

UNPACKING THE DRIVERS OF RACIAL DISPARITIES IN
SCHOOL SUSPENSION AND EXPULSION

Jayanti Owens
Brown University

Sara S. McLanahan
Princeton University

Fragile Families Working Paper WP18-04-FF

UNPACKING THE DRIVERS OF RACIAL DISPARITIES IN SCHOOL SUSPENSION AND EXPULSION

Jayanti Owens
Brown University

Sara S. McLanahan
Princeton University

February 2018*

ABSTRACT

School suspension and expulsion are important forms of punishment that disproportionately affect Black students. Punishment early in schooling has consequences for long term achievement, criminal justice interaction, and wellbeing. Prior research identifies three mechanisms that help account for racial disparities in suspension and expulsion: between-school sorting, differences in student behaviors, and differential treatment of students with the same behaviors. We build upon and extend prior research in four ways: (1) by comparing the relative importance of the three mechanisms in a single study, (2) by focusing on students early in elementary school, (3) by measuring students' behavior at the time they enter school, before they have been exposed to school disciplinary practices, and (4) by using both teacher and parent reports of student behaviors. Using data from the Fragile Families Study and decomposition techniques, we find that between-school sorting accounts for 13% of the Black/White gap in suspension/expulsion, differences in behavior account for 9% of the gap, and differential treatment accounts for 39% of the gap. Behavior differences measured at age 9, which are likely to be endogenous to suspension and expulsion, reinforce the importance of differential punishment. Theoretically, our findings highlight differential treatment as a mechanism of early criminalization.

*This replaces earlier versions of this manuscript from 2016 and 2017.

INTRODUCTION

Over the last forty years, schools' use of exclusionary discipline tactics, such as suspension and expulsion, increased by nearly 50%. In 1980, 12% of 8th-10th grade students reported having been suspended at some point in their lives. By 2006, this figure had increased to 18% (Bertrand and Pan 2013). In the 2011-12 academic year alone, 10% of students in kindergarten through 12th grade, totaling more than 3.5 million children, were suspended outside of school or expelled (Losen et al. 2015). The incidence rate is over 2.5 times greater (over 25%) when in-school suspensions are counted (Losen et al. 2015; Wallace et al. 2008).

These aggregate rates mask considerable heterogeneity by race and gender. Minority students, particularly boys and those who attend schools with minority and low-income populations, are three times more likely to be suspended or expelled than other students (Ramey 2015; Skiba and Williams 2014). National incidence rates for the 2011-12 academic year alone indicate that 39% of Black boys in kindergarten through 12th grade experienced in-school or out-of-school suspension or expulsion as compared to 15% of White boys (authors' calculations are from the U.S. Department of Education's Office for Civil Rights Data Collection, 2011-12). Racial gaps between Black and White girls were equally striking, though absolute rates were roughly half those of boys (Morris and Perry 2017; Office for Civil Rights 2014).

The disproportionate suspension and expulsion of Black students shapes racial gaps in educational attainment, interaction with the criminal justice system, and other indicators of economic and social wellbeing. Using a range of datasets, estimation strategies, and outcomes, researchers have shown that being suspended or expelled from school – and the associated negative labeling plus loss of instruction time – is associated with poor school performance and a higher risk of school dropout, arrest, incarceration and unemployment (Andrew and Blake 2017;

Fabelo et al. 2011; Mittleman 2017; Wolf and Kupchik 2017). Studies using within-student variation, which controls for stable unobserved risk factors for suspension, reach similar conclusions (Morris and Perry 2016).

Some scholars have pointed to the existence of a ‘school-to-prison pipeline’ to highlight the parallel between school suspension and incarceration at both the aggregate and individual levels (Gregory, Skiba and Noguera 2010; Wald and Losen 2003). These increasingly intertwined forms of social control disproportionately affect the lives of Black males, their families and their communities (Hirschfield 2008; Kupchik et al. 2009; Pager 2003; Perry and Morris 2014). Despite the obvious link between exclusionary school discipline and incarceration, empirical research on the causes of racial disparities in these two domains remains somewhat siloed. Notably, whereas differential treatment on the basis of race (e.g., racial discrimination) is widely accepted as an important cause of racial disparities in police stops, arrests, use of force, and judicial sentencing in the criminal justice system, this explanation has received much less attention in the literature on disparities in school suspension and expulsion. Instead, sociologists have focused primarily on structural discrimination in the form of differences in the composition and characteristics of the schools that Black and White children attend, differences in student behaviors and, to a lesser extent, differences in children’s family backgrounds (Skiba et al. 2002; Skiba and Williams 2014). Moreover, the handful of studies that examine differential treatment as a predictor of suspension and expulsion have not attempted to assess how much of the racial gap in suspension is due to this mechanism as compared to other explanations.

The most compelling evidence for the differential treatment mechanism comes from a laboratory experiment by psychologists who use written vignettes describing behavioral infractions to examine racial differences in how teachers perceive children’s behaviors and

differences in the punishments they recommend (Okonofua and Eberhardt 2015). This study finds that teachers perceive minority students' behavior more negatively and sanction it more harshly.¹ Although compelling, this study does not tell us whether the results obtained in the laboratory extend to the general population; nor does it tell us how much of the racial gap is due to differential treatment as opposed to other explanations.

Understanding the relative importance of different explanations for the racial gap in suspension and expulsion is important, not only for developing theories about the processes underlying racial disparities but also for developing effective policies and practices. Although not mutually exclusive, arguments about student behavior would suggest that training teachers to more effectively manage student behaviors may serve as an important entry point, whereas arguments that emphasize school composition and policies would point to rethinking disciplinary systems in schools serving minority and poor populations as a starting place. By contrast, arguments that point to differential treatment (e.g., racial discrimination or “tough love”) by school officials of children with the same behaviors in similar schools would suggest that increasing educator empathy around the effects of structural disadvantage, negative stereotyping, and (often implicit) bias would be efficacious (Okonofua, Paunesku and Walton 2016).

In this paper we examine the relative contributions of these three mechanisms for explaining race differences in school suspension and expulsion. As far as we know, ours is the first study to examine and assess all three mechanisms together. We use data from the Fragile Families and Child Wellbeing Study, a population-based, birth cohort study of children born in large U.S. cities at the turn of the 21st century. These data include a large sample of children born to unmarried parents, who are disproportionately from minority and disadvantaged backgrounds

and more likely to attend schools with high rates of suspension and expulsion. The children in the sample are spread across roughly 1,066 elementary schools in 350 counties.

We focus on Black and White boys. In addition to the fact that Black males experience the highest rates of both suspension/expulsion and criminal justice contact, research on intersectionality suggests that the mechanisms described above may operate differently for boys and girls (Collins 2015; Goff et al. 2014). We suspect that mechanisms also may operate differently for other ethnic groups (e.g., Latinos and Native Americans) (Dill and Zambrana 2009; Morris and Perry 2017; Office for Civil Rights 2014; Skiba et al. 2002).

We begin by examining children's behaviors at the time they enter kindergarten and before they have been exposed to primary school disciplinary practices. Research indicates that missing school during elementary school has lasting effects on children's academic achievement (Forrest et al. 2013; Ginsburg, Jordan and Chang 2014; Morris and Perry 2016). Consequences are more severe when missed days are due to suspension or expulsion, when negative labeling may further exacerbate academic consequences (Ferguson 2001; Okonofua, Walton and Eberhardt 2016; Rios 2011). The earlier in school a suspension occurs, the longer the time horizon over which the associated disadvantages can accumulate, leading to heightened risk of school dropout, interaction with the criminal justice system, family instability, and unemployment (DiPrete and Eirich 2006; Rios 2011).

We use parent as well as teacher reports of children's behavior and we focus on overall behavior rather than a specific infraction. Our approach has several advantages over studies that rely on teacher reports of a specific infraction. First, ethnographic research suggests that a teacher's decision to refer or not refer a child for suspension is often less about the specific infraction at hand and more about the teacher's overall impression of the child's general behavior

(Ferguson 2001; Rios 2011; Vavrus and Cole 2002). Second, the use of multiple reporters (teachers and parents) provides insight into differences in how parents and teachers view children and how this varies by child race (Okonofua, Walton and Eberhardt 2016). And finally, using parent as well as teacher reports protects against teacher stereotype bias toward students from low income and minority backgrounds (Okonofua and Eberhardt 2015).

Consistent with prior research, we find that the concentration of Black boys in schools serving low-income, minority students accounts for roughly 13% of the racial gap expulsion by the time children are nine years old (approximately third grade), while differences in boys' behavior at school-entry account for 9% percent of the gap. Most notably, we find that differential treatment of Black and White boys exhibiting similar behaviors and attending schools with similar compositions accounts for 39% of the Black/White gap in boys' suspension.

If black boys' behaviors worsen more than white boys' after school entry, our model might overestimate the role of differential treatment in accounting for race differences in suspension and expulsion. To account for this possibility, we re-estimate our model, using teacher and parent reports of behaviors at the end of our observational period, when the child is age nine. Because differences in age nine behaviors may be endogenous to suspension or expulsion, this second model should provide an upper-bound estimate of the role of race differences in boys' behavior. Even so, using the age 9 measures of behavior, we continue to find robust evidence for the differential treatment hypothesis. Ultimately, we argue that differential treatment accounts for a larger share of the racial gap than either between-school sorting or racial differences in boys' behavior and merits greater attention from researchers who study racial disparities in school suspension.

EXPLANATIONS FOR THE RACIAL GAP IN SUSPENSION AND EXPULSION***Differences in School Composition***

The most widely accepted explanation for racial disparities in suspension and expulsion is *between-school sorting* (Welch and Payne 2010). According to this argument, schools serving minority and low-income students are more likely than other schools to adopt ‘zero tolerance policies’ for dealing with student misbehavior. In their study of 294 public schools, Welch and Payne (2010) use principals’ judgements about “how often” their school uses various punitive or non-punitive approaches to handle student misconduct and find that schools with large enrollments of Black students are more likely to use zero tolerance and other exclusionary discipline practices than schools with large enrollments of White students. Importantly, these authors hold constant average levels of student delinquency and the percentage of students receiving free-or-reduced-price lunches at the school level, suggesting that there is something unique about schools that enroll high percentages of minority students above and beyond the fact that their students are disproportionately from low income families (Welch and Payne 2010).

Related research makes the additional claim that the combination of serving *both* poor and minority students leads to the highest levels of suspension and expulsion because these schools are the least likely to have the resources and supports necessary to help teachers handle misbehavior effectively (Gregory, Skiba and Noguera 2010).² For example, Ramey (2015) finds that schools serving either majority-minority or a combination of majority-minority and poor students are more likely to use proactive *exclusionary* discipline tactics like suspension and expulsion or arrest. By contrast, in schools serving predominantly low-income *White* students, Ramey (2015) and Hinshaw and Scheffler (2014) are more likely to pressure working-class parents to seek ADHD and other behavioral diagnoses for their children in order to qualify for

Individualized Education Plans (IEPs) and special education services through the Individuals with Disabilities Education Act (IDEA). Qualification for services through IDEA helps schools alleviate the pressure of high-stakes testing and the threat of school takeover or closure, both of which are particularly acute in low-performing schools (Jacob 2005). Finally, the claim that between-school differences in approaches to handling misbehavior are correlated with school racial and socioeconomic composition is further supported by work explicitly testing racial gaps in suspension within and across schools. Kinsler (2011) finds that the racial gap in school suspension *conditional on referral* is due to differences *between* schools. Racial gaps, then, are exacerbated by residential segregation, whereby minority students are systematically sorted into more punitive schools. Based on the association between school composition and the use of punitive disciplinary tactics, we hypothesize that:

[Hypothesis 1:] Race differences in school composition (i.e., percent minority and percent poor) account for a large share of the racial gap in suspension/expulsion by year 9.

Differences in Children's Behavior

An alternative perspective argues that racial differences in suspension and expulsion are due to *differences in children's behaviors*, such as rule-following, attention, self-regulation, aggression, and the ability to get along with peers and teachers (Gregory, Skiba and Noguera 2010; Raffaele-Mendez 2003). Racial differences in children's behavior are well documented (Entwisle and Alexander 1993; Entwisle, Alexander and Olson 2005; McLeod and Nonnemaker 2000; Wright et al. 2014) and result from differences in exposure to stressful environments (e.g., violence), variation in parenting styles, and differences in pre-school experiences (Bates et al.

1991; Brooks-Gunn and Duncan 1997; Dance 2002; Deater-Deckard and Dodge 1997; Magnuson and Waldfogel 2005; Robinson 2014).³

Existing research provides conflicting evidence about the extent to which student behaviors account for the racial gap in suspension and expulsion, with some studies finding that behavior accounts for almost *none* of the gap (Skiba et al. 2014) and other studies finding that it accounts for nearly *all* of the gap (Wright et al. 2014). This discrepancy is driven by differences in data sources and measures of behaviors across studies, as well as by modeling strategies. Studies that condition on being referred for sanction and model the association between the severity of sanctions and suspension/expulsion typically find that differences in behavior account for very little of the racial gap in punishment (Gregory, Skiba and Noguera 2010; Skiba et al. 2014; Skiba et al. 2002; Skiba and Williams 2014). Note, however, that if Black boys are more likely than White boys to be referred for less severe infractions, as some studies suggest (Kinsler 2011; Skiba et al. 2002; Skiba and Williams 2014), conditioning on referral would bias the association between race and severity of sanction toward zero.

In contrast, studies that use population-based data and do not condition on referral find that student behavior is strongly associated with suspension and expulsion (Bradshaw et al. 2010; Rocque 2010). Raffaele-Mendez (2003), for example, finds that teachers' ratings of students' attention, school attitudes, and classroom behavior in grades 3 through 5 are strong predictors of 6th grade out-of-school suspension for both Black and White students. Most strikingly, Wright et al. (2014) find that racial differences in behaviors at school entry and third grade account for *all* of the gap in suspension by 8th grade.⁴ Importantly, both of these studies include measures of behavior *after* the child enters school, which raises questions about the causal ordering of the child's behavior and school punishment. Insofar as behavior is endogenous to how children are

treated by teachers and school officials, and insofar as Black boys are treated more punitively than White boys, the studies described above would overstate the extent to which racial differences in behavior account for differences in suspension and expulsion (Kinsler 2011; Mittleman 2017; Okonofua, Walton and Eberhardt 2016; Okonofua and Eberhardt 2015; Okonofua, Paunesku and Walton 2016). That is, if Black boys' behavior worsens than White boys' as a response to increased punishment, behavior measured after school entry will overstate behavior differences in accounting for the racial gap in suspension. Given that we measure behavior prior to suspension and given that we do not condition on referral, we hypothesize that:

[Hypothesis 2:] Race differences in children's behavior problems account for more of the racial gap in suspension/expulsion at year 9 than what is suggested by studies that condition on referral but less of the gap than what is suggested by studies that measure behavior after the child has been exposed to school discipline.

Differences in the Treatment of Black Boys

Finally, the *differential treatment perspective* argues that racial differences in suspension and expulsion occur because teachers and school officials impose harsher punishment for Black boys with the same behaviors and the same school composition as White boys (Okonofua, Paunesku and Walton 2016). We draw a distinction between unequal treatment that results from “structural discrimination,” such as the greater use of exclusionary discipline in minority-serving schools (e.g., between-school sorting) and unequal treatment that results from “differential treatment/interpersonal discrimination.” The latter may result from varying cognitive mechanisms, including implicit or explicit interpersonal discrimination fueled by racial stereotypes, or from a desire to practice “tough love,” in which advocates of students of color

may assign harsher sanctions in an effort to help prepare students for the realities of an unjust society (Pager and Shepherd 2008).

The best evidence for *differential treatment* comes from laboratory experiments where teachers were asked to rate the severity and appropriateness of sanctions for identical misbehaviors among Black and White boys (Gilliam et al. 2016; Okonofua and Eberhardt 2015). To study discrimination, Okonofua and Eberhardt (2015) presented teachers with short, written vignettes about student behaviors and asked them to assess the behaviors and match them with sanctions. To signal race, they used racially coded names. They found that teachers viewed behaviors as more negative and recommended harsher sanctions when the student had a racially-coded 'Black' name. The idea that certain teachers are more likely to recommend harsher sanctions for Black boys as compared with White boys has also been tested outside the laboratory. In line with the notion of "tough love," Lindsay and Hart (2017), for example, found that White teachers were more likely than Black teachers to punish Black students. However, lacking prospective measures of student behavior prior to any suspension, this study cannot rule out the possibility that Black students behave better around Black teachers than White teachers (Egalite and Kisida 2017).

In elementary school, most misbehavior that culminates in referral and suspension is relatively minor, consisting of defiance, disruption, or noncompliance (Ferguson 2001; Lindsay and Hart 2017).⁵ Importantly, research shows that it is precisely with these relatively minor forms of misbehavior that discretion over whether to refer or recommend for punishment is greatest (Dovidio and Gaertner 2004). When it comes to misbehavior that is of questionable levels of severity, teachers may be more likely to rely on stereotypes to guide their decisions, which can lead to unequal treatment by race. Although the expression of racist attitudes is less

common today than it was in the past, culturally embedded racial stereotypes that depict Black boys as ‘lazy,’ ‘violence-prone,’ and ‘aggressive,’ continue to exist (Bobo and Kluegel 1997; Ferguson 2001; Pager and Shepherd 2008). These stereotypes may be conscious or unconscious, and may be based on implicit bias, animus, or more apathetic statistical discrimination (Allport 1979; Ayres and Siegelman 1995). Even statistical discrimination on the part of educators involves using stereotypes as cognitive shortcuts for instrumental purposes like efficiency when filling in missing information on behavioral intent. Educators may find it more efficient in situations of ambiguous or minor misbehavior to make predictions about the appropriate disciplinary response based on racial stereotypes about behavioral intent (Ferguson 2001). Whatever its cognitive basis, stereotyping may lead to differential treatment in the form of disciplinary decision-making. Building on findings described above, we hypothesize that:

[Hypothesis 3:] Race differences in the disciplinary treatment of Black boys with the same behaviors at school entry, the same family socioeconomic resources and the same school contexts explain a large share of the racial gap in suspension/expulsion at year 9.

Taken together, our study goes beyond prior work in several ways. To the best of our knowledge, we are the first to parse the relative contributions of the three key explanations for the racial disproportionality in school suspension and expulsion. By focusing on children in elementary school, we are able to shed light on a key part of the life course when behavior and behavioral labels take hold and lay the groundwork for possible suspension and cumulative (dis)advantage trajectories in school. Additionally, we examine the likelihood of suspension and expulsion *unconditional* on having been referred and suspended⁶ and measure children’s behaviors at the time they enter school, before they have been suspended/expelled and/or labeled

as ‘troublemakers’ by teachers and school administrators (Okonofua, Walton and Eberhardt 2016).⁷ Finally, our study draws on both teacher and parent reports of child behaviors, providing a more comprehensive picture of children’s behaviors (Ferguson 2001; Gilliam et al. 2016).⁸

DATA AND METHODS

Data and Sample

Our study draws on data from the Fragile Families and Child Wellbeing Study (FFCWS), a longitudinal birth cohort study of 4,898 children born between 1998 and 2000 in 20 U.S. cities with populations greater than 200,000 (Reichman et al. 2001).⁹ Baseline interviews with mothers and most fathers were conducted in hospitals shortly after the child’s birth. The baseline response rates were 86 percent for mothers and 79 percent for fathers (conditional on enrolling the mother). Follow-up phone interviews with both parents were conducted when the child was approximately one, three, five, and nine years old. In-home observations of the child and the home environment as well as interviews with the primary caregiver were also conducted beginning with the year 3 follow-up survey. Additionally, and essential for our purposes, teachers were surveyed at years 5 and 9 about the focal child’s behaviors and achievement. School administrative data were also collected about the school context at year 9. Finally, at year 9, the child was directly interviewed about their school and home experiences, including ever having been suspended/expelled.

The FFCWS is well-suited for our study of the drivers of racial disparities in elementary school suspension for four reasons. First, the availability of school suspension information at this early age is unique to the FFCWS; other large, contemporary datasets that follow children through elementary school – such as the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K) – do not contain information on school suspension and expulsion until eighth

grade, if at all. Second, the FFCWS data are unique in that they contain a relatively large number of Black children, which is essential given our focus on racial disparities and given that suspension and expulsion are relatively rare events in elementary school. Third, the FFCWS data are also unique in that they originate at the child's birth and continue to trace children, their families, and their schools through elementary school. Other datasets that follow children from birth, such as the ECLS-Birth Cohort (ECLS-B), terminate at kindergarten.

Interviews with parents and the primary caregiver (usually the mother) provide detailed information on family, home, and early care contexts, family socioeconomic status, and family structure since birth. This information is essential for isolating the role of school context and behavior independent of family socioeconomic resources and family structure instability. In addition to the well-documented role of family economic resources in shaping school selection, an experiment that manipulated family structure found that teachers were significantly more likely to recommend suspension for boys without a father at home (Santrock and Tracy 1978). Whereas teachers and school officials may *not* have explicit information about a child's family situation, they often have a sense of which children are receiving free-or-reduced-price-lunch (a proxy for family income) or are experiencing divorce at home. Calarco (2011), for example, finds that guidance counselors often hold "lunch bunch" group for kids going through family breakup. Finally, Dallaire, Ciccone and Wilson (2010) find that children with incarcerated parents are seen as being less competent and more prone to academic failure than those without exposure to parental incarceration, suggesting that this form of family structure instability may also influence child suspension. Whatever limited information teachers and schools have about families, we want to isolate the role of between-school sorting and behavior differences

independent of differences in family socioeconomic and structure resources (Lewis and Diamond 2015).

Fourth, we combine the rich information on family and home contexts with detailed *prospective* reports from teachers and parents about the child's behavioral development both at school entry (year 5) and at the end of the observational period (year 9), or roughly third grade. Behavior reports at school entry clearly predate any elementary school suspension and any labeling on the part of teachers that may occur. By contrast, reports at the end of the observational period occur after any suspension and therefore are likely to be contaminated by the negative behavioral reputations that form quickly and are disproportionately sticky for black boys (Okonofua and Eberhardt 2015). Thus, behavior reports at the end of the observation period provide upper-bound estimates of the role of behavior differences in accounting for the racial gap in suspension.

The year 9 sample includes 3,515 children, about 72 percent of the FFCWS sample at baseline.¹⁰ Because detailed information on school demographic and discipline context are essential for accurately identifying the role of school factors in contributing to the racial gap in suspension (hypothesis 1), we merge into the year 9 FFCWS sample additional variables from two administrative data sources using the National Center for Education Statistics (NCES) school identification numbers: (1) the 2008-2010 Common Core of Data (CCD)/Private School Universe Survey (PSS), and (2) the 2009-2013 Civil Rights Data Collection (CRDC) data, both collected through the NCES of the U.S. Department of Education. Using NCES school ID numbers, we successfully matched all 3,515 children in the year 9 FFCWS sample to one of the 2,743 schools listed in the CCD and CRDC.¹¹

Of the 3,515 children in the year 9 FFCWS sample, our primary outcome variable, child-reported ever suspended or expelled at year 9, was missing for 176 (or 5% of) children. We exclude this 5% of children from our analysis. However, to address item-missingness on our predictors among the 3,339 children with complete information on our outcome and our school context variables from the CCD at year 9, we use multiple imputation of 20 datasets based on the MI suite in Stata 14. We included our outcome variable of suspension/expulsion in the imputation equation, but then drop the 5% of boys with imputed values for suspension from our analytic sample before conducting our analyses, following Von Hippel (2007). On all predictors other than teacher ratings of child behavior at year 5, item-missingness was no more than 19 percent of the 3,515 children in the year 9 sample (see the Appendix and Appendix Table A.1 for a discussion of missingness on teacher ratings of behavior at year 5).

Building on work by Collins (2015), Goff et al. (2014), and others, we believe that intersectional processes across lines of race and gender may lead to differing mechanisms of the Black/White racial gap between boys and girls. In this study, therefore, we restrict our analytic sample to Black (n=880) and White (n=368) *boys* with complete information on our outcome variable at year 9.

Measures

Elementary School Suspension/Expulsion: The dependent variable in our analysis is a binary indicator based on a yes/no question asked of the child during the year 9 survey (roughly third grade): “Have you ever been suspended or expelled from school?” A comparison to rates of in-school and out-of-school suspension using national data from the U.S. Department of Education suggests that this measure includes both in-school and out-of-school suspension as well as expulsion. However, because expulsion is a very rare event (affecting roughly 1% of U.S.

children in kindergarten through third grade), for brevity we refer to our outcome as primarily capturing in- and out-of-school suspension (“suspension” for short).

Certainly, selection into suspension is not random. However, by looking at suspensions in the first three years of K-12 schooling, we are able to be more mindful to issues of selection into suspension because all children start this period without a K-12 suspension record. Because any suspension that does occur is necessarily a first-time primary school suspension, we avoid the potential of omitted variables bias that would arise if our estimates of suspension during our observational period were to be confounded by the heightened risk of repeat suspension that results when some children have pre-existing suspension records (Okonofua and Eberhardt 2015). Although children technically could be suspended multiple times over this period, we cannot model the numbers of suspensions. However, most children who are suspended in early elementary school are only suspended once. Because early suspensions lay the ground work for cumulative disadvantages over the schooling years, capturing "ever being suspended" during these early years is important (DiPrete and Eirich 2006).

Child Race: Child race is a dummy variable coded 1 for “Black”, 0 for “White” and is derived from the baseline mother survey. For 93% of children, both the biological mother and father identified as “Black, non-Hispanic.” In 4% of cases, the biological mother identified as “Black, non-Hispanic” whereas the father identified as “Hispanic.” In less than 1% of cases where the mother identified as “Black, non-Hispanic” did the father identify as “White, non-Hispanic.” Almost all of these children were likely to have been identified as Black by the educators in charge of making disciplinary decisions. For children classified as White, both biological parents identified as “White, non-Hispanic” in 76% of cases. In 10% of cases where the mother identified as “White, non-Hispanic”, the biological father identified as “Black, non-

Hispanic.” In another 10% of cases he identified as “Hispanic,” and in 4% of cases he identified as “Other.” Results excluding the 24% of children we classify as White and the 7% of children we classify as Black but who may in fact be multiracial do not change substantive findings and are available upon request.

School Racial/Ethnic and Socioeconomic Composition: Based on prior research, we operationalize school racial and socioeconomic composition in two ways. First, prior research highlights that suspension/expulsion is more widely used in schools with large enrollments of students from minority or low-income backgrounds. Second, other research suggests that there is an additional disadvantage for children who attend schools that serve a majority of students who are *both* poor and minority. Based on the results of analyses discussed in the Appendix and displayed in Appendix Table A.2., we operationalize school racial and socioeconomic composition using a set of dummy variables that equal 1 if the student attends a school in the top 50% of: Black and Hispanic enrollments, or the top 50% of free or reduced-price lunch enrollment (FRPL), or the top 50% of both. Results are robust to the use of continuous measures of school racial and FRPL composition.

Child Behavior Problems (Parent and Teacher Reports of Externalizing Problems): We measure child behavior using separate parent and teacher reports of the child’s externalizing problems at year 5/kindergarten and year 9, using items from the Child Behavior Checklist (CBCL). In the main part of our analyses, we rely only on year 5/kindergarten behaviors as these are causally prior to any suspension. Both teacher and parent CBCL scales are strongly predictive of child suspension/expulsion (Bradshaw et al. 2010; Raffaele-Mendez 2003). We include both teacher and parent reports separately because the correlation between the two scales is relatively low ($r=0.23$ at year 5 and $r=0.38$ at year 9). Parents are more familiar with their

child's behavior at year 5/kindergarten entry. At year 9/third grade, Black children are more likely to be stereotyped as 'troublemakers' (e.g., as a result of having been suspended), potentially leading to bias in teacher reports (Kinsler 2011; Okonofua and Eberhardt 2015). However, teachers are more closely involved in disciplinary decisions and therefore are important reporters. Including both teacher and parent reports separately therefore provides a fuller picture of child behaviors.

Equally important, the use of these holistic and detailed measures of child behavior based on multiple reporters enable us to capture the range of both subtle and readily apparent forms of misbehavior that in reality shape teacher decisions about whether to refer/suspend a student for any single disciplinary infraction (Vavrus and Cole 2002). Studies based on single infraction classifications from discipline records introduce issues of non-random selection into referral and may produce biased estimates of the role of behavior differences in accounting for suspension because, by definition, behavior is only reported for students who are referred/suspended.¹² By contrast, the externalizing problems scales used here are not conditional on suspension. Instead, they consist of summed indices containing age-appropriate items from four sub-scales: *social problems* (year 5: 8 parent-reported items, 6 kindergarten teacher items; year 9: 11 parent-reported, 11 teacher-reported), *attention problems* (year 5: 5 parent-reported, 9 teacher-reported; year 9: 10 parent-reported, 10 teacher-reported), *aggression* (year 5: 14 parent-reported, 19 teacher-reported; year 9: 18 parent-reported, 18 teacher-reported), and *rule-breaking* (year 5: 9 parent-reported, 0 teacher-reported; year 9: 17 parent-reported, 12 teacher-reported) (Achenbach 1991).¹³ Responses to each year 5 item range from 0 "not true", 1 "somewhat/sometimes true", to 2 "very often or often true." Responses to each year 9 item range from 0 "never," 1 "sometimes," 2 "often," to 3 "very often." Items are reverse-coded as necessary so higher scores

indicate greater problems. There are 36 items in the year 5 parent-reported scale ($\alpha=0.80$), 34 items in the year 5 teacher-reported scale ($\alpha=0.92$), 54 items in the year 9 parent-reported scale ($\alpha=0.93$), and 51 items in the year 9 teacher-reported scale ($\alpha=0.94$).

Controls: Analyses also adjust for a number of child and family factors that prior research has found to be correlated with child race as well as school suspension/expulsion.

Parents' Socio-Economic Resources. We use mother-reported household income-to-poverty ratio at year 9, mother's education at child's birth (dummy for "some college/college degree or higher" relative to "high school or less"), and mother's age at the child's birth (to account for differences in social context of childbearing and in genetic factors that may influence early development). *Father Absence.* An indicator variable is coded 1 to capture all family types involving at least one episode of biological father absence across any of the five survey waves. Note that stable cohabiting families with both biological parents living in the household at each survey wave are coded as father present. Father absent families include stable, single mother families and families where the mother re-partnered or re-married at least once (i.e., stepparent families). Note that due to small sample sizes, father-headed and foster parent households are excluded. The reference (coded 0) encompasses stable two-(biological) parent family, based on mother's report of living with the child's biological father in each of the five waves (birth through year 9). *Paternal Incarceration:* Prior research shows that parental incarceration is associated with suspension as well as externalizing behavior problems, especially among boys, and that Black boys are more likely than white boys to be exposed to this form of family structure instability (Wildeman 2010). Moreover, paternal incarceration may also serve as a mechanism of social exclusion that exerts intergenerational effects on children's suspension in the next generation. As such, we include a dummy variable equal to 1 if the child's biological

father has ever been in jail or prison at any wave up to year 9. *Other Child Characteristics.*

Lastly, we control for the child's Peabody Picture Vocabulary Test (PPVT) score at year 5 (to account for differences in academic skills at school entry) and the child's age in months in year 9 (to account for differences in age at school entry and in grade retention over the first few years of schooling, as well as differences in the child's age at interview; children were interviewed up to six months before or after their birthday).

Analytic Approach

We employ a two-stage Oaxaca-Blinder decomposition analysis to identify the contributions to the racial gap in suspension that are associated with differences in the racial and socioeconomic composition of children's schools (hypothesis 1), racial differences in children's behaviors (hypothesis 2), and differences in suspension between Black and White boys with the same behaviors who attend schools with the same racial and socioeconomic compositions (hypothesis 3) (Jann 2008).¹⁴ The decomposition models a counterfactual scenario that displays how large the racial gap in suspension would be for hypothetical Black and White boys if they were to have the same levels/exposures to the factors in our model, but different coefficients/slopes, versus if they were to have the same coefficients/slopes, but different levels/exposures observed in our sample (i.e., "differential treatment") (Jann 2008). For each factor, this decomposition parses the racial gap in suspension into two components: (1) the portion of the gap associated with the race difference in mean levels of a given factor and (2) the portion of the gap associated with differences in responses to Black versus White boys when both have the same mean levels of a given factor (Jann 2008). This counterfactual analysis allows us to estimate how large the racial gap in

suspension would be if Black and White boys had the same mean levels of exposure to each factor, but the association for each factor varied by race, versus if Black and White boys had the same coefficients for each factor but different levels of exposure, as shown in Equation 1:

$$Suspension_W - Suspension_B = \underbrace{(x'_W - x'_B)\beta_B}_{\text{Exposure/levels}} + \underbrace{\bar{x}'_B(\beta_W - \beta_B)}_{\text{Vulnerability/ "effects"}} \quad (1)$$

where $(x'_W - x'_B)\beta_B$ is the contribution of race differences in levels of exposure to the observed predictors, and $\bar{x}'_B(\beta_W - \beta_B)$ is the contribution of race differences to their slopes or coefficients/‘effects’.¹⁵ In the main text we present results using means and coefficients for Black boys as the reference (in the Appendix we discuss results using White boys’ means and coefficients as the reference).

We estimate two main decomposition models. In the first decomposition, we focus on children’s behavior at the time they enter school (year 5). Year 5 behaviors capture perceptions of boys’ behavior concurrent with or preceding any primary school suspension.¹⁶ Next, to account for the possibility that Black students’ behavior problems may disproportionately worsen between school entry and later elementary school, we estimate a second decomposition model that uses reports of behaviors at year 5 and year 9.¹⁷ Because behaviors at year 9 may be endogenous to suspension and because teacher reports of behavior may reflect labeling bias, this second model provides an upper bound estimate of the extent to which race differences in behavior account for the Black/White gap in suspension and a lower bound estimate of the role of differential punishment for the same behavior. Additionally, at both years 5 and 9, we use teacher as well as parent reports in order to capture a more holistic portrait of boys’ behavior, since the use of teacher reports alone may overlook racial bias in how teachers interpret and

assign meaning to the behaviors of minority students (Ferguson 2001; Gilliam et al. 2016; Vavrus and Cole 2002).

Importantly, the Oaxaca-Blinder decomposition allows us to *simultaneously* estimate within a single equation the contributions to the racial gap in suspension that which would arise from racial differences in *levels of exposure to* family socioeconomic background, family structure instability, and child age in months and cognitive skills, as opposed to that which would arise from the *differential associations between* each of these factors and school suspension (including differential treatment for the same behaviors). The decomposition approach has two primary advantages given our research questions. First, because the two-stage Oaxaca-Blinder decomposition interacts each observed predictor with child race (Black=1), it minimizes the risk of upwardly biased estimates of ‘differential treatment for the same behaviors’ that could arise if not all predictors were interacted with ‘Black’ (i.e., if other predictors were not also allowed to vary by race). Second, because the decomposition simultaneously estimates contributions of differences in levels and coefficients of each predictor within a single equation, it avoids making assumptions about the causal ordering of mediators that is a common issue with traditional mediation analyses (for discussion of this issue, see Legewie and DiPrete (2014), Morgan (2005), and Owens (2016)). Finally, we discuss results using a non-linear expansion to the Oaxaca-Blinder method in the Appendix; substantive results remain unchanged.

RESULTS

Descriptive Statistics

Table 1 reports means and standard deviations (or proportions) for each of the variables in our analyses. As shown in the table, Black boys are over three and a half times more likely

than White boys to report ever having been suspended in- or out-of-school or expelled by the year nine interview (37% of Black boys as compared to 10% of White boys). Our measure encompasses ever being suspended (in or out of school) or expelled, *between kindergarten and third grade*. For context, we compare these estimates to those from the Office of Civil Rights Data Collection (CRDC) of the U.S. Department of Education, which reports incidence rates across K-12 students for the 2011-12 academic year alone. Despite differences in measures, racial gaps are of similar magnitude: Black boys in K-12 from the CRDC are 2.6 times more likely than White boys in K-12 to be suspended in or out of school in the 2011-12 academic year alone (35.4% compared to 13.5%) (Office for Civil Rights 2014).

[TABLE 1 ABOUT HERE]

Descriptive differences in school composition and behaviors are also in line with our first and second hypotheses. The average Black boy in our sample attends a school in which 75% of students are Black or Latino/Hispanic and 73% receive free-or-reduced-price lunch. The numbers for White boys are 28% and 41%, respectively. Differences in school composition are further magnified when we look at the percent of students attending schools that are both minority and poor: 54% of Black boys compared to only 8% of White boys.

In line with hypothesis 2, Black boys score higher than White boys on behavior problems at year 5, based on both parent and teacher ratings. The racial gap in teacher reports of boys' behaviors grows from roughly 0.15 SD to 0.52 SD between years 5 and 9. Strikingly, the Black/White gap in behaviors at year 5 is of roughly equal magnitude using both teacher and parent reports (0.15 SD), whereas by year 9 the teacher reported gap is much larger than the parent reported gap (0.52 SD). This finding may be due to a disproportionate worsening of Black

boys' behaviors (perhaps in response to a suspension) or it may be due to an increase in teacher bias in assigning meaning to the behaviors of Black boys.

Finally, we observe statistically significant racial differences in the *levels* of virtually all of the control variables, including parental socioeconomic status, family instability/composition, paternal incarceration, and demographic controls. Exceptions are child's age and enrollment in schools serving a majority of poor.

Differential Associations between Behaviors and Suspension by Race

In our third hypothesis we posit that Black boys are more likely to be suspended than White boys with the same levels of behavior problems, *ceteris paribus*. Columns 4-5 of **Table 2** present coefficients for all variables in the analyses stratified by race. For teacher reports, each unit increase in year 5 teacher-rated behavior problems is associated with a 0.8 percent increase in suspension among White boys and a 1.0 percent increase in suspension among Black boys. For parent reports, the magnitude of the association is much weaker for White boys: each unit increase in behavior problems is associated with only a 0.2 percent increase in suspension. However, for Black boys, each unit increase in parent-rated behaviors is associated with a 0.9 percentage increase in suspension. Results from an interaction model shown in Appendix Table A.3 indicate that this race difference in the association between each unit increase in parent-rated behavior problems and suspension is statistically significant, inclusive of all other main effects and interactions with race shown in Table 2.

[TABLE 2 ABOUT HERE]

Tests of Hypotheses

How much of the racial gap in suspension/expulsion is driven by racial differences in school composition (hypothesis 1), student behaviors (hypothesis 2), and differential responses to student with the same behaviors (hypothesis 3)? **Table 2** and **Figure 1** display results from our decomposition analysis. Black-White differences in school composition and boys' behaviors are shown in the first 5 rows of Column 3. Differences in slopes/'effects' associated with school composition and student behaviors are reported in Columns 4 and 5. Column 6 displays the proportion of the total racial gap in suspension that can be attributed to *differences in levels of exposure*, and Column 7 displays the proportion that can be attributed to *differences in the slopes/'effects'* (i.e., differences in treatment by race for the same exposures). Together, the proportions in columns 6 and 7 sum to 0.27, or a 27 percentage-point gap in suspension. Positive values refer to factors that widen the racial gap in suspension, while negative values refer to factors that narrow the gap. Importantly, the estimates of contributions are net of differences in the means and slopes/effects of each of the other variables in the model.

[FIGURE 1 ABOUT HERE]

Hypothesis #1. *Race differences in school composition (i.e., percent minority and percent poor) account for a large share of the racial gap in suspension/expulsion by year 9.*

According to our estimates, racial differences in the composition of the schools that Black and White boys attend account for 3.0 percentage-points (or 13.1%) of the 27 percentage-point gap in suspension/expulsion. Throughout our reporting of results, gap widening and gap narrowing contributions are displayed in percentage-point units in column 6 of Table 2. For school composition, the numbers are $(0.035-0.001+0.001)/0.268=0.131\%$. Importantly, the sorting of Black boys into schools that serve students from *both* low-income and minority

backgrounds accounts for this entire 13% contribution. The first three bars of Figure 1 illustrate these findings. To summarize, our results not support our hypothesis that school sorting accounts for a large portion of the racial gap in suspension. (In Appendix Table A.4 we show that schools serving poor and minority students also have high rates of punitive discipline.)

Hypothesis #2. *Race differences in children’s problem behaviors account for more of the racial gap in suspension/expulsion at year 9 than what is suggested by studies that use school records data, but less of the gap than what is suggested by studies that measure behavior after the child has been exposed to school discipline.*

According to our estimates, differences in teachers’ reports of boys’ behavior at year 5 can account for 1.6 percentage points (6.0%) of the racial gap in school suspension ($0.016/0.268=0.060$), while differences in parents’ reports can account for an additional 0.9 percentage-points (or 3.3%) ($0.009/0.268=0.033$). Together, parent and teacher reports of boys’ behavior at school entry account for 2.5 percentage points (9.3%) of the racial gap in suspension, *ceteris paribus*. The second panel of Figure 1 illustrates these findings for parent and teacher reports. To summarize, these results are in line with hypothesis 2, which suggests that race differences in behaviors at school entry account for a small but non-zero component of the racial gap in suspension.

Hypothesis #3. *Race differences in the disciplinary treatment of Black boys with the same behavior at school entry, the same family socioeconomic resources and the same school contexts explain a large share of the racial gap in suspension/expulsion at year 9.*

Results for our third hypothesis, which focuses on differences in slopes/effects rather than differences in levels, are presented in column 7 of Table 2. According to our estimates, 10.4 percentage points (38.8%) of the racial gap in suspension can be attributed to differential

treatment of Black and White boys who enter school with the same behaviors and who have similar means on all the other variables in the model $[(0.024+0.080)/0.268=0.388]$.¹⁸ In particular, 2.4 percentage-points (8.9%) of the 27 percentage-point gap is due to differential treatment for the same teacher reports $(.024/0.268=0.089)$, whereas 8 percentage-points (30.0%) is associated with differential treatment for the same parent reported behaviors, *ceteris paribus* $(0.080/0.268=0.300)$. We expand on this finding in the Discussion. Altogether, race differences in between-school sorting (hypothesis 1), in children's behaviors at year 5 (hypothesis 2), and in the treatment of boys who have the same behaviors and attend similar schools (hypothesis 3) account for nearly two thirds of the 27 percentage-point racial gap in suspension at year 9.¹⁹

Robustness Checks

We view the results reported in Table 2 as lower-bound estimates of the role of differences in behaviors in accounting for the racial gap in suspension insofar as our model does not allow boys' behaviors to change over time. To the extent that Black boys' behaviors worsens more than White boys', our use of the year 5 measure would lead to an underestimate of the role of race differences in behavior, and an overestimate the role of differential treatment for the same behaviors at school entry.

As shown in **Figure 2**, Black and White boys' behaviors worsened between years 5 and 9, with the largest increase appearing in teacher ratings of Black boys' behaviors. One explanation for this trend might be that Black boys are more likely than White boys to experience family structure instability and/or neighborhood violence, all of which are likely to increase behavior problems. Alternatively, the disproportionate worsening of Black boys' behaviors may be due to differences in exposure to negative school environments. If Black boys are more likely than White boys to be suspended, and if suspension leads to an increase in

behavior problems, we would expect the racial gap in behaviors to increase over time (Okonofua and Eberhardt 2015). Finally, between school entry and third grade, Black students may be more likely to garner negative reputations in the eyes of teachers, which could account for their more negative behavior ratings and more suspension as they progress through elementary school (Ferguson 2001).

[FIGURE 2 ABOUT HERE]

To take account of changes in boys' behaviors after they enter school, we re-estimated our decomposition model and included parent and teacher reports of boys' behavior at the time of the year 9 survey. **Table 3** and **Figure 3** display results from this second analysis.

[TABLE 3 ABOUT HERE]

[FIGURE 3 ABOUT HERE]

The second panel of Table 3 and Figure 3 displays results from the new model. For ease of comparison, the circles overlaid on Figure 3 indicate estimates of contributions using year 5 behaviors only. Our estimates of the role of between-school sorting (hypothesis 1) decrease only very slightly when we add parent and teacher reports of year 9 behaviors to the model. Assuming similar behavior and similar responses to behavior, race differences in the racial and income composition of the schools Black and White boys attend account for 3.1 percentage-points (11.9%) of the 27 percentage-point race gap in suspension/expulsion $[(0.030+0.000+0.001)/0.268=0.119]$, as opposed to 3.5 percentage-points (13.1%), using year 5 behaviors alone. As before, the sorting of Black boys into schools that serve students who are from both low-income and minority backgrounds accounts for almost all of the 3.1 percentage-point (11.9%) contribution.

As expected, the estimated contribution of differences in behavior (hypothesis 2) increases substantially when year 9 behaviors are included in the model. According to column 6, race differences in behaviors account for 6.3 percentage-points (23.5%) of the 27 percentage-point racial gap in behavior ($[0.010+0.006+0.048-0.001]/0.268=0.235$), as compared with the 2.5 percentage-points (9.3%) reported Table 2 and Figure 1 based on year 5 reports alone. Importantly, 5.8 percentage-points (21.6%) of the 27 percentage-point gap is due to *teachers'* higher ratings of Black compared to White boys ($[0.010+0.048]/0.268=0.216$), as compared to 1.6 percentage-points (6.0%), using year 5 behaviors alone. By contrast, only 0.5 percentage-points (1.9%) of the gap is due to parents' ratings as opposed to 0.9 percentage-points (3.0%) in the Table 2. In sum, when year 9 behaviors are included in the model, the difference in behaviors can account for a much larger share of the racial gap in suspension than when we rely on year 5 behaviors alone. These findings suggest that the *time* at which child behaviors are measured is paramount. We also elaborate on this point in the discussion.

Surprisingly, we continue to find strong support for hypothesis 3. Results in column 7 of the second panel in Table 3 (and panel 3 of Figure 3) indicate that 18.2 percentage-points (67.9%) of the race gap in suspension can be attributed to the differential treatment of Black relative to White boys ($[0.000+0.073+0.096+0.013]/0.268=0.679$) as compared to 10.4 percentage-points (38.8%) in the model using year 5 behaviors alone. Specifically, 9.6 percentage-points (35.8%) of the 27 percentage-point gap is due to differential treatment for the same teacher reports ($[0.00+0.096]/0.268$), whereas another 9.0 percentage-points (32.1%) is associated with differential treatment for the same parent reported behaviors, *ceteris paribus*. This is in contrast to 2.4 percentage-points (8.9%) associated with teachers' year 5 behavior reports and 8.0 percentage-points (30%) associated with parents' year 5 behavior reports.

While we expected that the inclusion of the year 9 behaviors would bring behavior and suspension more closely in line with one another and increase the relative importance of behavior differences in accounting for the racial gap in suspension, we did not expect the estimates for differential treatment to persist and to even increase. The latter finding tells us that even after including the more liberal measures of behaviors, which are likely to be endogenous to suspension, we continue to find strong evidence that Black boys are treated differently than White boys for the same holistic behaviors. We discuss the implications of this point in the Discussion.

The estimates reported in Table 3 treat Black boys as the reference category. If White boys were treated as the reference category, differences in school sorting would account for 33.6% of the gap (as compared to 11.9% in the previous analysis), differences in behavior would account for 8.6% of the gap (as compared to 23.5% in the previous analysis), and differential treatment would account for 82.5% of the gap (as compared to 67.9% in the previous analysis), as shown in Appendix Table A.5.²⁰ If 80% of Black were, like White boys, to attend the non-poor White schools with low levels of use of punitive discipline, Black boys' suspension/expulsion rates would decrease dramatically, and thus between-school sorting would increase substantially in its contribution to accounting for the racial gap in suspension/expulsion. Similarly, the large difference in coefficients would mean that, even if Black and White boys were to have White boys' lower mean levels of behavior problems, Black boys would be suspended/expelled at much higher rates, leading to an increase in the contribution of differential treatment to the race gap. By contrast, if the relationship between behaviors and suspension/expulsion was as small for Black boys as for White boys, the mean race differences in behaviors would decrease in their overall contribution to the race gap. Viewing these two sets

of estimates as providing a range for the contribution of the three drivers of the racial gap, results are consistent about the important role of differential treatment as a primary driver of the racial gap in elementary school suspension/expulsion. Finally, in Appendix Table A.6 we show that results are also consistent across the Oaxaca-Blinder linear decomposition and the Fairlie non-linear decomposition expansion.

DISCUSSION

In this study, we test three hypotheses that may account for the Black-White gap in school suspension and expulsion: differences in the racial and socioeconomic composition of the schools children attend (i.e., between-school sorting), differences in children's behavior, and differential treatment of children with the same overall behaviors. Our study makes four important contributions to the literatures on education, stratification, delinquency and social control, and social psychology. First, whereas past research on delinquency and social control has focused primarily on racial disparities among adolescents and adults and on punishment within the criminal justice system, we document a large racial gap in punishment among children in elementary school.²¹ Using data on a cohort of children born in large US cities at the turn of the 21st century, we document a 27 percentage-point racial gap in suspension/expulsion – 37% of Black boys versus 10% of White boys – by the third grade.

Second, we help disentangle among the *mechanisms* that underlie the racial gap in school suspension/expulsion. Whereas prior research has focused primarily on racial differences in student behaviors and differences in the composition of the schools that Black and White children attend, we find that these factors account for only a modest share of the racial gap in suspension during elementary school. Instead, we find that the largest portion of the racial gap is

due to the differential treatment of Black and White boys who attend similar schools and who exhibit similar behaviors at the time they enter school (Ferguson 2001; Rocque 2010; Skiba and Williams 2014). Specifically, differential treatment accounts for 39% of the racial gap in suspension, while between-school sorting and behavior differences account for roughly 13% and 9% of the gap respectively.

The broad support we uncover for the differential treatment hypothesis is consistent with prior research on interpersonal racial discrimination (Pager and Quillian 2005; Pager and Shepherd 2008). In elementary school in particular, when misbehavior tends to be both relatively common and relatively minor, educators may exercise high levels of discretion in determining sanctions for inappropriate behavior (Dovidio and Gaertner 2004). Moreover, in cases where there is a lack of concrete information about the extent to which parents and others at home or in the community can help address a child's misbehavior, educators are more likely to rely on racial stereotypes to fill in missing information. Importantly, differential treatment may reflect "tough love," a distinct but related mechanism that is motivated not by implicit/explicit bias but rather by a desire to help prepare Black boys for the challenges they are likely to encounter in the wider society (Gilliam et al. 2016).

Third, our analyses help reconcile conflicting results reported in prior research by showing how variation in the nature of the sample and the time at which child behaviors are measured can dramatically change results. For example, using a sample of all children entering kindergarten, we find that behavior differences play a larger role than what is found in studies based on a sample of children referred for punishment. If Black boys are more likely than White boys to be referred for minor misbehaviors, the latter studies are likely to *underestimate* the role of race differences in behavior.

Similarly, measuring behaviors at the time children enter school, we find that differences in behaviors play a smaller role in accounting for the racial disparities in suspension than what is found that in studies that measuring behaviors later in a child's school career. This happens because Black boys are more likely than White boys to be suspended and because suspension is likely to negatively affect behaviors (Dance 2002; Rios 2011). Thus studies that measure behaviors when children are further along in school and after suspension has occurred are likely to overstate the role of behavior differences in accounting for the racial gap in suspension. This insight is consistent with a rich ethnographic literature, which documents how worsening behavior often follows from suspension as youth act out to gain dignity and seize their own agency in response to harsh treatment in schools, including both stereotyping and harassment or discrimination by school officials (Dance 2002; Ferguson 2001; Rios 2011).

Our data allow us to assess children's behaviors at two points – when children enter kindergarten and in third grade. As we expected, the behavior differences explain more of the racial gap in suspension when behaviors are measured in third grade. The difference between the two time points could be due to non-school factors that lead to a greater increase in the behavior problems of Black boys, or it could be due to within school factors, such as harsh punishment and suspension. Alternatively, the disproportionate suspension of Black boys may lead teachers to rate their behaviors more negatively. Although we cannot differentiate between these sources of differential change in behavior, it is notable that we continue to find evidence of differential treatment, even when we use our more liberal measure of boys' behaviors.

Fourth, and related, we show that parents and teachers provide relatively independent insights into child behaviors. The two measures have low correlations, and teacher and parent reports carry different implications for the contributions of race differences in behaviors.

Moreover, we find that differential treatment operates differently for teacher relative to parent reports. Therefore, although both parent and teacher reports of behavior are reliable predictors of a range of later child outcomes (Duncan et al. 2007; Owens 2016), our findings suggest that future research should analyze teacher and parent reports of behavior separately. Separate reports provide a more accurate picture of the role of child behaviors in stratification processes.

Our study also has limitations. First, the decomposition approach and the use of observational data preclude causal claims about the role of the different mechanisms in accounting for the racial gap in suspension. By design, our decomposition approach leverages the selection processes that produce racial differences in between-school sorting and behavior to create counterfactual scenarios that model the gaps we would expect to see if the selection processes guiding one group (e.g., Blacks) were applied to the other group. Furthermore, our data do not allow us to use school fixed effects because our data are based on a birth cohort rather than a school sample and do not contain sufficient clustering within schools. By contrast, school-based samples like the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K:98) that would facilitate a school fixed effects approach do not contain measures of suspension until older ages.

Second, our holistic reports of children's behaviors do not allow us to link suspension to a particular infraction and thus to determine whether Black and White boys are differentially suspended for the same infraction. Our year 9 measures of behavior, however, help compensate for this lack of information by capturing changes in teacher and parent impressions of a child's behavior following the specific infraction that led a child to be suspended. Thus we are able to help guard against the possibility that our findings about the role of differential treatment are driven by a lack of data on specific infractions.

A third limitation of our study is that we are unable to disentangle the extent to which differences between parent and teacher reports of behavior reflect variation in behaviors across contexts (school vs. home) from differences in the ways identical behavior is perceived in the eyes of different observers (teachers vs. parents). Future research would benefit from experimental techniques to investigating this question. Finally, our analysis is based on a sample of children born in large cities and therefore our results should not be interpreted as representing the dynamics that occur in suburban and rural settings.

CONCLUSION: IMPLICATIONS FOR POLICY

At the broadest level, our findings suggest that the processes leading to racial gaps in suspension and expulsion, and the role of differential treatment in particular, begin much earlier in the life course than previously documented in a population-based sample and are firmly entrenched in the education system, the nation's main institution for promoting upward mobility. Specifically, our findings suggest that differential treatment plays an important role in accounting for differences in school suspension between Black and White boys. While our analyses are descriptive rather than causal, the findings are consistent with laboratory experiments conducted by Okonofua and Eberhardt (2015). Together, these findings merit consideration in re-thinking the processes governing the assignment of disciplinary sanctions.

Given the complex constellation of factors that contribute to the racial gap in suspension, it would be overly simplistic to say that policy efforts should focus on a single mechanism. Nor do our results say how policies or practices should be structured at the individual school or district levels. Therefore, in certain contexts, facilitating school involvement from minority parents may be the most efficacious way to reduce racial disparities in suspension. In other

contexts, increasing empathy among teachers especially when relating to their students of color, may be more effective. In the latter case, recent work by Okonofua, Paunesku and Walton (2016) shows promising effects of an empathetic mindset intervention with teachers for reducing suspension rates. This work suggests that creative efforts to increase teacher empathy may be another effective strategy.

Finally, expanding on Okonofua and Eberhardt's (2015) experiment with mostly White teachers, future work should use an experimental design to identify whether differential treatment is indicative of racial discrimination, which implies the presence of racial prejudice, or whether it is driven by “tough love” or other forms of differential treatment that may not stem from prejudicial beliefs about African Americans. By further disentangling the mechanisms of bias and discrimination, future work will be able to weigh in more precisely on policy approaches that might help reduce it.

REFERENCES

- Achenbach, Thomas M. 1991. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*: Department of Psychiatry, University of Vermont Burlington, VT.
- Allport, Gordon W. 1979. *The Nature of Prejudice*. New York: Basic Books.
- Andrew, Megan, and Mary Kate Blake. 2017. "The Long-Term Scars of Exclusionary School Discipline and Their Non-Cognitive, Cognitive, and Academic Mechanisms." in *Working Paper Series*, edited by Center for Research on Educational Opportunity: University of Notre Dame.
- Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car." *The American Economic Review* 85(3):304-21.
- Bates, John E., Kathryn Bayles, David S. Bennett, Beth Ridge, and Melissa M. Brown. 1991. "Origins of Externalizing Behavior Problems at Eight Years of Age." Pp. 93-120 in *The Development and Treatment of Childhood Aggression*, edited by Debra J. Pepler and Kenneth H. Rubin. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Bertrand, Marianne, and Jessica Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5(1):32-64.
- Bobo, Lawrence, and James Kluegel. 1997. "Status, Ideology, and Dimensions of Whites' Racial Beliefs and Attitudes: Progress and Stagnation." Pp. 93-120 in *Racial Attitudes in the 1990s: Continuity and Change*, edited by Steven A. Tuch and Jack Martin. Westport, CT: Praeger.
- Bradshaw, Catherine P., Mary M. Mitchell, Lindsey M. O'Brennan, and Philip J. Leaf. 2010. "Multilevel Exploration of Factors Contributing to the Overrepresentation of Black Students in Office Disciplinary Referrals." *Journal of Educational Psychology* 102(2):508-20.
- Brooks-Gunn, Jeanne, and Greg J. Duncan. 1997. "The Effects of Poverty on Children." *The Future of Children* 7(2):55-71.
- Calarco, Jessica McCrory. 2011. "'I Need Help!'" Social Class and Children's Help-Seeking in Elementary School." *American Sociological Review* 76(6):862-82.
- Collins, Patricia Hill. 2015. "Intersectionality's Definitional Dilemmas." *Annual Review of Sociology* 41:1-20.
- Dallaire, Danielle H., Anne Ciccone, and Laura C. Wilson. 2010. "Teachers' Experiences with and Expectations of Children with Incarcerated Parents." *Journal of Applied Developmental Psychology* 31(4):281-90.
- Dance, Lory Janelle. 2002. *Tough Fronts: The Impact of Street Culture on Schooling*. New York: RoutledgeFarmer.
- Deater-Deckard, Kirby, and Kenneth A. Dodge. 1997. "Externalizing Behavior Problems and Discipline Revisited: Nonlinear Effects and Variation by Culture, Context, and Gender." *Psychological Inquiry* 8(3):161-75.
- Dill, Bonnie Thornton, and Ruth Enid Zambrana (Eds.). 2009. *Emerging Intersections: Race, Class, and Gender in Theory, Policy, and Practice*. New Brunswick, NJ: Rutgers University Press.
- DiPrete, Thomas A., and Gregory M. Eirich. 2006. "Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments." *Annual Review of Sociology* 32:271-97.

- Dovidio, John F., and Samuel L. Gaertner. 2004. "Aversive Racism." *Advances in Experimental Social Psychology* 36:4-56.
- Duncan, Greg J., Chantelle J. Dowsett, Amy Claessens, Katherine Magnuson, Aletha C. Huston, Pamela Klebanov, Linda S. Pagani, Leon Feinstein, Mimi Engel, and Jeanne Brooks-Gunn. 2007. "School Readiness and Later Achievement." *Developmental Psychology* 43(6):1428.
- Egalite, Anna J., and Brian Kisida. 2017. "The Effects of Teacher Match on Students' Academic Perceptions and Attitudes." in *Department of Educational Leadership Working Paper Series*: North Carolina State University.
- Entwisle, Doris R., and Karl L. Alexander. 1993. "Entry into School: The Beginning School Transition and Educational Stratification in the United States." *Annual Review of Sociology* 19(1):401-23.
- Entwisle, Doris R., Karl L. Alexander, and Linda Steffel Olson. 2005. "First Grade and Educational Attainment by Age 22: A New Story." *American Journal of Sociology* 110(5):1458-502.
- Fabelo, Tony, Michael D. Thompson, Martha Plotkin, Dottie Carmichael, Miner P. Marchbanks III, and Eric A. Booth. 2011. "Breaking Schools' Rules: A Statewide Study of How School Discipline Relates to Students' Success and Juvenile Justice Involvement." Pp. 54-55 in *Council of State Governments Justice Center*. New York: Council of State Governments Justice Center.
- Ferguson, Ann A. 2001. *Bad Boys: Public Schools in the Making of Black Masculinity*. Ann Arbor: University of Michigan Press.
- Forrest, Christopher B., Katherine B. Bevans, Anne W. Riley, Richard Crespo, and Thomas A. Louis. 2013. "Health and School Outcomes During Children's Transition Into Adolescence." *Journal of Adolescent Health* 52(2):186-94.
- Gilliam, Walter S., Angela N. Maupin, Chin R. Reyes, Maria Accavitti, and Frederick Shic. 2016. "Do Early Educators' Implicit Biases Regarding Sex and Race Relate to Behavior Expectations and Recommendations of Preschool Expulsions and Suspensions?" Pp. 991-1013 in *Yale Child Study Center*. New Haven, CT: Yale University.
- Ginsburg, Alan, Phyllis Jordan, and Hedy Chang. 2014. "Absences Add Up: How School Attendance Influences Student Success." edited by Attendance Works. Portland, OR: Attendance Works.
- Goff, Phillip A., Matthew C. Jackson, Di Leone, Brooke A. Lewis, Carmen M. Culotta, and Natalie A. DiTomasso. 2014. "The Essence of Innocence: Consequences of Dehumanizing Black Children." *Journal of Personality and Social Psychology* 106(4):526.
- Gregory, Anne, Dewey Cornell, and Xitao Fan. 2011. "The Relationship of School Structure and Support to Suspension Rates for Black and White High School Students." *American Educational Research Journal* 48(4):904-34.
- Gregory, Anne, Russell J. Skiba, and Pedro A. Noguera. 2010. "The Achievement Gap and the Discipline Gap Two Sides of the Same Coin?" *Educational Researcher* 39(1):59-68.
- Hinshaw, Stephen P., and Richard M. Scheffler. 2014. *The ADHD Explosion: Myths, Medication, Money, and Today's Push for Performance*: Oxford University Press.
- Hirschfield, Paul J. 2008. "Preparing for Prison? The Criminalization of School Discipline in the USA." *Theoretical Criminology* 12(1):79-101.

- Jacob, Brian A. 2005. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89(5):761-96.
- Jann, Ben. 2008. "The Blinder-Oaxaca Decomposition for Linear Regression Models." *The Stata Journal* 8(4):453-79.
- Kinsler, Josh. 2011. "Understanding the Black–White School Discipline Gap." *Economics of Education Review* 30(6):1370-83.
- Kupchik, Aaron, Nicole Bracy, Michael Apple, Paul Hirschfield, Ronnie Casella, John Gilliom, Andrew Hope, Tyson Lewis, Pauline Lipman, and Richard Matthew. 2009. *Schools Under Surveillance: Cultures of Control in Public Education*. New Brunswick, NJ: Rutgers University Press.
- Legewie, Joscha, and Thomas A. DiPrete. 2014. "Pathways to Science and Engineering Bachelor's Degrees for Men and Women." *Sociological Science* 1:41-48.
- Lewis, Amanda E., and John B. Diamond. 2015. *Despite the Best Intentions: How Racial Inequality Thrives in Good Schools*. New York, NY: Oxford University Press.
- Lindsay, Constance A., and Cassandra M. D. Hart. 2017. "Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina." *Educational Evaluation and Policy Analysis* Forthcoming.
- Losen, Daniel J., Cheri L. Hodson, Michael A. Keith II, Katrina Morrison, and Shakti Belway. 2015. "Are We Closing the School Discipline Gap?", edited by The Civil Rights Project. Los Angeles: University of California, Los Angeles.
- Magnuson, Katherine A., and Jane Waldfogel. 2005. "Early Childhood Care and Education: Effects on Ethnic and Racial Gaps in School Readiness." *The Future of Children* 15(1):169-96.
- McLeod, Jane D., and James M. Nonnemaker. 2000. "Poverty and Child Emotional and Behavioral Problems: Racial/Ethnic Differences in Processes and Effects." *Journal of Health and Social Behavior* 41(2):137-61.
- Mittleman, Joel. 2017. "A School-to-Prison Pipeline?: Exclusionary School Discipline and Juvenile Justice Contact." Population Association of America Annual Meeting Archives: Population Association of America.
- Morgan, Stephen L. 2005. *On the Edge of Commitment: Educational Attainment and Race in the United States*. Stanford, CA: Stanford University Press.
- Morris, Edward W., and Brea L. Perry. 2016. "The Punishment Gap: School Suspension and Racial Disparities in Achievement." *Social Problems* 63(1):68-86.
- . 2017. "Girls Behaving Badly? Race, Gender, and Subjective Evaluation in the Discipline of African American Girls." *Sociology of Education* 90(2):127-48.
- Office for Civil Rights, U. S. Department of Education. 2014. "Civil Rights Data Collection: Data Snapshot (School Discipline)." Washington, D.C.: U.S. Department of Education.
- Okonofua, Jason A, Gregory M Walton, and Jennifer L Eberhardt. 2016. "A Vicious Cycle A Social–Psychological Account of Extreme Racial Disparities in School Discipline." *Perspectives on Psychological Science* 11(3):381-98.
- Okonofua, Jason A., and Jennifer L. Eberhardt. 2015. "Two Strikes: Race and the Disciplining of Young Students." *Psychological Science* 26(5):617-24.
- Okonofua, Jason A., David Paunesku, and Gregory M. Walton. 2016. "Brief Intervention to Encourage Empathic Discipline Cuts Suspension Rates in Half Among Adolescents." *Proceedings of the National Academy of Sciences* 113(19):5221-26.

- Owens, Jayanti. 2016. "Early Childhood Behavior Problems and the Gender Gap in Educational Attainment in the United States." *Sociology of Education* 89(3):236-58.
- Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108(5):937-75.
- Pager, Devah, and Lincoln Quillian. 2005. "Walking the Talk? What Employers Say Versus What They Do." *American Sociological Review* 70(3):355-80.
- Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets." *Annual Review of Sociology* 34:181-209.
- Perry, Brea L., and Edward W. Morris. 2014. "Suspending Progress Collateral Consequences of Exclusionary Punishment in Public Schools." *American Sociological Review* 79(6):1067-87.
- Raffaele-Mendez, Linda M. 2003. "Predictors of Suspension and Negative School Outcomes: A Longitudinal Investigation." *New Directions for Youth Development* 99:17-33.
- Ramey, David M. 2015. "The Social Structure of Criminalized and Medicalized School Discipline." *Sociology of Education* 88(3):181-201.
- Reichman, Nancy E., Julien O. Teitler, Irwin Garfinkel, and Sara S. McLanahan. 2001. "Fragile Families: Sample and Design." *Children and Youth Services Review* 23(4):303-26.
- Rios, Victor M. 2011. *Punished: Policing the Lives of Black and Latino Boys*. New York: NYU Press.
- Robinson, Keith and Angel L. Harris. 2014. *The Broken Compass*. Cambridge, MA: Harvard University Press.
- Rocque, Michael. 2010. "Office Discipline and Student Behavior: Does Race Matter?" *American Journal of Education* 116(4):557-81.
- Santrock, John W., and Russel L. Tracy. 1978. "Effects of Children's Family Structure Status on the Development of Stereotypes by Teachers." *Journal of Educational Psychology* 70(5):754-57.
- Skiba, Russell J., Choong-Geun Chung, Megan Trachok, Timberly L. Baker, Adam Sheya, and Robin L. Hughes. 2014. "Parsing Disciplinary Disproportionality: Contributions of Infraction, Student, and School Characteristics to Out-of-School Suspension and Expulsion." *American Educational Research Journal* 51(4):640-70.
- Skiba, Russell J., Heather Edl, and M. Karega Rausch. 2007. "The Disciplinary Practices Survey: Principal Attitudes Towards Suspension and Expulsion." in *Annual Meeting of the American Educational Research Association*. Chicago, IL.
- Skiba, Russell J., Robert S. Michael, Abra Carroll Nardo, and Reece L. Peterson. 2002. "The Color of Discipline: Sources of Racial and Gender Disproportionality in School Punishment." *The Urban Review* 34(4):317-42.
- Skiba, Russell J., and Natasha T. Williams. 2014. "Are Black Kids Worse? Myths and Facts about Racial Differences in Behavior." in *The Equity Project at Indiana University*. Bloomington, IL: Indiana University.
- Vavrus, Frances, and KimMarie Cole. 2002. "'I Didn't Do Nothin'": The Discursive Construction of School Suspension." *The Urban Review* 34(2):87-111.
- Von Hippel, Paul T. 2007. "Regression with Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data." *Sociological Methodology* 37(1):83-117.
- Wald, Johanna, and Daniel J. Losen. 2003. "Defining and Redirecting a School-to-Prison Pipeline." *New Directions for Youth Development* 99:9-15.

- Wallace, John M., Sara Goodkind, Cynthia M. Wallace, and Jerald G. Bachman. 2008. "Racial, Ethnic, and Gender Differences in School Discipline among U.S. High School Students: 1991-2005." *The Negro Educational Review* 59(1-2):47-62.
- Welch, Kelly, and Allison A. Payne. 2010. "Racial Threat and Punitive School Discipline." *Social Problems* 57(1):25-48.
- Wildeman, Christopher. 2010. "Paternal Incarceration and Children's Physically Aggressive Behaviors: Evidence from the Fragile Families and Child Wellbeing Study." *Social Forces* 89(1):285-309.
- Wolf, Kerrin C., and Aaron Kupchik. 2017. "School Suspensions and Adverse Experiences in Adulthood." *Justice Quarterly* 34(3):407-30.
- Wright, John Paul, Mark Alden Morgan, Michelle A. Coyne, Kevin M. Beaver, and J.C. Barnes. 2014. "Prior Problem Behavior Accounts for the Racial Gap in School Suspensions." *Journal of Criminal Justice* 42(3):257-66.

Table 1. Means and Standard Deviations or Proportions of Variables Used in the Analyses, By Race

	Black (N=880)		White (N=368)		Min	Max	Sig. Diff.
	Mean	SD	Mean	SD			
<u>Outcome</u>							
Child Ever Suspended/Expelled Between Years 5 and 9	0.37		0.10		0	1	***
<u>Child Behavior Problems</u>							
Teacher-Rated Externalizing Problems Score, Year 5	12.78	9.95	11.17	9.91	0	55	**
Parent-Rated Externalizing Problems Score, Year 5	11.29	7.39	10.36	7.20	0	43	*
Teacher-Rated Externalizing Problems Score, Year 9	34.98	17.09	26.04	14.86	0	81	***
Parent-Rated Externalizing Problems Score, Year 9	13.74	11.57	13.89	11.27	0	93	
<u>School Racial and Socioeconomic Composition Factors</u>							
Percent of School Enrollment Black or Hispanic, Year 9	75.43	29.82	27.62	27.39	0	100	***
Percent of School Enrollment FRPL, Year 9	72.97	24.33	41.42	27.39	0	100	***
School is in Top Half of Minority and FRPL Enrollments, Year 9	0.54		0.08		0	1	***
School is in Top Half of Minority Enrollments in Sample, Year 9	0.10		0.01		0	1	***
School is in Top Half of FRPL Enrollments in Sample, Year 9	0.09		0.07		0	1	
School-Level Prevalence of Suspension and Expulsion, Year 9	0.08	0.07	0.03	0.04	0	0	***
<u>Family Socioeconomic and Stability Factors</u>							
Family Income-to-Poverty Ratio, Year 5	1.43	1.42	3.30	3.27	0	28	***
Mother Has Some College or College Degree, Year 1	0.30		0.56		0	1	***
Father Absent From Household at Any Wave, Year 5	0.85		0.49		0	1	***
Father has Ever Been in Jail or Prison, Year 9	0.61		0.38		0	1	***
Mother's Age, Year 1	24.07	5.51	27.02	6.50	15	43	***
<u>Other Child Factors</u>							
Child's PPVT Cognitive Score, Year 5	89.66	15.31	102.78	14.37	40	139	***
Child's Age, Year 5	112.31	4.50	111.74	3.87	104	130	*

Notes: Differences in means/proportions are based on two-tailed t-tests; *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

Source: Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 1,248 Black and White boys who remained in the study from birth (wave 1) through year 9 (wave 5). Multiple imputation of 20 datasets is used to handle item-missingness on all but the dependent variable (suspension/expulsion) and on measures of school composition. The outcome variable was included in the imputation equation but observations with imputed values were deleted before analysis.

Table 2. Contributions of Racial Differences in School Composition (H1) and Child Behaviors at Year 5 (H2) and in the "Effects" of the Same Behaviors in Similar Schools (H3) to the Black/White Gap in Probability of Suspension or Expulsion by Year 9, Based on a Two-Way Decomposition Model (Reference: Blacks)^a

Factors	Predictor	Means		Difference in Means	OLS Regression Coefficients				Contribution of Differences in Levels of Exposure	Contribution of Differences in 'Effects' / Slopes
		(1) \bar{x}_W	(2) \bar{x}_B		(3) $(\bar{x}_W - \bar{x}_B)$	(4) β_W	Sig	(5) β_B		
School Factors (H1)	School is in Top Half of Minority and FRPL Enrollments in Sample, Year 9	0.080	0.540	-0.460	0.160	**	0.076	*	0.035	-0.007
	School is in Top Half of Minority Enrollments, Year 9	0.010	0.100	-0.090	0.223		-0.008		-0.001	-0.002
	School is in Top Half of FRPL Enrollments in Sample, Year 9	0.070	0.090	-0.020	0.000		0.032		0.001	0.002
Behavior Factors (H2-H3)	Teacher-Reported Externalizing Problems Score, Year 5	11.170	12.780	-1.610	0.008	***	0.010	***	0.016	0.024
	Parent-Reported Externalizing Score, Year 5	10.360	11.290	-0.930	0.002		0.009	***	0.009	0.080
Controls	Family Income-to-Poverty Ratio, Year 5	3.300	1.430	1.870	-0.007		-0.026	*	0.049	-0.064
	Mother Has Some College or College Degree, Year 1	0.560	0.300	0.260	-0.036		-0.020		0.005	0.009
	Father Absent from Household at Any Wave, Year 5	0.490	0.850	-0.360	0.022		0.061		0.022	0.019
	Father has Ever Been in Jail or Prison, Year 9	0.380	0.610	-0.230	0.006		0.029		0.007	0.009
	Child's PPVT Cognitive Score, Year 5	102.780	89.660	13.120	0.002		0.003	*	-0.035	0.090
	Child's Age (in Months), Year 9	111.740	112.310	-0.570	0.001		-0.002		-0.001	-0.340
	Mother's Age, Year 1	27.020	24.070	2.950	-0.002		0.001		-0.002	0.068

Constant	1.000	1.000	0.000	-0.188	0.088	0.000	0.276
Observations (N)	368	880		368	880		
Overall Contribution of to the Racial Gap of Differences in Levels vs Slopes in Percentage-Point Units (/100):						0.105	0.163
Proportion of the Overall Race Gap Driven by Differences in Levels vs. Effects/Slopes:						0.391	0.601

*** p<0.001, **p<0.01, * p<0.05, + p<0.10 (two-tailed t-tests for a statistically significant difference from 0).

^aThis model uses Black boys' coefficients as the reference when calculating each variable's contribution to the gap in schooling due to racial differences in mean levels and Black boys' means as the reference when calculating each variable's contribution due to racial differences in coefficients (i.e., "effects").

^bValues in Column (6) are multiplied by -1 (to achieve positive values for gap-widening contributions and vice versa for gap-narrowing contributions).

^cValues in Column (7) are multiplied by -1 (to achieve positive values for gap-widening contributions and negative values for gap-narrowing contributions).

Source: The Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 1,248 Black and White boys who remained in the study from birth (wave 1) through year 9 (wave 5). Multiple imputation of 20 datasets is used to handle item-missingness on all but the dependent variable (suspension/expulsion), which is included in the imputation equation but for which observations with imputed values are deleted.

Table 3. Contributions of Racial Differences in School Composition (H1) and Child Behaviors at Years 5 and 9 (H2) and in the "Effects" of the Same Behaviors in Similar Schools (H3) to the Black/White Gap in Probability of Suspension or Expulsion by Year 9, Based on a Two-Way Decomposition Model (Reference: Blacks)^a

Factors	Predictor	Means		Differ- ence in Means (3) $(\bar{x}_W - \bar{x}_B)$	OLS Regression Coefficients			Contribution of Differences in Levels of Exposure (6) ^b $(\bar{x}_W - \bar{x}_B)\beta_B$	Contribution of Differences in 'Effects' / Slopes (7) ^c $(\beta_W - \beta_B)\bar{x}_B$	
		(1) \bar{x}_W	(2) \bar{x}_B		(4) β_W	Sig	(5) β_B			Sig
School Factors (H1)	School is in Top Half of Minority and FRPL Enrollments in Sample, Year 9	0.080	0.540	-0.460	0.153	**	0.066	0.030	-0.007	
	School is in Top Half of Minority Enrollments, Year 9	0.010	0.100	-0.090	0.218		0.004	0.000	-0.002	
	School is in Top Half of FRPL Enrollments in Sample, Year 9	0.070	0.090	-0.020	-0.007		0.028	0.001	0.002	
Behavior Factors (H2-H3)	Teacher-Reported Externalizing Problems Score, Year 5	11.170	12.780	-1.610	0.006	**	0.006	***	0.010	0.000
	Parent-Reported Externalizing Score, Year 5	10.360	11.290	-0.930	-0.001		0.006	**	0.006	0.073
	Teacher-Reported Externalizing Score, Year 9	26.040	34.980	-8.940	0.002		0.005	***	0.048	0.096
	Parent-Reported Externalizing Score, Year 9	13.890	13.740	0.150	0.003	*	0.004	**	-0.001	0.013
Controls	Family Income-to-Poverty Ratio, Year 5	3.300	1.430	1.870	-0.006		-0.020		0.037	-0.045
	Mother Has Some College or College Degree, Year 1	0.560	0.300	0.260	-0.042		-0.020		0.005	0.013
	Father Absent from Household at Any Wave, Year 5	0.490	0.850	-0.360	0.017		0.037		0.013	0.010
	Father has Ever Been in Jail or Prison, Year 9	0.380	0.610	-0.230	-0.001		0.005		0.001	0.002

Child's PPVT Cognitive Score, Year 5	102.780	89.660	13.120	0.001	0.003	**	-0.036	0.139	
Child's Age (in Months), Year 9	111.740	112.310	-0.570	0.001	-0.003		-0.001	-0.399	
Mother's Age, Year 1	27.020	24.070	2.950	-0.002	0.000		0.000	0.045	
Constant	1.000	1.000	0.000	-0.246	-0.030		0.000	0.216	
Observations (N)	368	880		368	880				
Overall Contribution of to the Racial Gap of Differences in Levels vs Slopes in Percentage-Point Units (/100):								0.112	0.156
Proportion of the Overall Race Gap Driven by Differences in Levels vs. Effects/Slopes:								0.418	0.582

*** p<0.001, **p<0.01, * p<0.05, + p<0.10 (two-tailed t-tests for a statistically significant difference from 0).

^aThis model uses Black boys' coefficients as the reference when calculating each variable's contribution to the gap in schooling due to racial differences in mean levels and Black boys' means as the reference when calculating each variable's contribution due to racial differences in coefficients (i.e., "effects").

^bValues in Column (6) are multiplied by -1 (to achieve positive values for gap-widening contributions and vice versa for gap-narrowing contributions).

^cValues in Column (7) are multiplied by -1 (to achieve positive values for gap-widening contributions and negative values for gap-narrowing contributions).

Source: The Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 1,248 Black and White boys who remained in the study from birth (wave 1) through year 9 (wave 5). Multiple imputation of 20 datasets is used to handle item-missingness on all but the dependent variable (suspension/expulsion), which is included in the imputation equation but for which observations with imputed values are deleted.

Figure 1. Lower-Bound Estimate of the Role of Behavior Differences: Teacher and Parent Behavior Ratings Only at Year 5

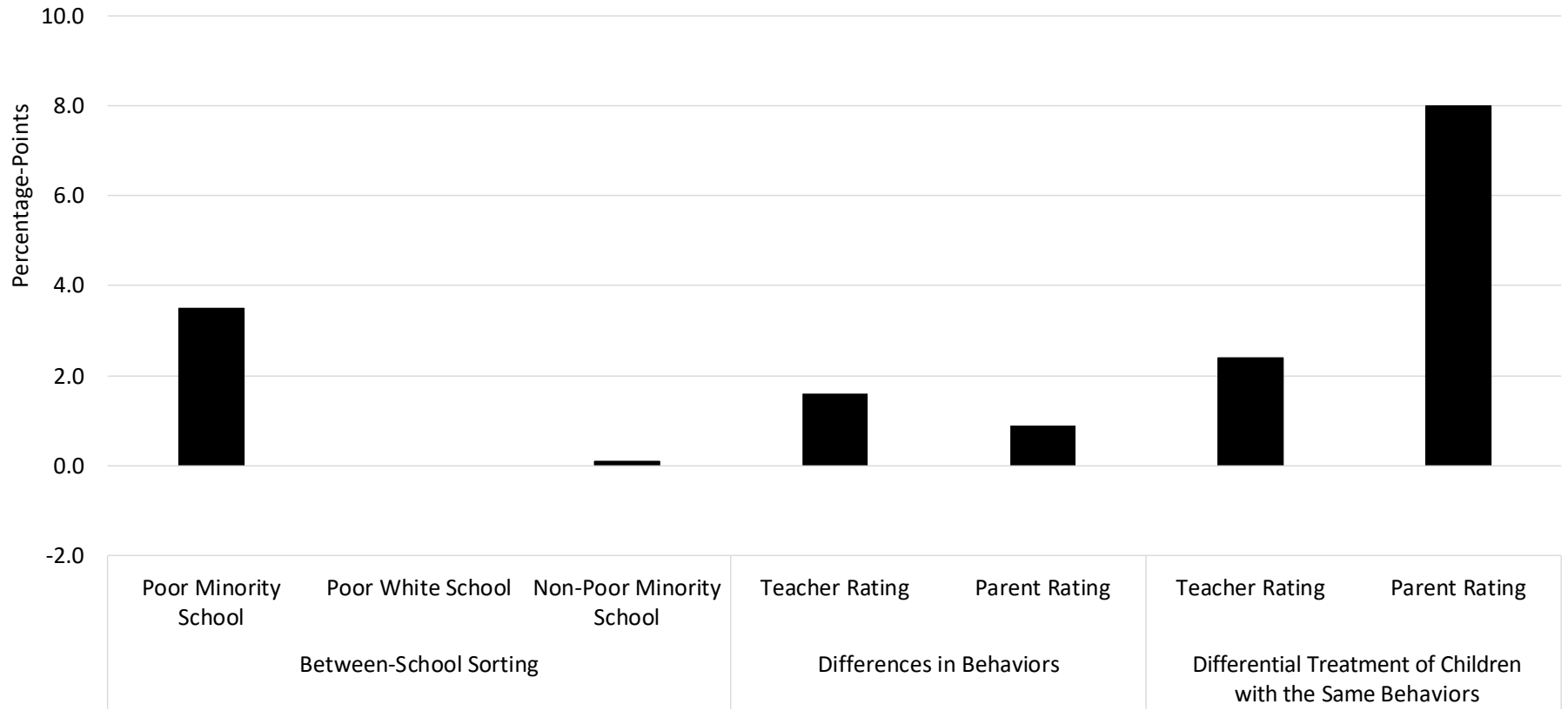


Figure 2. Differences in Behavior Trajectories of Black and White Boys Based on Teacher versus Parent Report

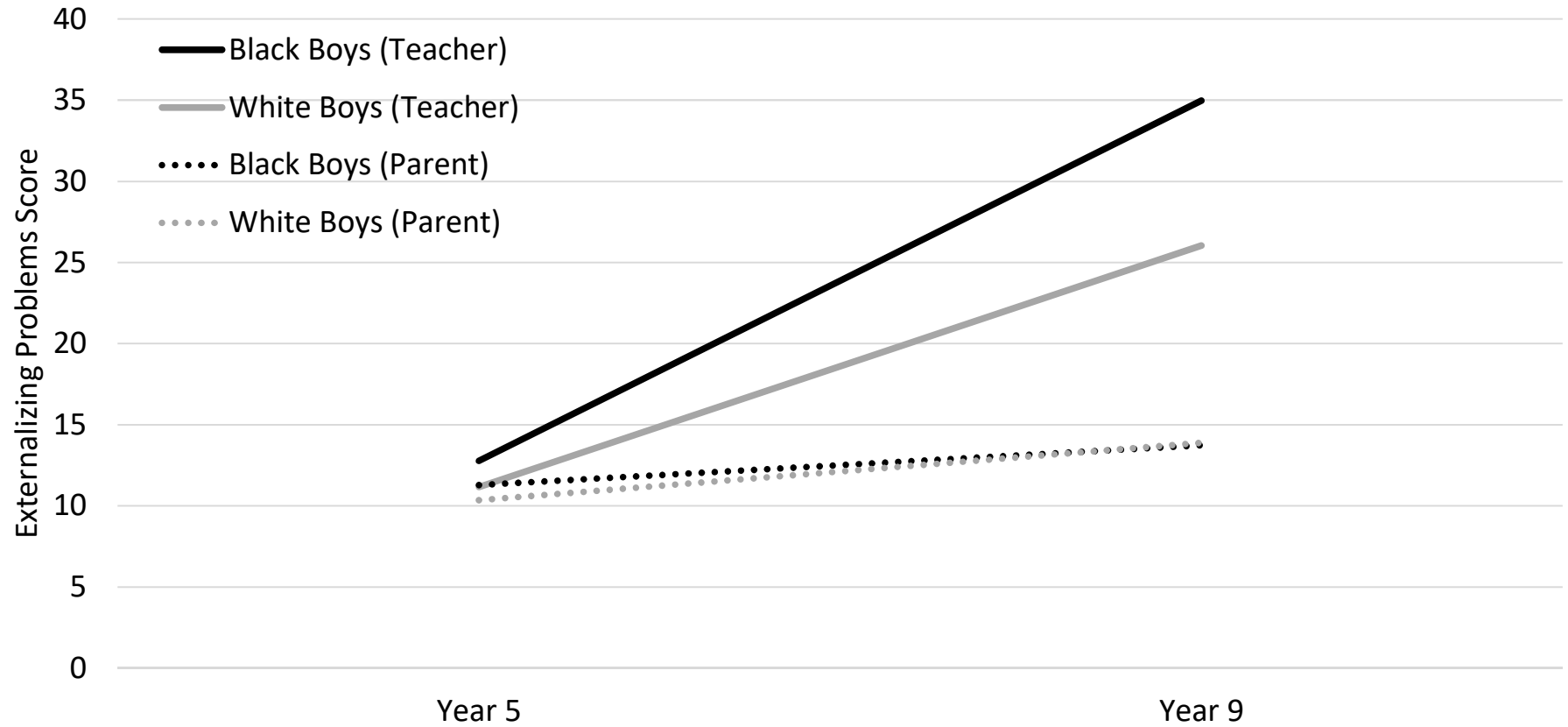
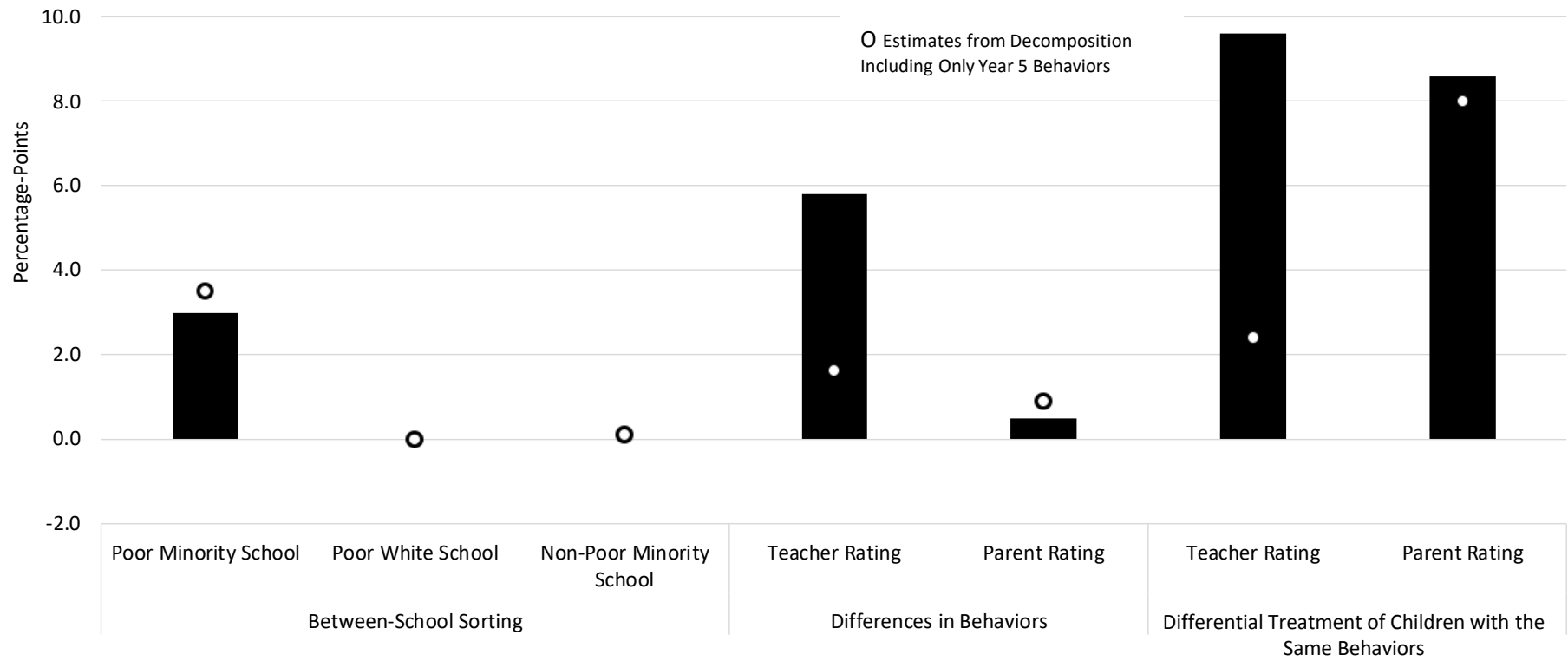


Figure 3. Upper-Bound Estimate of the Role of Behavior Differences: Including Teacher and Parent Behavior Ratings at Year 5 and Year 9



ENDNOTES

¹ Another related vignette experiment focuses on the question of how teacher race influences teacher evaluation and sanctions on behavior (Gilliam et al. 2016). Although we do not focus on teacher race effects in this study, these effects are captured through our differential treatment mechanism. Differential treatment can be driven by prejudice or “tough love” – the imposition of harsher sanctions on minority students by educators who are supporters of minority students but who believe that stiff punishments may help prepare students of color for the barriers they will encounter throughout society. Consistent with “tough love,” Gilliam et al. (2016) finds evidence for both differential behavior perception and sanctioning by student race, however, identifies Black teachers as being the harshest in their evaluations of Black students. Because of their specific focus on estimating the difference in scores from Black compared to White teachers, they do not display baseline ratings for the reference category (White teachers), thus making it difficult to infer average patterns of ratings across all teachers.

² Factors such as principals’ attitudes toward school exclusion can in theory set precedent for and help guide a school’s approach to discipline and may set precedent for non-punitive approaches to handling misbehavior even without adequate economic resources (Gregory, Cornell, and Fan 2011; Skiba, Edl and Rausch 2007), but principal attitudes are nonetheless likely to be correlated with school racial and socioeconomic composition.

³ Although extreme misbehavior or criminal activity are more highly predictive of suspension/expulsion, because of the infrequency of these acts, more minor behavioral infractions account for a larger share of school suspensions (Skiba, Chung, Trachok et al. 2014).

⁴ This study also has the significant limitation of comparing logit coefficients across nested models, including models with interaction terms (for issues with this approach, see Karlson, Holm, and Breen (2012)).

⁵ In most schools, teachers make referral decisions and at times can assign in-school suspension, whereas principals make final sanctioning decisions (Lewis and Diamond 2015). However, as the authority figures on the ground, teachers' accounts of disciplinary incidents are typically factored heavily into most higher-level administrative disciplinary decisions (Lewis and Diamond 2015).

⁶ A shortcoming of school discipline record data is that they include only students who have received some type of sanction. By excluding the (disproportionately white) students without a referral or discipline record in a given school or population cohort, prior studies of this kind tend to underestimate racial disproportionality, as discussed previously.

⁷ By later elementary school, teachers are more likely to have formed reputational assessments of students, which may disproportionately affect black students (Okonofua and Eberhardt 2015). Therefore, we treat behaviors measured at school entry as providing a lowerbound estimate of the role of behavior differences in contributing to the racial gap, while behaviors measured at the end of our observational period provide upperbound estimates of the role of behavior differences. The latter provide upperbound estimates because they make the assumption that all behavior change since school entry occurred prior to any suspension.

⁸ That teachers and parents provide ratings of children's every day behaviors on detailed behavioral indices provides a more nuanced and realistic prospective picture of teacher's impressions of a child. This nuanced picture of behavior is crucial because teachers' decisions about whether to initiate sanctioning for any specific incident does not happen in a vacuum, but

rather reflect the broader impressions of a child that develop over a series of interactions (Vavrus and Cole 2002).

⁹ Sixteen cities were drawn from a stratified random sample of cities of 200,000 or more, and four additional cities were selected for special interest to specific foundations. Hospitals within these 20 cities were then sampled after stratifying on labor market conditions and policy environments (see Reichman et al. 2001). Within hospitals, marital and non-marital births were randomly selected with a three to one oversample of non-marital births. The unweighted sample therefore over-represents the relatively disadvantaged families whose children are more likely to experience school suspension and expulsion at an early age, making it ideal for examining the drivers of racial disproportionality between relatively disadvantaged black compared to white children.

¹⁰ Response rates for these follow-up waves were 91% (year 1), 88% (year 3), 87% (year 5), and 76% (year 9) of the baseline sample. These rates reflect sample attrition (n=1,122), ineligibility due to child's death or adoption (n=159), or living less than half-time with biological parents (n=102).

¹¹ More information on the merged CCD/PSS variables can be found at:

<http://fragilefamilies.princeton.edu/restricted/school>. More information on the CRDC school context data is available at: <https://ocrdata.ed.gov/>.

¹² Despite their advantages, a limitation of these behavior measures is that they are unable to identify the specific event that culminated in a referral or suspension. However, prior work helps address this limitation by showing that Black elementary school students are more likely to be suspended than White students *for the same rule violation* (and in the exact same elementary school) (Kinsler 2011).

¹³ Because the decision to begin collecting kindergarten teacher surveys was made halfway through the year 5 data collection, time and budgetary constraints meant that the rule-breaking sub-scale was not separately fielded. However, key items from this sub-scale were folded into the social problems and aggression sub-scales, so a truncated version of the rule-breaking battery is still included in our year 5 kindergarten teacher externalizing scale.

¹⁴ 63% of the children in the year 9 sample do not attend schools with any other children in the FFCWS sample, 20% attend schools with only one other child in the FFCWS sample, and an additional 9% attend schools with two other FFCWS children. Due to the limited clustering of FFCWS children within the same schools, we are unable to estimate a meaningful school fixed effects model. However, prior work using school fixed effects shows that Black elementary school students *in the exact same school* are more likely to be suspended than White students for the same rule violations (Kinsler 2011).

¹⁵ The term ‘effects’ should not be interpreted causally. Rather, ‘effects’ in the decomposition framework refers to differences in slopes or coefficients between Black and White males, under the counterfactual in which they experience the same levels of exposure to a given factor.

¹⁶ Year 5 teacher reports of child behavior would *not* precede any suspension/expulsion that occurred in pre-school, but less than 1% of children are suspended prior to kindergarten. Additionally, prior research shows that student reputations formed by a prior suspension have much less sticking power between schools rather than within schools (Kinsler 2011). Reputations are likely even less sticky at such young ages as the transition between daycare/preschool and primary school.

¹⁷ Even in the absence of a suspension or expulsion, Black children’s behavior problems may increase more than White children’s, potentially due to greater challenges in or outside school,

such as family instability, exposure to violence, or subtle or overt negative treatment by teachers and school officials (Hemphill, Toumbourou, Herrenkohl et al. 2006; Mittleman 2017; Okonofua and Eberhardt 2015).

¹⁸ Results shown in Appendix Table A.3 indicate that there is a significant interaction between Black*parent-reported externalizing problems score at year 5. This interaction is statistically significant even when controlling for school racial and socioeconomic composition, regardless of whether school composition is measured based on continuous or categorical measures of school racial and socioeconomic composition. This indicates that Black boys are more likely to be suspended with the same behavior ratings in similar schools.

¹⁹ Strikingly, of the remaining 38.8% of the gap that would not be explained by hypotheses 1-3, the majority would instead be attributed to the higher suspension of Black boys than White boys with higher levels of academic ability at school entry (i.e., cognitive development as measured by PPVT score), with older mothers, and with a history of biological father absence from the home. We investigate these drivers in other work, but here focus on results that pertain to our main hypotheses.

²⁰ In our sample, roughly 80% of White boys attend schools that serve non-poor White populations. Therefore, when treating White boys' means and coefficients as the reference, we include the dummy variable for "non-poor White school" in our decomposition and instead treat non-poor minority school as the omitted category/reference group.

²¹ Important research by Gilliam et al. (2016) highlights that racial disparities are also present in pre-kindergarten, however, data come from a single pre-kindergarten provider in the Northeast, and rates of suspension/expulsion are much lower – roughly 1% of children are suspended or expelled from pre-kindergarten.

APPENDIX: SUPPLEMENTARY ANALYSES

We provide several supplementary analyses that provide greater context for our results. Specifically, we ask: 1) How does missing data on teacher reports of child behaviors at year 5 shape results? 2) Do our result change if we allow for a non-linear relationship between school racial and socioeconomic composition and child suspension? 3) Is our assumption that the association between school racial and socioeconomic composition and suspension is due to differences in punitive policies consistent with the data? 4) Do our results change when we employ the non-linear decomposition method designed by Fairlie (2005)?

1) How does missing data on year 5 teacher reports of child behavior shape results?

Teacher ratings of behavior at year 5 were only available for 1,039, or 30%, of the 3,515 families that participated in the year 9 survey. This is because the decision to conduct teacher surveys occurred over halfway through the year 5 survey. Analysis of patterns of item-missingness reveals that missingness on teacher reports of child behavior at school entry (year 5) was higher among older children, children with a history of father absence, and children who attended schools with higher suspension rates. To deal with this possible bias, we conduct supplementary analyses, using children with complete cases on year 5 teacher-rated behaviors (Appendix Table A.1). Substantive patterns of results remain unchanged.

2) Do our result change if we allow for a non-linear relationship between school racial and socioeconomic composition and child suspension?

Based on prior research, we examine two different methods of operationalizing school racial and socioeconomic composition. First, prior research highlights that suspension/expulsion is more widely used in schools with large enrollments of students from minority and low-income

backgrounds. Second, another set of studies suggest an additional disadvantage for children in schools serving a majority of students who are *both* poor and minority. We test this claim by comparing models 2 and 3 of Appendix Table A.2. We find that students who attend schools that enroll *either* high percentages of Black and Hispanic students *or* high percentages of students receiving free-or-reduced-price lunch (FRPL) are not more likely to be suspended than other students (model 1), whereas students who attend schools serving students who are *both* poor and minority are more likely to be suspended. Relative to model 1, both the linear (model 2) and non-linear (model 3) specifications of the percent minority*percent FRPL interaction mediate similar amounts (17%-18%) of the racial gap in suspension/expulsion and explain similar amounts of variance in suspension (approximately 25%). In addition, both the continuous measures (percent Black or Latino/Hispanic and percent FRPL) and the dummy indicator for high minority and high FRPL enrollments are similarly correlated with the CRDC measure for school-level suspension and expulsion rate (all three correlation coefficients are approximately 0.45).

For our purposes, the non-linear specification has the advantage of avoiding the need for a three-way interaction between school percent minority*school percent FRPL*student Black, which would complicate interpretation of the decomposition results. For this reason we use a set of dummy variables that indicate whether a student attends a school in the top 50% of Black and Hispanic enrollments, the top 50% of free or reduced-price lunch enrollments (FRPL), or both. Results are robust to the use of continuous measures of school racial and FRPL composition (see Appendix Table A.3).

- 3) *Is our assumption that the association between school racial and socioeconomic composition and suspension is due to differences in punitive policies consistent with the data?*

The premise of prior research on school racial and socioeconomic composition is that schools serving higher percentages of minority and poor students (or higher percentages of students that are both poor and minority) are more likely to practice punitive forms of discipline (Welch and Payne 2010). We test this claim by linking school-level rates of suspension and expulsion in 2009-11, taken from the CRDC data described in the main text, to the FFCWS survey data. As shown in Appendix Table A.4, school-level discipline mediates the association between school racial and socioeconomic composition and suspension. Importantly, we do not include the school-level suspension and expulsion rate variable in our decomposition analysis because prior work points to a causal process whereby poor and minority school enrollments *lead to* (i.e., are endogenous to) higher levels of child suspension at the individual level, which in turn produces higher suspension/expulsion rates at the school level. By contrast, to include school-level suspension rates in our decompositions, we would need to assume that school discipline is causally prior to both child behavior/suspension and school racial and socioeconomic composition. As this is not the claim of prior research, we instead control only for factors that are causally prior to a child's behavior/suspension at the individual level.

4) *Do our results change when we employ the non-linear decomposition method?*

Fairlie (2005) rightfully points out that the Oaxaca-Blinder decomposition method may *underestimate* the contribution of differences in levels of exposure between groups if the racial gap in suspension occurs at the tails of the distribution, or when there are large race differences in predictor variables within the decomposition. As such, we conduct the extension proposed by Fairlie (2005) in order to examine whether estimates of contributions due to differences in levels of exposure are similar. We find similar results using both the linear decomposition and the

nonlinear extension (see Appendix A.6). Given this finding, and because the extension developed by Fairlie (2005) does not provide a straightforward method for differentiating between the contributions of differences in levels of exposure from differences in coefficients, which are central for testing our hypotheses, we report results from the Oaxaca-Blinder decomposition in our main analyses.

Appendix Table A.1. Oaxaca-Blinder Decomposition of the Mean Race Gap in Suspension Among Black and White Boys with Complete Responses on All Variables in Analyses (Reference=Blacks)^a

Overall Decomposition Results		
White Suspension Rate	0.094***	
	(0.027)	
Black Suspension Rate	0.330***	
	(0.035)	
Mean Percentage-Point Difference in Suspension Rates	-0.235***	
	(0.044)	
Percentage-Points Attributable to Differences in Levels of Exposure	-0.153	
	(0.111)	
Percentage-Points Attributable to Differences in Coefficients	-0.082	
	(0.117)	
Detailed Decomposition Results		
	Levels Diff.	Coef. Diff.
School is both in Top Half of Minority and FRPL Enrollments Year 9	-0.153	0.138
	(0.082)	(0.105)
School is in Top Half of FRPL Enrollments in Sample Year 9	0.001	0.013
	(0.007)	(0.017)
School is in Top Half of Minority Enrollments in Sample Year 9	-0.015	0.031
	(0.012)	(0.024)
Teacher-Rated Externalizing Problems Score Year 5	-0.003	0.023
	(0.015)	(0.045)
Parent-Rated Externalizing Problems Score Year 5	0.009	-0.111
	(0.012)	(0.074)
Family Income-to-Poverty Ratio Year 5	0.040	-0.001
	(0.065)	(0.010)
Mother Has Some College or College Degree Year 1	-0.024	-0.010
	(0.033)	(0.018)
Father Absent From Household at Any Wave Year 5	0.051	-0.093
	(0.063)	(0.135)
Father has Ever Been in Jail or Prison Year 9	-0.014	0.032
	(0.032)	(0.075)
Child's PPVT Cognitive Score Year 5	-0.060	-0.253
	(0.062)	(0.238)
Child's Age Year 5	0.007	-0.861
	(0.011)	(1.499)
Mother's Age Year 1	0.009	0.017
	(0.052)	(0.167)
Constant		0.993
		(1.598)
Observations (N) ^b		321

*** p<0.001, **p<0.01, * p<0.05, + p<0.10 (two-tailed t-tests for a statistically significant difference from 0).

^aThis model uses Black boys' coefficients as the reference when calculating each variable's contribution to the gap in schooling due to racial differences in mean levels and Black boys' means as the reference when calculating each variable's contribution due to racial differences in coefficients (i.e., "effects").

^bComplete case analysis, including teacher ratings of behaviors at year 5. *Source: Fragile Families, Waves 1-5.*

Appendix Table A.2. Is there an Additional Penalty of Attending a School that is Both Majority-Poor and Majority-Minority Serving Above and Beyond the Increased Risk of Child Suspension and Expulsion Associated with Attending a School that is Either Majority-Poor or Majority-Minority?

	(1)	(2)	(3)
Non-Hispanic Black	0.180*** (0.033)	0.179*** (0.033)	0.182*** (0.030)
Percent of School Enrollment Black or Hispanic, Year 9 (/10)	0.001 (0.001)	0.001+ (0.001)	
Percent of School Enrollment FRPL, Year 9 (/10)	0.001 (0.001)	0.002* (0.001)	
Percent Black*Percent FRPL (/10)		0.000* (0.000)	
School is in Top Half of Minority Enrollments in Sample, Year 9			0.012 (0.054)
School is in Top Half of FRPL Enrollments in Sample, Year 9 (i.e., Poor White School)			0.034 (0.045)
School is both in Top Half of Minority and FRPL Enrollments, Year 9			0.099** (0.033)
Teacher-Rated Externalizing Problems Score, Year 5	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)
Parent-Rated Externalizing Problems Score, Year 5	0.007*** (0.002)	0.007*** (0.002)	0.007*** (0.002)
Family Income-to-Poverty Ratio, Year 5	-0.011* (0.005)	-0.012* (0.005)	-0.012* (0.005)
Mother Has Some College or College Degree, Year 1	-0.030 (0.029)	-0.029 (0.029)	-0.028 (0.029)
Father Absent From Household at Any Wave, Year 5	0.037 (0.029)	0.037 (0.029)	0.038 (0.029)
Father has Ever Been in Jail or Prison, Year 9	0.028 (0.028)	0.029 (0.028)	0.028 (0.028)
Child's PPVT Cognitive Score, Year 5	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)
Child's Age, Year 5	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)
Mother's Age, Year 1	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)
Constant	0.404*** (0.056)	0.418*** (0.057)	0.305*** (0.047)
Observations	1,248	1,248	1,248
R-squared	0.171	0.173	0.174

*** p<0.001, **p<0.01, * p<0.05, + p<0.10 (two-tailed t-tests for a statistically significant difference from 0).

Source: The Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 1,248 Black and White boys who remained in the study from birth (wave 1) through year 9 (wave 5). Multiple imputation of 20 datasets is used to handle item-missingness on all but the dependent variable (suspension/expulsion), which is included in the imputation equation but for which observations with imputed values are deleted.

Appendix Table A.3. LPM Predicting Child Ever Suspended/Expelled, Year 9: Estimates of 'Differential Treatment' by Race for the Same Behavior Problems, Additionally Including Main Effects and Interactions Between Race and School Context and Family Factors, Among Black and White Boys

	(1)	(2)
Black	0.113 (0.081)	0.228* (0.099)
<i>Main effects for Behavior Problems, Year 5</i>		
Teacher-Rated Externalizing Problems Score, Year 5	0.008*** (0.002)	0.008*** (0.002)
Parent-Rated Externalizing Problems Score, Year 5	0.001 (0.003)	0.002 (0.003)
<i>Interactions between Race and Behavior Problems, Year 5</i>		
Black*Teacher-Rated Externalizing Problems Score, Year 5	0.002 (0.003)	0.002 (0.003)
Black*Parent-Rated Externalizing Problems Score, Year 5	0.008* (0.004)	0.008* (0.004)
<i>Continuous Measures of School Context, Interactions with Child Race (Black)</i>		
Continuous Measures of School % Black/Hispanic, School %FRPL and Interaction %Black/Hispanic*% FRPL, Year 9	X	
Black* School % Black/Hispanic, Black*School % FRPL and Black*School % Black/Hispanic*School % FRPL, Year 9	X	
<i>Dummies for Non-Linear Effects of School Context, Interactions with Black</i>		
Dummies for Poor Minority School, Poor White School, Non-Poor Minority School (ref=Non-Poor White School), Year 9		X
Black* Dummies for Poor Minority School, Poor White School, Non-Poor Minority School (ref=Black*Non-Poor White School), Year 9		X
<i>Family and Child Controls, Interactions with Child Race (Black)</i>		
Main Effects for Family Social Class, Year 5	X	X
Interactions between Race and Family Social Class, Year 5	X	X
Controls for Child's Standardized PPVT Score and Age, and Mother's Age, Year 5	X	X
Constant	0.305*** (0.062)	0.239*** (0.063)
Observations	1,248	1,248
R-squared	0.180	0.182

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

¹Note that the coefficients on the main effect for Black cannot meaningfully be compared across models because the differing Black*[school racial composition] interactions across models yield distinct reference categories. Importantly, one cannot claim that the racial gap is smaller with the linear specification using the continuous variables relative to the non-linear specification using a series of dummy variables.

Source: Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 1,248 Black and White males who remained in the study from birth (wave 1) through age 9 (wave 5). Multiple imputation of 20 datasets is used to handle item-missingness on all but the dependent variable (suspension/expulsion), which is included in the imputation equation but for which observations with imputed values are deleted.

Appendix Table A.4. Linear Probability Models (LPM) Predicting School Suspension or Expulsion by Year 9: Does School Punitiveness Mediate School Racial and SES Composition?

	(1)	(2)	(3)
Black	0.152*** (0.034)	0.148*** (0.034)	0.146*** (0.034)
School is in Top Half of Minority Enrollments in Sample, Year 9	0.033 (0.055)	0.006 (0.056)	-0.025 (0.057)
School is in Top Half of FRPL Enrollments in Sample, Year 9 (Poor White School)	0.061 (0.049)	0.036 (0.051)	0.011 (0.051)
School is both in Top Half of Minority and FRPL Enrollments, Year 9	0.160*** (0.035)	0.132*** (0.038)	0.101** (0.038)
School-Level Proportion of Students Suspended Out of School or Expelled, Year 9		0.515* (0.250)	1.530*** (0.405)
School-Level Proportion of Students Suspended Out of School or Expelled, Year 9, Squared (Quadratic)			-2.770** (0.942)
Teacher-Rated Externalizing Problems Score, Year 5	0.011*** (0.002)	0.010*** (0.002)	0.010*** (0.002)
Parent-Rated Externalizing Problems Score, Year 5	0.007*** (0.002)	0.007** (0.002)	0.006** (0.002)
Father Absent from Household at Any Wave, Year 5	0.045 (0.031)	0.044 (0.031)	0.044 (0.031)
Family Income-to-Poverty Ratio, Year 5	-0.014* (0.006)	-0.013* (0.006)	-0.012* (0.006)
Mother Has Some College or College Degree, Year 1	-0.021 (0.032)	-0.015 (0.032)	-0.014 (0.032)
Child's PPVT Cognitive Score, Year 5	0.003** (0.001)	0.003*** (0.001)	0.003*** (0.001)
Child's Age, Wave 5	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)
Mother's Age, Year 1	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)
Constant	0.089** (0.033)	0.074* (0.033)	0.054 (0.033)
Observations	1,248	1,248	1,248
R-squared	0.183	0.189	0.194

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

Source: The Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 1,248 Black and White boys who remained in the study from birth (wave 1) through year 9 (wave 5). Multiple imputation of 20 datasets is used to handle item-missingness on all but the dependent variable (suspension/expulsion) and on measures of school composition. The outcome variable was included in the imputation equation but observations with imputed values were deleted before analysis.

Appendix Table A.5. Contributions of Racial Differences in School Composition (H1) and Child Behaviors at Years 5 and 9 (H2) and in the "Effects" of the Same Behaviors in Similar Schools (H3) to the Black/White Gap in Probability of Suspension or Expulsion by Year 9, Based on a Two-Way Decomposition Model (Reference: Whites)^a

Overall Decomposition Results		
White Suspension Rate	0.103*** (0.016)	
Black Suspension Rate	0.372*** (0.016)	
Mean Percentage-Point Difference in Suspension Rates	-0.268*** (0.023)	
Percentage-Points Attributable to Differences in Levels of Exposure	-0.128*** (0.035)	
Percentage-Points Attributable to Differences in Coefficients	-0.140*** (0.040)	
Detailed Decomposition Results		
	Diff. in Levels	Diff. in Coefs.
School is in Top Half of Minority Enrollments in Sample Year 9 (i.e., Poor Minority School)	-0.006 (0.016)	0.013 (0.018)
School is in Top Half of FRPL Enrollments in Sample Year 9 (i.e., Poor White School)	0.002 (0.003)	-0.010 (0.008)
School is Neither in Top Half of Minority nor Top Half of FRPL Enrollments Year 9 (i.e., Non-Poor White School)	-0.086** (0.033)	-0.024 (0.019)
Teacher-Rated Externalizing Problems Score Year 5	-0.010* (0.005)	-0.000 (0.031)
Parent-Rated Externalizing Problems Score Year 5	0.001 (0.002)	-0.079* (0.039)
Teacher-Rated Externalizing Problems Score Year 9	-0.015 (0.011)	-0.129* (0.055)
Parent-Rated Externalizing Problems Score Year 9	0.001 (0.002)	-0.013 (0.032)
Family Income-to-Poverty Ratio Year 5	-0.011 (0.010)	0.020 (0.019)
Mother Has Some College or College Degree Year 1	-0.011 (0.010)	-0.007 (0.016)
Father Absent From Household at Any Wave Year 5	-0.006 (0.013)	-0.017 (0.049)
Father has Ever Been in Jail or Prison Year 9	0.000 (0.009)	-0.004 (0.031)

Child's PPVT Cognitive Score Year 5	0.019 (0.015)	-0.122 (0.140)
Child's Age Year 5	-0.001 (0.002)	0.401 (0.582)
Mother's Age Year 1	-0.004 (0.009)	-0.040 (0.100)
Constant		-0.130 (0.628)

Observations	1248
--------------	------

*** p<0.001, **p<0.01, * p<0.05, + p<0.10 (two-tailed t-tests for a statistically significant difference from 0).

^aThis model uses White boys' coefficients as the reference when calculating each variable's contribution to the gap in schooling due to racial differences in mean levels and White boys' means as the reference when calculating each variable's contribution due to racial differences in coefficients (i.e., "effects"). Note that non-poor, White schools are now observed in the model (and non-poor minority schools are the omitted category) since a majority of the White boys who are now the reference group attend non-poor White schools.

Source: The Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 1,248 Black and White boys who remained in the study from birth (wave 1) through year 9 (wave 5). Multiple imputation of 20 datasets is used to handle item-missingness on all but the dependent variable (suspension/expulsion), which is included in the imputation equation but for which observations with imputed values are deleted.

Appendix Table A.6. Comparison of Fairlie Non-Linear Decomposition and Oaxaca-Blinder Linear Decomposition Estimates of the Drivers of the Racial Gap in Suspension/Expulsion at Year 9^a

	(1)	(2)
	Fairlie Non-Linear Decomposition	Oaxaca-Blinder Decomposition
	<i>Contributions of Differences in Levels (i.e., Means)</i>	
School is both in Top Half of Minority and FRPL Enrollments Year 9	-0.028 (0.017)	-0.030 (0.017)
School is in Top Half of FRPL Enrollments in Sample Year 9	-0.000 (0.001)	-0.000 (0.001)
School is in Top Half of Minority Enrollments in Sample Year 9	0.000 (0.004)	-0.000 (0.005)
Teacher-Rated Externalizing Problems Score Year 5	-0.010*** (0.003)	-0.010* (0.005)
Parent-Rated Externalizing Problems Score Year 5	-0.005* (0.002)	-0.006 (0.003)
Teacher-Rated Externalizing Problems Score Year 9	-0.046*** (0.009)	-0.048*** (0.011)
Parent-Rated Externalizing Problems Score Year 9	0.001 (0.001)	0.001 (0.003)
Family Income-to-Poverty Ratio Year 5	-0.032 (0.018)	-0.037 (0.023)
Mother Has Some College or College Degree Year 1	-0.004 (0.009)	-0.005 (0.010)
Father Absent From Household at Any Wave Year 5	-0.010 (0.013)	-0.013 (0.016)
Father has Ever Been in Jail or Prison Year 9	-0.001 (0.007)	-0.001 (0.008)
Child's PPVT Cognitive Score Year 5	0.031* (0.012)	0.036** (0.014)
Child's Age Year 5	0.001 (0.002)	0.001 (0.002)
Mother's Age Year 1	0.000 (0.007)	0.000 (0.009)
Observations	1248	1248

*** p<0.001, **p<0.01, * p<0.05, + p<0.10 (two-tailed t-tests for a statistically significant difference from 0).

^aThese models use Black boys' coefficients as the reference.

Source: The Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 1,248 Black and White boys who remained in the study from birth (wave 1) through year 9 (wave 5). Multiple imputation of 20 datasets is used to handle item-missingness on all but the dependent variable (suspension/expulsion), which is included in the imputation equation but for which observations with imputed values are deleted.