no direct interaction with students they are not listed in the personnel files that include basic demographic information. To sidestep this issue I utilize the fact that before becoming principals, most individuals start as teachers. The teacher personnel data go back to 1995, at which point some of the principals in 2001 were still teachers. Using this strategy I am able to determine principal race for about a third of the schools in the sample. Because these principals were teachers relatively recently, this sub-sample is skewed towards younger, inexperienced principals. One reassuring result is that using this selected sample of schools, the estimated-cross school discipline gaps for all schooling levels and discipline margins are quite similar to the estimates obtained using the full sample.<sup>33</sup>

If racial bias is driving the cross school variation in discipline, there should be significant differences in the way white and black principals treat black student misbehavior. To test for this pattern in the data, I estimate a model that includes interactions between the race of the principal and the offending student. Infraction type dummies and basic student demographic variables are included in all regressions. The models also include the proportion of the school that is black and the average lagged math score in the school. These variables are added to ensure that the behavioral comparison between black and white principals comes from observably similar schools.<sup>34</sup> Standard errors are clustered at the student level. Tables 12 and 13 list the student-principal race interaction coefficients and a set of F-statistics testing

---

<sup>33</sup>Results available upon request.

<sup>34</sup>Excluding these variables does not change the results of the test for whether principal race is related to the racial gap in discipline.

the relationships among these coefficients for each discipline margin. For the F-tests,  $\beta_{XY}$  is the interaction coefficient where the principal's race is  $X$  and the student's race is  $Y$ .

The results pertaining to the extensive discipline margin are listed in Table 12. The pattern of student-principal race interaction coefficients is rather inconsistent across schooling levels. In elementary school, black students are more likely to be suspended conditional on a discipline referral, and black principals are more likely to use suspensions. The exact opposite pattern is true in high school. Increasing the proportion of black students in a school leads to an increased probability of suspension in elementary and middle school, and a decrease in the probability of suspension in high school. The final row of the table tests whether the difference in the likelihood of suspension between black and white students varies with the race of the principal. In each case, either there is no significant difference, or the gap is actually greater for black principals compared to white principals. Thus, the results suggest that the statewide racial gap in the probability of suspension conditional on a discipline referral is not related to variation in principal race.

The student-principal race interaction coefficients in Table 13 are considerably more consistent across schooling levels. Black students regularly receive longer suspensions, and the average black principal suspends offending students for longer periods of time. For all school types, suspension length is positively correlated with the proportion of the student body that is black. Again the critical test is whether the gap in suspension length between black and white students varies with the race of the principal. The final row of Table 13 shows that there is no significant difference in elementary or middle school, and a marginally significant difference in high school. Again, it appears as if principal race plays a very small role in generating the statewide discipline gap.

The cross school variation in discipline policy does not appear to be driven by white principals severely punishing black students. The fact that black principals also tend to punish black students more harshly suggests that there might be other reasons why schools with high proportions of black students find it advantageous to use lengthy suspensions.

It might also reflect the theory espoused by Townsend (2000) that black principals are so drastically different from their students that they are as culturally biased as white principals.

## 6 Conclusion

Racial bias is often cited as a source of the black-white disparity in the unconditional likelihood of suspension and in the application of punishment conditional on behavior. This is troubling given the links between suspension and low achievement, dropout, and the likelihood of arrest. In this paper, I show that student race plays a very small role in principal punishment decisions. Within schools, black and white students are equally likely to be suspended and receive similar suspension durations conditional on behavior and a host of other observable characteristics. This result is robust across schools, schooling levels, and infraction types. In addition, the within-school equity result is largely maintained even when infraction type and behavioral history are excluded from the model.

The finding that black and white students are treated equally within schools is a positive result. However, significant racial gaps in discipline across schools could be explained in part by racial bias. The previous section shows that the cross-school discipline gap between black and white students does not vary with principal race. Again, this cannot rule out racial bias, but it certainly suggests that it plays a much smaller role than previously believed. The question remains, what drives the cross-school variation in discipline? In addition, do schools consider the long-term consequences associated with suspension when making their decisions? To answer these question, we need to consider how discipline affects the school environment through its impact on student behavior and achievement, and how suspensions impact the community at large. Future work will address these issues.

## References

- Ayres, Ian and Joel Waldfogel**, “A Market Test for Race Discrimination in Bail Setting,” *Stanford Law Review*, May 1994, *46*, 987–1048.
- Bowditch, Christine**, “Getting Rid of Troublemakers: High School Disciplinary Procedures and the Production of Dropouts,” *Social Problems*, November 1993, *40* (4), 493–509.
- Braden, Jeffrey and Steven Shaw**, “Race and Gender Bias in the Administration of Corporal Punishment,” *School Psychology Review*, 1990, *19* (3).
- Children’s Defense Fund**, “School Suspensions: Are they helping children?,” Technical Report, Cambridge,MA: Washington Research Project 1975.
- Costenbader, Virginia and Samia Markson**, “School Suspension: A Study with Secondary School Students,” *Journal of School Psychology*, 1998, *36* (1), 59–82.
- Currie, Janet and Duncan Thomas**, “Early Test Scores, Socioeconomic Status and Future Outcomes,” 1999. NBER Working Paper 6943.
- Eitle, Tamela M. and David J. Eitle**, “Inequality, Segregation, and the Overrepresentation of African Americans in School Suspensions,” *Sociological Perspectives*, 2004, *47*, 269–287.
- Everett, Ronald S. and Roger A. Wojtkiewicz**, “Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing,” *Journal of Quantitative Criminology*, June 2002, *18* (2), 189–211.
- Gregory, James F.**, “The Crime of Punishment: Racial and Gender Disparities in the use of Corporal Punishment in U.S Public Schools,” *The Journal of Negro Education*, Autumn 1995, *64* (4), 454–462.

- Grogger, Jeffrey**, “The Effect of Arrests on the Employment and Earnings of Young Men,” *The Quarterly Journal of Economics*, 1995, 110 (1).
- Hjalmarsson, Randi**, “Criminal Justice Involvement and High School Completion,” *Journal of Urban Economics*, 2008, 63.
- Jacob, Brian and Lars Lefgren**, “Are Idle Hands the Devil’s Workshop? Incapacitation, Concentration, and Juvenile Crime,” *American Economic Review*, 2003, 93 (5), 1560–1577.
- Lochner, Lance and Enrico Moretti**, “The Effects of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports,” *The American Economic Review*, 2004, 94 (1).
- McCarthy, John and Dean Hoge**, “The Social Construction of School Punishment: Racial Disadvantage Out of Universalistic Process,” *Social Forces*, 1987, 65 (4).
- McFadden, Anna C. and George E. Marsh**, “A Study of Race and Gender Bias in the Punishment of School Children,” *Education and Treatment of Children*, May 1992, 15 (2).
- Mendez, Linda M. Raffaele**, “Predictors of Suspension and Negative School Outcomes: A Longitudinal Investigation,” *New Directions for Youth Development*, 2003, 99.
- , **Howard M. Knoff, and John M. Ferron**, “School Demographic Variables and Out-of-School Suspension Rates: A Quantitative and Qualitative Analysis of a Large, Ethnically Diverse School District,” *Psychology in the School*, 2002, 39 (3), 259–277.
- Mustard, David B.**, “Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts,” *Journal of Law and Economics*, April 2001, XLIV, 285–314.
- Neal, Derek and William Johnson**, “The Role of Pre-Market Factors in Black-White Wage Differences,” *Journal of Political Economy*, 1996, 104 (5).

**Schanzenbach, Max**, “Racial and Gender Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics,” *Journal of Legal Studies*, January 2005, *34* (1), 57–92.

**Skiba, Russell J.**, “Zero Tolerance, Zero Evidence: An Analysis of School Disciplinary Practice,” Technical Report, Indiana Education Policy Center August 2000.

**Steffensmeier, Darrell and Stephen Demuth**, “Ethnicity and Sentencing Outcomes in U.S. Federal Courts: Who is Punished More Harshly?,” *American Sociological Review*, October 2000, *65* (5), 705–729.

**Townsend, B.**, “Disproportionate discipline of African American children and youth: Culturally-responsive strategies for reducing school suspensions and expulsions,” *Exceptional Children*, 2000, *66*, 381–391.

**Waldfoegel, Joel**, “The Effect of Criminal Conviction on Income and the Trust Reposed in the Workmen,” *Education and Urban Society*, 1994, *29* (1).

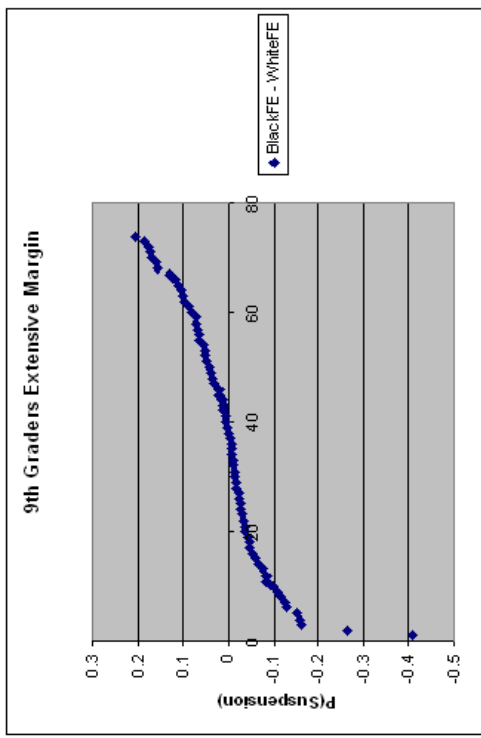
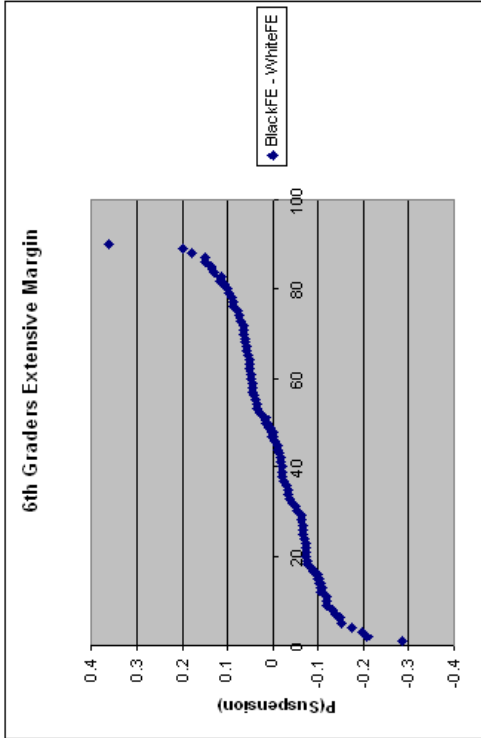
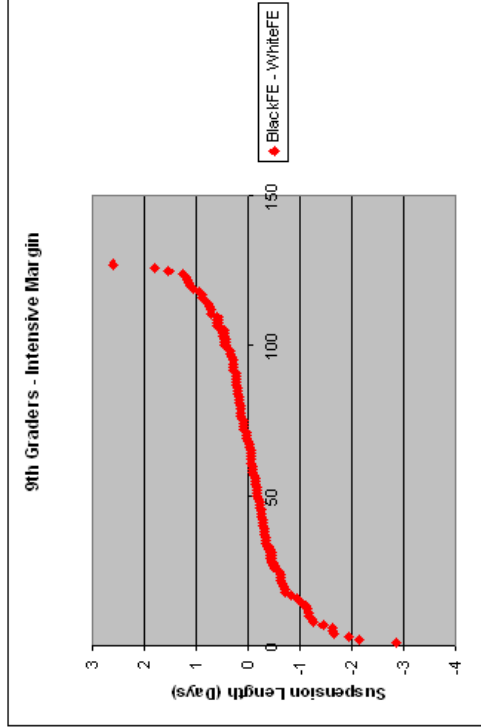
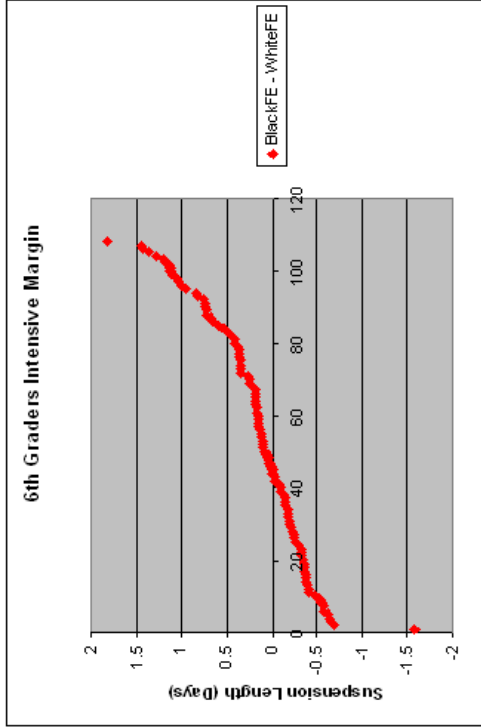


Fig 1: Distribution of Differences in Black-White FE

Table 1: School Infraction Summary

Infraction Type	Quantity	Proportion	Mean Suspension (Days)
Rule violation	32,286	31.78	2.21
Aggressive behavior (Fighting)	28,429	27.98	3.19
Rowdy (Undisciplined)	22,108	21.76	2.38
Other	6,278	6.18	2.21
Truancy	4,038	3.97	2.19
Theft	2,007	1.98	2.89
Possession of a controlled substance	1,408	1.39	6.59
Substance abuse	1,168	1.15	4.46
Possession of a weapon	1,025	1.01	5.59
Property damage	747	0.74	3.06

Note: Unit of observation is an infraction resulting in a short-term suspension from school. A short-term suspension is defined as an out-of-school suspension lasting for ten days or less.

Table 2: Sample Statistics by School Type

	Elementary	Middle	High School
Grades	3-5	6-8	9-12
Schools	568	226	208
Students	148,444	149,558	194,561
Black	0.27	0.27	0.25
Male	0.51	0.51	0.50
Old for Grade	0.21	0.21	0.23
Physically Disabled	0.075	0.064	0.019
Learning Disabled	0.072	0.074	0.038
Free/Reduced Price Lunch	0.43	0.37	-
Lagged Math Score	0.019	0.023	-0.0032
Suspended at least once	0.045	0.13	0.12

Note: Unit of observation is a public school student in North Carolina that satisfies the sample restrictions cited in the text. The lagged score for high school students is their normalized 8th grade test score.

Table 3: Suspended Student Characteristics

	Elementary	Middle	High School
Students	6,718	18,999	20,902
Total Suspensions	1.64	2.02	1.88
Total Days Suspended	3.29	6.23	6.26
Black	0.56	0.47	0.39
Male	0.81	0.70	0.68
Old for Grade	0.39	0.38	0.42
Physically Disabled	0.20	0.15	0.047
Learning Disabled	0.12	0.12	0.063
Free/Reduced Price Lunch	0.71	0.58	-
Lagged Math Score	-0.47	-0.45	-0.35

Note: Unit of observation is a North Carolina public school student who received at least one out of school suspension during the school year.

Table 4: Racial Disparities in Discipline

	Elementary		Middle		High School	
	Black	White	Black	White	Black	White
<b>Extensive Margin</b>						
Pr(Suspension   Discipline Referral)	0.45*	0.35	0.40*	0.32	0.34*	0.33
Pr(Suspension   Rule Violation)	0.23*	0.17	0.28*	0.19	0.30*	0.27
Pr(Suspension   Undisciplined)	0.53*	0.37	0.35*	0.26	0.34*	0.29
Pr(Suspension   Aggressive Behavior)	0.63*	0.54	0.67*	0.58	0.83	0.85
<b>Intensive Margin</b>						
E(Days Suspended   OSS)	2.06*	1.83	2.89*	2.74	2.90*	2.56
E(Days Suspended   OSS, Rule Violation)	1.89*	1.76	2.56*	2.25	2.20*	1.91
E(Days Suspended   OSS, Undisciplined)	1.95*	1.72	2.51*	2.35	2.66*	2.30
E(Days Suspended   OSS, Aggressive Behavior)	2.02*	1.78	3.38*	2.87	4.77*	3.72

Note: \* Indicates that the difference in punishment between black and white students at a particular schooling level is significant at a 5% level. Unit of observation is an individual student infraction in North Carolina. Discipline referral indicates all infraction categories are included. OSS is out of school suspension.

Table 5: Probability of At Least One Suspension

	Elementary	Middle	High School
Black	0.038*	0.066*	0.057*
	(0.003)	(0.005)	(0.005)
Asian	-0.027*	-0.042*	-0.037*
	(0.004)	(0.007)	(0.006)
Hispanic	-0.022*	-0.023*	-0.026*
	(0.003)	(0.005)	(0.006)
Other Non-White	0.011*	0.029*	0.032*
	(0.004)	(0.007)	(0.006)
Male	0.048*	0.087*	0.065*
	(0.002)	(0.004)	(0.003)
Repeat	0.014*	0.121*	0.102*
	(0.005)	(0.012)	(0.007)
Old for Grade	0.020*	0.056*	0.035*
	(0.002)	(0.004)	(0.003)
Physically Disabled	0.053*	0.096*	0.080*
	(0.004)	(0.007)	(0.009)
Learning Disabled	0.010*	0.032*	0.015*
	(0.003)	(0.005)	(0.005)
Lag Math	-0.010*	-0.033*	-0.028*
	(0.001)	(0.002)	(0.001)
Free/Reduce	0.026*	0.059*	-
	(0.002)	(0.003)	
Constant	-0.020*	0.009*	0.112*
	(0.003)	(0.005)	(0.003)
N	148,444	149,558	194,561
School Fixed Effects	Y	Y	Y
Grade Dummies	Y	Y	Y

Note: \* Indicates a coefficient significant at a 5% level. Unit of analysis is a public school student in a North Carolina elementary, middle, or high school. Dependent variable indicates whether student received at least one suspension during the year. Standard errors are clustered at the school level.

Table 6: Probability of At Least One Suspension

Student-Teacher Race Interactions	Elementary	
White Teacher x Black Student	0.046*	0.037*
	(0.002)	(0.002)
Black Teacher x White Student	0.006*	0.005
	(0.003)	(0.003)
Black Teacher x Black Student	0.040*	0.042*
	(0.005)	(0.005)
Male Teacher	0.004	0.006
	(0.004)	(0.004)
Tenure*100	-0.100*	-0.079*
	(0.032)	(0.028)
Tenure <sup>2</sup> *100	0.003*	0.002*
	(0.001)	(0.001)
N	120,844	120,844
School Fixed Effects	N	Y
Grade Dummies	Y	Y
Student Demographics	Y	Y

Note: \* Indicates a coefficient significant at a 5% level. Unit of observation is a North Carolina elementary school student. Dependent variable indicates whether student received at least one suspension during the year. Standard errors are clustered at the teacher level.

Table 7: Extensive and Intensive Discipline Margins for 6th and 9th Graders

Discipline Margin	6th Graders		9th Graders					
	Extensive	Intensive	Extensive	Intensive				
Black	0.073* (0.010)	-0.004 (0.009)	0.251* (0.050)	0.043 (0.047)	-0.008 (0.009)	0.007 (0.008)	0.391* (0.047)	-0.032 (0.047)
Male	0.000 (0.009)	-0.002 (0.007)	-0.049 (0.055)	-0.027 (0.043)	0.011 (0.009)	0.010 (0.007)	0.015 (0.045)	-0.010* (0.038)
Old for Grade	0.025* (0.010)	0.019* (0.008)	-0.018 (0.049)	-0.007 (0.039)	0.021* (0.010)	0.016* (0.007)	0.101* (0.047)	0.031 (0.040)
Physical Disability	0.064* (0.015)	0.061* (0.011)	-0.213* (0.066)	-0.244* (0.053)	0.034 (0.019)	0.031* (0.015)	-0.195* (0.088)	-0.102 (0.074)
Learning Disability	-0.010 (0.014)	-0.012 (0.011)	-0.108 (0.068)	-0.048 (0.052)	-0.020 (0.015)	-0.016 (0.012)	-0.124 (0.068)	-0.060 (0.059)
Free/Reduce Lunch	0.032* (0.010)	0.023* (0.008)	0.139* (0.052)	0.104* (0.045)				
Lagged Math	0.017* (0.006)	0.003 (0.005)	0.007 (0.029)	-0.007 (0.025)	-0.022* (0.006)	-0.008 (0.005)	-0.001 (0.030)	0.015 (0.026)
Previous Offenses	0.020* (0.002)	0.030* (0.002)	0.037* (0.008)	0.083* (0.008)	0.016* (0.002)	0.027* (0.002)	0.045* (0.009)	0.096* (0.009)
Multiple Offenses	0.069 (0.043)	0.100* (0.040)	0.801* (0.258)	0.891* (0.234)	0.207* (0.036)	0.273* (0.039)	0.824* (0.159)	1.081* (0.154)
N	18,532	10,626	18,812	14,428				
School FE	N	Y	N	Y	N	Y	N	Y
Infraction Type	Y	Y	Y	Y	Y	Y	Y	Y

Note: \* Indicates a coefficient significant at a 5% level. Unit of observation is a student infraction committed by a 6th or 9th grader in a new school. The dependent variable for the extensive margin is an indicator for whether the student was suspended. The dependent variable for the intensive margin is the length of suspension in days. Standard errors are clustered at the student level.

Table 8: Heterogeneity in Punishment - Race Interactions

Discipline Margin	6th Graders		9th Graders	
	Extensive	Intensive	Extensive	Intensive
Black	-0.015 (0.018)	0.125 (0.104)	0.009 (0.012)	-0.127 (0.075)
Black x Male	0.008 (0.015)	0.009 (0.094)	-0.001 (0.014)	0.085 (0.078)
Black x Lag Math	-0.014 (0.009)	0.041 (0.045)	0.004 (0.009)	-0.070 (0.049)
Black x Free/Reduced	-0.004 (0.017)	-0.104 (0.082)		
N	18,532	10,626	18,812	14,428
School FE	Y	Y	Y	Y
Infraction Type	Y	Y	Y	Y
Other Demographic Characteristics	Y	Y	Y	Y

Note: \* Indicates a coefficient significant at a 5% level. Unit of observation is a student infraction committed by a 6th or 9th grader in a new school. The dependent variable for the extensive margin is an indicator for whether the student was suspended. The dependent variable for the intensive margin is the length of suspension in days. Standard errors are clustered at the student level.

Table 9: Heterogeneity in Punishment - Offense Interactions

Discipline Margin	6th Graders		9th Graders	
	Extensive	Intensive	Extensive	Intensive
Rule Violation x Black	-0.017 (0.013)	0.007 (0.084)	0.005 (0.011)	-0.160* (0.056)
Other Offense x Black	-0.004 (0.043)	-0.039 (0.155)	-0.020 (0.035)	-0.001 (0.165)
Undisciplined x Black	0.012 (0.014)	-0.070 (0.070)	0.025 (0.017)	-0.138 (0.092)
Theft x Black	0.028 (0.053)	0.151 (0.233)	-0.087 (0.064)	-0.046 (0.313)
Fighting x Black	0.000 (0.016)	0.131* (0.064)	0.024 (0.021)	0.418* (0.113)
Truancy x Black	-0.110* (0.035)	1.016* (0.257)	-0.021 (0.020)	-0.216 (0.117)
Property Damage x Black	0.075 (0.111)	-0.622 (0.564)	-0.101 (0.130)	0.087 (0.681)
Major x Black	-0.134* (0.063)	-0.395 (0.413)	0.021 (0.047)	0.275 (0.317)
N	17,735	9,984	17,499	13,399
School Fixed Effects	Y	Y	Y	Y
Infraction Type	Y	Y	Y	Y
Other Demographic Characteristics	Y	Y	Y	Y

Note: \* Indicates a coefficient significant at a 5% level. Unit of observation is a student infraction committed by a 6th or 9th grader in a new school. Standard errors are clustered at the student level.

Table 10: Testing for Bias in Conditioning Variables

	6th Graders				9th Graders			
<b>Extensive Margin</b>								
Black	-0.004 (0.009)	0.007 (0.009)	0.002 (0.010)	0.012 (0.010)	0.007 (0.008)	0.015 (0.008)	0.019* (0.009)	0.026* (0.009)
Previous Offenses	Y	N	Y	N	Y	N	Y	N
Infraction Type	Y	Y	N	N	Y	Y	N	N
<b>Intensive Margin</b>								
Black	0.043 (0.047)	0.071 (0.047)	0.018 (0.050)	0.037 (0.050)	-0.032 (0.047)	-0.007 (0.048)	-0.007 (0.051)	0.003 (0.051)
Previous Offenses	Y	N	Y	N	Y	N	Y	N
Infraction Type	Y	Y	N	N	Y	Y	N	N

Note: \* Indicates a coefficient significant at a 5% level. Unit of observation is a student infraction committed by a 6th or 9th grader in a new school. Each coefficient is from a separate regression that also includes school fixed effects and other observable student characteristics. Standard errors are clustered at the student level.

Table 11: Extensive and Intensive Discipline for All Grades

Grades	Elementary 3-5		Middle 6-8		High School 9-12		
<b>Extensive Margin</b>							
Black	0.099* (0.014)	0.015 (0.012)	0.077* (0.005)	-0.005 (0.005)	0.012* (0.005)	0.009 (0.005)	
N	8,842	8,842	65,148	65,148	53,681	53,681	
School FE	N	Y	N	Y	N	Y	
Infraction Type	Y	Y	Y	Y	Y	Y	
<b>Intensive Margin</b>							
Black	0.228* (0.034)	0.055 (0.034)	0.260* (0.030)	-0.009 (0.028)	0.428* (0.027)	0.074* (0.027)	
N	10,879	10,879	37,860	37,860	38,769	38,769	
School FE	N	Y	N	Y	N	Y	
Infraction Type	Y	Y	Y	Y	Y	Y	

Note: \* Indicates a coefficient significant at a 5% level. Unit of observation is a student infraction committed by a public school student. The dependent variable for the extensive margin is an indicator for whether the student was suspended. The dependent variable for the intensive model is the length of suspension in days. Each coefficient is from a separate regression that also includes other observable student characteristics. Standard errors are clustered at the student level.

Table 12: Extensive Discipline Margin - Principal-Student Race Interactions

	Elementary	Middle	High School
White Principal x Black Student	0.050*	0.037*	0.003
	(0.025)	(0.012)	(0.008)
Black Principal x Black Student	0.230*	-0.004	-0.028*
	(0.044)	(0.014)	(0.014)
Black Principal x White Student	0.284*	-0.022	-0.136*
	(0.078)	(0.012)	(0.012)
Proportion Black in School	0.333*	0.359*	-0.305*
	(0.055)	(0.031)	(0.027)
Mean Lagged Math Score in School	-0.019	0.145*	-0.326*
	(0.044)	(0.025)	(0.024)
N	3,345	19,115	27,132
Infraction Type Dummies	Y	Y	Y
Other Student Demographics	Y	Y	Y
F-statistics and p-values testing student-principal coefficients			
$\beta_{WB} = \beta_{BB}?$	20.8	8.57	5.03
	(0.00)	(0.00)	(0.03)
$\beta_{BB} = \beta_{BW}?$	0.40	1.54	42.79
	(0.53)	(0.21)	(0.00)
$\beta_{BB} - \beta_{BW} = \beta_{WB}?$	1.48	1.43	33.72
	(0.22)	(0.23)	(0.00)

Note: \* Indicates a coefficient significant at a 5% level. Unit of observation is a student infraction committed by a public school student. The dependent variable is an indicator for whether the student was suspended. The sample is limited to schools whose principal's race can be identified. Standard errors are clustered at the student level.  $\beta_{WB}$  is the interaction coefficient for a white principal paired with a black student. Other  $\beta$ 's are defined similarly.

Table 13: Intensive Discipline Margin - Principal-Student Race Interactions

	Elementary	Middle	High School
White Principal x Black Student	0.129*	0.101*	0.236*
	(0.071)	(0.064)	(0.041)
Black Principal x Black Student	0.132	0.095	0.097
	(0.126)	(0.083)	(0.084)
Black Principal x White Student	0.007	0.120	0.066
	(0.165)	(0.088)	(0.092)
Proportion Black in School	0.895*	1.969*	0.661*
	(0.191)	(0.155)	(0.112)
Mean Lagged Math Score in School	0.286	0.857*	-0.011
	(0.155)	(0.140)	(0.094)
N	2,945	10,168	18,917
Infraction Type Dummies	Y	Y	Y
Other Student Demographics	Y	Y	Y
F-statistics and p-values testing student-principal coefficients			
$\beta_{WB} = \beta_{BB}?$	0.00	0.01	3.03
	(0.97)	(0.94)	(0.08)
$\beta_{BB} = \beta_{BW}?$	0.47	0.06	0.07
	(0.49)	(0.80)	(0.79)
$\beta_{BB} - \beta_{BW} = \beta_{WB}?$	0.00	1.36	2.95
	(0.99)	(0.24)	(0.09)

Note: \* Indicates a coefficient significant at a 5% level. Unit of observation is a student infraction committed by a public school student. The dependent variable is the length of suspension in days. The sample is limited to schools whose principal's race can be identified. Standard errors are clustered at the student level.  $\beta_{WB}$  is the interaction coefficient for a white principal paired with a black student. Other  $\beta$ 's are defined similarly.