Unpacking the Drivers of Racial Disparities in School Suspension and Expulsion

Jayanti Owens, *Brown University*
Sara S. McLanahan, *Princeton University*

School suspension and expulsion are important forms of punishment that disproportionately affect Black students, with long-term consequences for educational attainment and other indicators of wellbeing. Prior research identifies three mechanisms that help account for racial disparities in suspension and expulsion: between-school sorting, differences in student behaviors, and differences in the treatment and support of students with similar behaviors. We extend this literature by (1) comparing the contributions of these three mechanisms in a single study, (2) assessing behavior and school composition when children enter kindergarten and before most are exposed to school discipline, and (3) using both teacher and parent reports of student behaviors. Decomposition analyses reveal that differential treatment and support account for 46 percent of the Black/White gap in suspension/expulsion, while between-school sorting and differences in behavior account for 21 percent and 9 percent of the gap respectively. Results are similar for boys and girls and robust to the use of school fixed effects and measures of school composition and student behavior at ages 5 and 9. Theoretically, our findings highlight differential treatment/support after children enter school as an important but understudied mechanism in the early criminalization of Black students.

This work has been supported by the National Academy of Education/Spencer Foundation, the Population Studies and Training Center at Brown University, which receives funding from the NIH (P2C HD041020), the Robert Wood Johnson Foundation Health and Society Scholars Program at the University of Wisconsin-Madison, and the Office of Population Research and Center for Research on Child Wellbeing (CRCW) at Princeton University, which receives funding from the NIH (P2CHD047879). All content is the responsibility of the authors and does not necessarily reflect the views of funding organizations. The authors thank Laura Doering, Louis Donnelly, Daniel Hirschman, Kate Jaeger, Joel Mittleman, Garrett Pace, Shiva Rouhani, and Germán Rodríguez for invaluable feedback on this manuscript. This paper also benefited from presentations at the Population Association of America and the American Sociological Association Annual Meetings in 2017, the Department of Education at Brown University, the Departments of Sociology at Boston College, and the University of Notre Dame, the Steinhardt School at New York University, the Annenberg Institute at Brown University, and the CRCW at Princeton University. Direct correspondence to Jayanti Owens, 108 George St., Maxcy Hall, Brown University, Box 1916, Providence, RI 02912; email: jayanti_owens@brown.edu

Social Forces 1–30
doi: 10.1093/sf/soz095

Introduction

Schools' use of exclusionary discipline tactics, such as suspension and expulsion, increased by nearly 50 percent over the last forty years. In 1980, 12 percent of 8th–10th grade students reported having been suspended at some point in their lives. By 2006, this figure had increased to 18 percent (Bertrand and Pan 2013). Although suspension rates have been on the decline since 2010-11, absolute levels remain high (Office for Civil Rights 2014). In the 2011-12 academic year, 10 percent of students in kindergarten through 12th grade—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 percent) when in-school suspensions are counted (Losen et al. 2015; Wallace et al. 2008).

Aggregate rates of suspension and expulsion mask considerable heterogeneity by race and gender. As compared to White students, Black students are 3.2 times more likely to be suspended or expelled, Native American students are 2.0 times more likely, and Hispanic/Latinx students are 1.3 times more likely (Government Accountability Office 2018). Although girls of each racial/ethnic group experience roughly 50 percent lower rates of out-of-school suspension and expulsion than boys, racial gaps among Black, Latinx, or Native American versus White girls are similarly large in proportionate terms (Morris and Perry 2017; Office for Civil Rights 2014).

Using a range of datasets, estimation strategies, and outcomes, researchers have shown that being suspended or expelled from school, and the associated negative labeling and loss of instruction, is associated with poor school performance and a higher risk of school dropout, arrest, incarceration, and unemployment (Fabelo et al. 2011; Mittleman 2018; Wolf and Kupchik 2017). Studies using within-student variation, which controls for stable unobserved risk factors for suspension and expulsion, reach similar conclusions (Morris and Perry 2016). Since the effects of suspension and expulsion are likely to accumulate over time, the earlier in the life course they occur, the more negative the consequences are likely to be (DiPrete and Eirich 2006).

Some scholars have pointed to the existence of a "school-to-prison pipeline" to highlight the parallel between school disciplinary practices 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 youth, 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 siloed. Specifically, whereas "differential treatment/support on the basis of race" 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, hereafter referred to as "suspension." Instead, sociologists have focused on structural discrimination in the form of differences in the characteristics of the schools that Black and White children

attend (the "between-school sorting" mechanism) and differences in students' behaviors (the "behavior differences" mechanism) (Skiba and Williams 2014; Skiba et al. 2002).

Understanding the relative importance of these three explanations is important, not only for developing theories about the processes underlying racial disparities but also for developing effective policies and practices. If differences in students' school-entry behavior are a key factor, training teachers to more effectively manage behaviors may serve as an important entry point, whereas if differences in school composition and policies are the key driver, rethinking disciplinary systems in schools serving minority and poor populations would be an appropriate starting point. Finally, if differential treatment/support is critical, reducing educators' implicit and explicit biases and increasing children's access to services designed to support positive behavior as children progress through school would be a positive first step.

In this paper, we examine the relative contributions of these three mechanisms. We use data from the Fragile Families and Child Wellbeing Study (FFCWS), a population-based birth cohort study of children born in large U.S. cities at the turn of the 21st century. The sample design called for a large oversample of children born to unmarried parents, who are more likely to experience various forms of disadvantage, including poverty, family structure and housing instability, and underserved school environments. Because these characteristics place children at higher risk for suspension, the oversample of disadvantaged families enables greater precision in the estimation of the drivers of racial disparities in suspension.

We focus on Black and White children since Black students experience the highest rates of both school suspension and criminal justice contact relative to other groups (e.g., Latinos and Native Americans) (Morris and Perry 2017; Office for Civil Rights 2014; Skiba et al. 2002). Additionally, because studies suggest that the mechanisms described above may operate differently for boys and girls (Goff et al. 2014), we conduct separate analyses by gender. We measure behavior at age 5, when children are in kindergarten and before most students are exposed to school disciplinary practices. We use both parent and teacher reports of children's behavior, and we focus on overall behavior rather than a specific infraction.

Explanations for the Racial Gap in Suspension and Expulsion

Differences in School Composition

One prominent explanation for racial disparities in suspension is differences in school composition or what researchers refer to as *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' judgments 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 students are disproportionately from low-income families (Welch and Payne 2010).

Moreover, Ramey (2015) finds that schools serving either majority-minority or a combination of majority-minority and poor students are more likely to use *exclusionary* discipline tactics like suspension and expulsion or arrest. Finally, Kinsler (2011) finds that the racial gap in school suspension *conditional on referral to the principal's office* 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 the time children are age 9.*

Differences in Children's Behavior

An alternative perspective argues that racial differences in suspension are due to *differences in students'* behaviors, such as rule-breaking and aggression, inability to pay attention, and inability to get along with peers and teachers (Gregory, Skiba and Noguera 2010; Raffaele-Mendez 2003). Racial differences in students' 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 and extra-school experiences (Bates et al. 1991; Brooks-Gunn and Duncan 1997; Dance 2002; Deater-Deckard and Dodge 1997; Magnuson and Waldfogel 2005; Robinson and Harris 2014).

Existing research suggests that differences in student behaviors account for only a small share of the racial gap in suspension (Anyon et al. 2016; Kinsler 2011; Skiba et al. 2014). This result is based largely on studies that condition on being referred to the principal's office for misbehavior and that model the association between race and the severity of sanction net of behavioral infraction type (Gregory, Skiba and Noguera 2010; Skiba and Williams 2014; Skiba et al. 2002, 2014). Note, however, that if Black students are more likely than White students to be referred to the principal for minor infractions that are less likely to warrant suspension, as some studies suggest, conditioning on referral will lead to an underestimate of the association between race and severity of sanction.

In contrast, studies that use population-based data and do not condition on referral to the principal's office find that student behavior is strongly associated with suspension (Bradshaw et al. 2010; Rocque 2010; Shollenberger 2015). As examples, Raffaele-Mendez (2003) 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. Wright et al. (2014) find that racial differences in behaviors between school entry and 4th grade account for some but not all of the gap in suspension by 8th grade. Importantly, both of these studies measure behavior *after* the child enters school, which raises questions about the causal ordering of behavior and school punishment. Insofar as behavior is endogenous to how children are treated by teachers and school officials, and insofar as Black children are treated more punitively, the studies described above would overstate the extent to which racial differences in behavior account for differences in suspension (Jacobsen, Pace and Ramirez 2018; Okonofua and Eberhardt 2015; Okonofua, Walton and Eberhardt 2016). Given that we measure behavior prior to suspension and do not condition on referral to the principal's office, we hypothesize that:

> *[Hypothesis 2:] Race differences in children's behaviors account for a much smaller share of the racial gap in suspension/expulsion at age 9 than the between-school sorting explanation.*

Differences in the Treatment/Support of Black Students

Finally, the *differential treatment/support* perspective argues that Black students are more likely to be suspended than White students even when they enter school with the same behavior. Differences in punishment may be due to racial bias on the part of teachers and school officials (Okonofua and Eberhardt 2015), race differences in students' access to settings and resources that promote social emotional learning and school engagement (Lewis and Diamond 2015; Weinstein 2002), and differences in parents' ability to advocate for their child (Lareau and Horvat 1999). These factors can themselves instigate a "self-fulfilling prophecy" or a "vicious cycle" that leads to worsening behaviors (Okonofua, Walton and Eberhardt 2016; Weinstein 2002).

The strongest evidence for *differential treatment/support* 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), from school administrative records examining length of suspensions assigned to Black versus White students who fought with each other (Barrett et al. 2017), and from ethnographic studies based on close observation of teacher-student interactions in classrooms (Carter 2005; Ferguson 2001). To study discrimination in lab experiments, Okonofua and Eberhardt (2015) presented teachers with short, written vignettes about student behaviors and ask 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 students as compared with White students has also been tested outside the laboratory. Consistent with the notion of "tough love" or "protective parenting," some research finds that advocates of students of color may nonetheless evaluate or punish students of color more harshly in preparation for the realities of an unjust society (Farkas et al. 1990; Gilliam et al. 2016; Howard, Rose and Barbarin 2013). By contrast, Lindsay and Hart (2017) find that White teachers are 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 as compared to White teachers (Egalite and Kisida 2017).

Ethnographic research and research using administrative records find that, 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). Importantly, it is precisely with these relatively minor forms of misbehavior that discretion over whether to refer or recommend for punishment is greatest (Gregory and Weinstein 2008; Smolkowski et al. 2016). 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. Building on findings described above, we hypothesize that:

> *[Hypothesis 3:] Race differences in the disciplinary treatment of Black students with the same behaviors at school entry, the same family socioeconomic resources, and the same school contexts explain more of the racial gap in suspension/expulsion at age 9 than the behavior differences explanation and at least as much as the between-school sorting explanation.*

Our study extends prior work in several ways. First, expanding on insights from prior work (e.g., Barrett et al. 2017), we parse the relative contributions of the three mechanisms described above. 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. Second, we examine the likelihood of suspension and expulsion *unconditional* on having been referred and suspended, and we measure children's behaviors at the time they enter school, helping increase the chances that children have not previously been referred to the principal's office or otherwise disciplined given low rates of preschool and kindergarten discipline (Government Accountability Office 2018). Third, our study utilizes both teacher and parent reports of child behaviors, providing a more comprehensive picture of children's behaviors (Ferguson 2001; Gilliam et al. 2016). Finally, to address the possibility that stable but unobserved differences in the characteristics of schools may bias our estimates, we also examine race differences in suspension within the exact same school.

Data and Methods

Data and Sample

Our data come from the FFCWS, a stratified, multistage probability sample of 4,898 children born between 1998 and 2000 in 20 U.S. cities with populations equal to or greater than 200,000 and followed prospectively and longitudinally from birth. These data include a large oversample of children born to unmarried parents (around 3 to 1), resulting in a disproportionately large number of children from low-income families. Baseline interviews were conducted with mothers and most fathers in the hospital shortly after the child's birth. 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 1, 3, 4, 9, and 15 years old. Teachers were also interviewed about students' behaviors and achievement, and school administrative records were collected during the age 5 and age 9 interviews. At age 9, children were interviewed about their school and home experiences, including whether they had ever been suspended or expelled. The age 9 sample includes 3,515 children, about 72 percent of the FFCWS sample at baseline, attending 1,729 public and 163 private schools located fairly evenly across the U.S. (roughly 1/3 each in the Northeast and South, and the remaining 1/3 split between the West and Midwest).

The FFCWS is well suited for our study of the drivers of racial disparities in elementary school suspension for several reasons. First, 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 8th grade, if at all. Second, the FFCWS data contain a 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 begin at birth and provide detailed information on children's families, early-care environments, socioeconomic status, and family structure since birth. Fourth, behaviors are reported *prospectively*, in most cases prior to any suspension or negative labeling of children, and they are reported by *both* teachers and parents, helping compensate for stereotype bias that might arise from teacher reports alone (Okonofua and Eberhardt 2015). Finally, the data include information on children's schools at the start of elementary school, *and* allow us to test whether differential treatment/support persists *within the exact same school*. School NCES ID numbers are used to match children to the 2004-06 administrative records from the Common Core of Data (CCD)/Private School Universe Survey (PSS).

Our primary outcome variable—child-report of ever being suspended or expelled by age 9—was missing for 176 children, or 5 percent of the sample. These cases were excluded from all analyses, yielding a sample of 3,339 children. Additionally, on most predictor variables, item-missingness does not exceed about 19 percent (predictors are measured between birth and school entry, or

waves 1–4). The exception is teacher ratings of child behavior at age 5, where item-missingness approaches 70 percent. To address item-missingness on predictors, we used multiple imputation of 20 datasets based on the MI suite in Stata 14. Following Von Hippel (2007), we included the outcome variable in the imputation equation but dropped children with imputed y-values from the analysis. Complete case analysis revealed a similar pattern of results, indicating that patterns of item-missingness do not alter substantive conclusions, including for teacher reports of behavior at age 5 (see Appendix Table A.1). Finally, we exclude 943 Latinx, Asian, and Native American/Pacific Islander children, yielding an analytic sample of 2,396 Black and White children.

Measures

Elementary School Suspension/Expulsion

The dependent variable is a binary indicator from the child's "age 9" response to the question: "Have you ever been suspended or expelled from school?" Because only 0.1 percent of elementary school children have been expelled (Government Accountability Office 2018, Table 16), this measure largely captures suspension. A comparison to rates of ever being suspended in-school and out-of-school among K-12th graders during the 2011-12 academic year (20 percent for Black students versus 9 percent for White students) suggests that our measure includes both in-school and out-of-school suspensions (Office for Civil Rights 2014). For this reason, we refer to our outcome as "suspension" for short.

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 percent of children classified as Black, the biological father also identified as "Black, non-Hispanic." For 76 percent of children classified as White, the biological father also identified as "White, non-Hispanic." To determine if children's multi-racial backgrounds may have shaped identification by school officials in discipline-related ways, we conducted sensitivity analyses excluding all multiracial children. Results remained unchanged and are available upon request.

School Racial/Ethnic and Socioeconomic Composition

School composition involves continuous measures of "percentage of students who are Black or Hispanic" and "percentage of students receiving free-or-reduced-price lunch" (FRPL) at the start of elementary school. Appendix Tables A.2-A.3 show that results are robust to possible non-linearities captured through a series of dummy variables and to changes in school composition.

Child Behavior Problems (Parent and Teacher Reports of Externalizing Problems)

Behavior problems consist of teacher and parent reports of children's externalizing behavior problems at age 5 using items from the Child Behavior Checklist

(CBCL) (Achenbach 1991), which strongly predict child suspension/expulsion (Bradshaw et al. 2010; Raffaele-Mendez 2003). Given that it is difficult to disentangle the extent to which low correlations across reporters ($r = 0.28$ in our sample at age 5) reflect differences in behavior across contexts versus differences in perceptions across raters, we follow Achenbach, McConaughy and Howell (1987), Achenbach (1991), and Verhulst, Koot and Van der Ende (1994) in averaging teacher and parent reports. Also, since teacher reports may reflect how teachers interpret and assign meaning to the behaviors of minority students, averaging helps to capture a more holistic portrait of children's behavior (Ferguson 2001; Gilliam et al. 2016; Vavrus and Cole 2002) and prevents overreliance on teacher reports given high item-missingness. Whereas behavior reports based on a single infraction introduce issues of non-random selection into referral, potentially producing biased model estimates, CBCL scales are not conditional on suspension. Age-appropriate items draw from four sub-scales: *social problems*, *attention problems*, *aggression*, and *rule-breaking* (Achenbach 1991). Each item ranges from 0 "not true/never", 1 "somewhat/sometimes true", to 2 "very often or often true." Items are reverse-coded as necessary so higher scores indicate greater problems. There are 36 items in the "age 5" parent-reported scale (alpha = 0.80) and 34 items in the "age 5" teacher-reported scale (alpha = 0.92). Appendix Table A.7 reports results using "age 9" behaviors (r = 0.39 across raters) to address the possibility that behaviors may have changed between age 5 and any suspension by age 9 (r = 0.53 across ages 5 and 9). There are 54 items in the age 9 parent-reported scale (alpha = 0.93) and 51 items in the age 9 teacher-reported scale (alpha = 0.94). Finally, given that some children may have had prior infraction histories before "age 5" behaviors were rated but such children cannot be identified/excluded, we were concerned about potential reverse causality. However, supplemental analyses of the CRDC data indicate that roughly 3 percent of K-5 suspended/expelled students nationwide are suspended/expelled in kindergarten, suggesting that the inclusion of such children is unlikely to drive our results.

Controls

Analyses also adjust for a number of child and family factors shown to be correlated with child race and school suspension. *Parents' Socioeconomic Resources.* Mother-reported household income-to-poverty ratio at age 5, mother's education at the 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 mother-reported biological father absence across survey waves 1-4 (birth to age 5). Father-absent families include stable single mother families and families where the mother repartnered or remarried at least once (i.e., stepparent families). Eight children lived in father-only headed or foster parent households (without the mother). Collinearity issues prohibited the inclusion of dummy variables to capture these

rare family types, but results do not differ, so they are retained in the sample. *Paternal Incarceration.* A dummy variable equal to 1 if the child's biological father has ever been in jail or prison at any wave up to age 5, causally prior to any suspension (based on mother or father report). *Other Child Characteristics.* Child Peabody Picture Vocabulary Test (PPVT) score at age 5 (to account for differences in academic skills at school entry) and the child's age in months at age 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). However, results do not change if we use child age at age 5, causally prior to any suspension.

Analytic Approach

We begin by employing a two-stage Oaxaca-Blinder decomposition that examines 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 children with the same behaviors who attend schools with the same racial and socioeconomic compositions (hypothesis 3) (Jann 2008). The decomposition models a counterfactual scenario that displays how large the racial gap in suspension would be for hypothetical Black and White children 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/support") (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 children 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 children had the same mean levels of exposure to each factor, but the association for each factor varied by race, versus if Black and White children had the same coefficients for each factor but different levels of exposure, as shown in Equation (1):

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

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." The term "effects" should not be interpreted causally. In addition to the possibility of omitted variables, "effects" in the decomposition framework refers to differences in slopes or coefficients between Black and White children, under the counterfactual in which they experience the same levels of exposure to a given factor.[1]

Importantly, the Oaxaca-Blinder decomposition allows us to *simultaneously* estimate the contributions to the racial gap in suspension that arise from racial differences in *levels of* behaviors, school composition, and family and child background factors, as opposed to racial differences in the "effect" of behavior (e.g., including differential treatment/support of children with 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/support" that could arise if not all predictors were interacted with "Black" (i.e., if other predictors were not also allowed to vary by race) (VanderWeele and Knol 2011). 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. Since 856 of the 2,396 children in the analytic sample attend one of 352 schools enrolling other sample children, standard errors are clustered at the school level.

To examine whether there is evidence of differential treatment/support of Black relative to White children *within the exact schools* (hypothesis 3), we estimate two types of linear probability models (LPMs). First, we establish baseline estimates of differential treatment/support in schools with similar racial and SES compositions. We regress school suspension by age 9 (our outcome) on race (an indicator for "Black"), the average of teacher- and parent-rated behavior at school entry, the interaction between race and behaviors, main effects for school composition at the start of elementary school, family socioeconomic status, and child characteristics, and interactions of each with "Black" (to guard against biased estimates of the focal behavior*Black interaction due to unobserved interactions, per VanderWeele and Knol (2011)). Second, we add controls for school fixed effects, which allow us to test for differential treatment/support of children attending the same school. This model controls for all stable observed and unobserved differences between schools.

Importantly, results are robust to the use of a non-linear expansion to the Oaxaca-Blinder method, which avoids the assumptions of linearity imposed by the LPMs and Oaxaca-Blinder decomposition (see Appendix Table A.4). We also examine and find that the drivers of the racial gap operate similarly within gender groups, except that there is some evidence of lesser differential treatment/support toward Black relative to White girls who have similar behaviors at school entry (see Appendix Tables A.9-A.10 and Appendix Figure A.2).

Results

Descriptive Statistics

Table 1 reports means and standard deviations (or proportions) for each of the variables in our analyses. Black children are four times more likely than White children to report having ever been suspended or expelled by age 9 (28 percent

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

	Black ($N = 1,696$)		White ($N = 700$)				
	Mean	SD	Mean	SD	Min	Max	Sig. Diff.
Outcome							
Child Ever Suspended/Expelled Between Ages 5 and 9	0.280		0.070		0	1	***
Child Behavior Problems							
Average of Teacher- and Parent-Reported Externalizing Problems Score, Age 5	11.200	6.613	10.091	6.769	0	40	***
School Racial and Socioeconomic Composition Factors							
Proportion of School Enrollment Black or Hispanic at Start of Elementary School	0.801	0.259	0.317	0.277	0	1[a]	***
Proportion of School Enrollment Free-or-Reduced-Price Lunch (FRPL) at Start of Elementary School	0.696	0.244	0.386	0.274	0	1[a]	***
Family Socioeconomic and Stability Factors							
Family Income-to-Poverty Ratio, Age 5	1.456	1.476	3.162	3.029	0	28	***
Mother Has Some College or College Degree, Age 1	0.305		0.556		0	1	***
Father Absent From Household at Any Wave, Age 5	0.832		0.476		0	1	***
Father has Ever Been in Jail or Prison, Age 5	0.550		0.354		0	1	***
Mother's Age, Age 1	24.287	5.624	26.931	6.449	15	43	***

(*Continued*)

Table 1. ***continued***

	Black ($N = 1{,}696$)		White ($N = 700$)				
	Mean	SD	Mean	SD	Min	Max	Sig. Diff.
Other Child Factors							
Child's PPVT Cognitive Score, Age 5	91.260	14.590	103.101	14.689	38	140	***
Child's Sex (Male = 1), Age 1	0.519		0.526		0	1	
Child's Age (in Months), Age 9	112.274	4.563	111.731	3.781	100	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.
[a]School composition variables are continuous within the 0 to 1 range (i.e., they are not binary/categorical 0 or 1 variables).
Source: Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 2,396 Black and White boys and girls 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).

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

		Means		Difference in Means	OLS Regression Coefficients				Contrib. of Differences in Levels of Exposure	Contrib. of Differences in "Effects"/ Slopes
		(1)	(2)	(3)	(4)		(5)		(6)[b]	(7)[c]
Factors	Predictor	$\bar{x}_W$	$\bar{x}_B$	$(\bar{x}_W-\bar{x}_B)$	β_W	Sig	β_B	Sig	$(\bar{x}_W-\bar{x}_B)\beta_B$	$(\beta_W-\beta_B)\bar{x}_B$
School Factors (H1)	Proportion Black or Hispanic Students at Start of Elementary School	0.317	0.801	−0.484	0.127	**	0.023		**0.011**	−0.033
	Proportion Free-or-Reduced-Price Lunch (FRPL) Students at Start of Elementary School	0.386	0.696	−0.310	−0.036		0.101		**0.031**	0.053
Behavior Factors (H2-H3)	Average of Teacher- and Parent-Reported Externalizing Problems Score, Age 5	10.091	11.200	−1.108	0.007	***	0.017	***	**0.018**	**0.095**
Controls	Family Income-to-Poverty Ratio, Age 5	3.162	1.456	1.707	−0.006	**	−0.011		0.019	−0.017
	Mother Has Some College or College Degree, Age 1	0.556	0.305	0.251	−0.029		−0.039		0.010	−0.005
	Father Absent from Household at Any Wave, Age 5	0.476	0.832	−0.356	0.009		0.032		0.011	0.011

	Father has Ever Been in Jail or Prison, Age 5	0.354	0.550	−0.196	0.027		0.021		0.004	−0.002
	Child's PPVT Cognitive Score, Age 5	103.101	91.260	11.841	0.001	*	0.002	**	−0.024	0.080
	Child's Sex (Male = 1), Age 1	0.526	0.519	0.007	0.053	**	0.168	***	−0.001	0.060
	Child's Age (in Months), Age 9	111.731	112.274	−0.543	0.000		−0.001		−0.001	−0.143
	Mother's Age, Age 1	26.931	24.287	2.644	−0.002		−0.001		0.002	0.025
	Constant	1.000	1.000	0.000	−0.143		−0.140		0.000	0.003
	Observations (N)	700	1696		700		1696			
Overall Contribution of to the Racial Gap of Differences in Levels vs Slopes in Percentage-Point Units (/100):									0.081	0.127
Proportion of the Overall Race Gap Driven by Differences in Levels vs. Effects/Slopes:									0.389	0.611

***p < 0.001, **p < 0.01, *p < 0.05, +p < 0.10 (two-tailed t-tests for a statistically significant difference from 0). Controls and the constant are included in the decomposition but not shown (see Appendix for complete decomposition table).

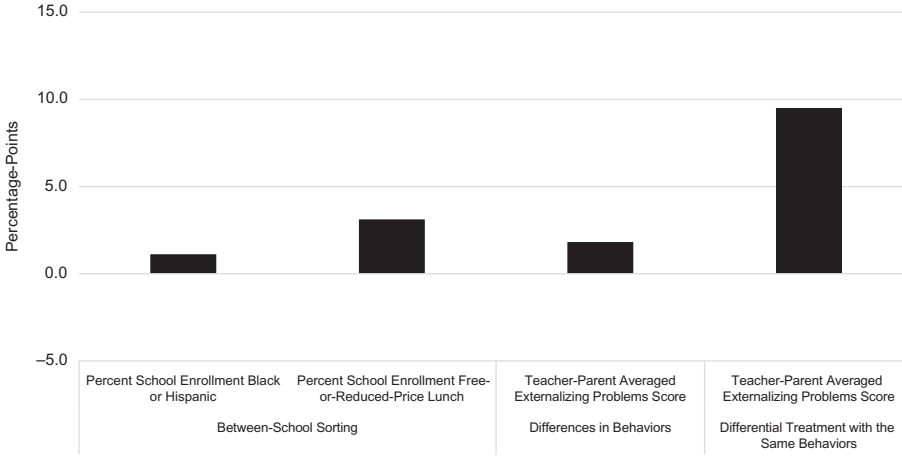
[a]This model uses Black children's coefficients as the reference when calculating each variable's contribution to the gap in schooling due to racial differences in mean levels and Black children's means as the reference when calculating each variable's contribution due to racial differences in coefficients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

[b]Values in Column (6) are multiplied by −1 (to achieve positive values for gap-widening contributions and vice versa for gap-narrowing contributions).

[c]Values in Column (7) are multiplied by −1 (to achieve positive values for gap-widening contributions and negative values for gap-narrowing contributions).

Source: Fragile Families and Child Well-Being Study, Waves 1-5. Sample is restricted to the 2,396 Black and White boys and girls 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).

Figure 1. Estimate of the Contributions of Between-School Sorting, Behavior Differences, and Differential Treatment/Support of Children who Entered School with Comparable Behaviors to the Racial Gap in Suspension/Expulsion: Behavior Ratings Only at Age 5.



of Black children as compared to 7 percent of White children). By gender, 37 percent of Black boys report ever having been suspended or expelled by age 9 compared to 10 percent of White boys, 17 percent of Black girls, and 4 percent of White girls.

At the start of elementary school, the average Black child in our sample attends a school in which 80 percent of students are Black or Latino/Hispanic and 70 percent receive free-or-reduced-price lunch. The numbers for White children are 32 percent and 39 percent, respectively. Differences in school composition are further magnified when we look at the percent of students attending schools that are both minority and poor: 52 percent of Black children compared to only 7 percent of White children.

Black children score higher than White children on behavior problems at age 5, based on the average of teacher and parent ratings (a roughly 1.1-point or 0.15 SD difference). 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 (with the exception of child's age).

Decomposition Analyses

How much of the racial gap in suspension is driven by racial differences in school composition (hypothesis 1), student behaviors (hypothesis 2), and differential treatment/support for students who enter school with the same behaviors (hypothesis 3)? Table 2 and Figure 1 display results from our decomposition analysis. Columns 1 and 2 of Table 2 display mean levels of children's early school and family exposures and behaviors at school entry by race. Column 3 displays White-Black differences in these means. Race-specific slopes/"effects"

associated with observed factors 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/coefficients* (i.e., differences in treatment/support by race for the same exposures). Together, the proportions in Columns 6 and 7 can be multiplied by 100 for interpretation in percentage-point units, which sum to 0.208, or the nearly 21 percentage-point gap in suspension. Positive values in Columns 6 and 7 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/coefficients of each of the other variables in the model.

> ***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 age 9.*

The first 5 rows of Table 2 highlight the school composition and behaviors of theoretical interest. According to our estimates, racial differences in the composition of the schools that Black and White children attend account for 4.4 percentage-points (or 21.2 percent) of the roughly 21 percentage-point gap in suspension [(0.031+0.011)/0.208 = 0.212*100 = 21.2 percent]. The fact that a larger share of Black children attend schools with high percentages of minority students accounts for 1.1 percentage-points (5.3 percent) of the gap, while the fact that a larger share of Black students come from lower-income families accounts for 3.1 percentage-points (14.9 percent) of the 21 percentage-point gap.

The first three bars of Figure 1 illustrate these findings. Consistent with prior work, we find that between school sorting accounts for a notable portion of the Black/White gap in suspension. Appendix Table A.2 indicates that the sorting of Black children into schools that serve students from *both* low-income and minority backgrounds accounts for the overwhelming majority (3.3 percentage-points) of the total 4.4 percentage-point contribution of between-school sorting. Mediation analysis (displayed in Appendix Table A.5) shows that these schools have high rates of punitive discipline.

Insofar as children may change schools and schools may change composition during our observation period (age 5 to age 9), our school composition variable may be measured with error, resulting in an underestimate of the role of school composition in accounting for racial differences in suspension. To address this possibility, we restricted our sample to children who did not change schools after 1st grade and used school composition measures from "age 9." Results are nearly identical to those in Table 2 (see Appendix Tables A.3 and A.6).

> ***Hypothesis #2.*** *Race differences in children's behavior problems account for a much smaller share of the racial gap in suspension/expulsion at age 9 than the between-school sorting explanation.*

According to our estimates in Table 2, differences in teachers' and parents' averaged reports of children's behavior at age 5 account for 1.8 percentage points (8.7 percent) of the racial gap in school suspension (0.018/0.208 = 0.087),

ceteris paribus. The second panel of Figure 1 graphically illustrates these findings. These results are in line with hypothesis 2, which suggests that race differences in behaviors at school entry contribute a relatively small share to the racial gap in suspension.

To the extent that Black children's behaviors worsen more than White children's after school entry and before any suspension (for reasons discussed in the Appendix), our use of the "age 5" measure would lead to an underestimate of the role of race differences in behavior and an overestimate of the role of differential treatment/support. To take account of changes in children's behaviors after they enter school, we re-estimated our decomposition model and included averaged teacher and parent reports of children's behavior at age 9 (Appendix Table A.7 and Appendix Figure A.1). Resulting estimates of the role of between school sorting (hypothesis 1) decreased only slightly, accounting for 4.3 percentage-points (20.7 percent) of the 21 percentage-point race gap in suspension, whereas estimates of differences in behavior (hypothesis 2) increased substantially, accounting for 6.3 percentage-points (23.5 percent) of the racial gap. Note that age 9 behavior is measured contemporaneously with suspension.

> ***Hypothesis #3.*** *Race differences in the disciplinary treatment of Black students with the same behaviors at school entry, the same family socioeconomic resources and the same school contexts explain more than the behavior differences explanation and at least as much of the racial gap in suspension/expulsion at age 9 as the between-school sorting explanation.*

Decomposition results for our third hypothesis, which focuses on differences in slopes/coefficients rather than differences in levels, are presented in column 7 of Table 2 and panel 3 of Figure 1. According to our estimates, 9.5 percentage points (45.7 percent) of the racial gap in suspension can be attributed to the differential treatment/support of Black and White children who enter school with the same behaviors, holding constant all children's means on the other variables in the model at the levels observed for Black children on average (0.095/0.208 = 0.457*100 = 45.7 percent). This contribution remains statistically significant at a more conservative p-value < 0.016 (two-sided test) with a Bonferroni correction for our testing of three hypotheses. Altogether, race differences in between-school sorting (hypothesis 1), in children's behaviors at age 5 (hypothesis 2), and in the treatment of Black children who enter school with the same behaviors and attend similar schools as their White counterparts (hypothesis 3) account for nearly three quarters of the roughly 21 percentage-point racial gap in suspension at age 9. An identical decomposition using White (rather than Black) children's means and coefficients as the reference group are shown in Appendix Table A.8; substantive patterns of results remain consistent, but the contributions of between school sorting and differential treatment/support to the racial gap both increase.

Linear Probability Models

Finally, while the above decomposition results lend support to the differential treatment/support hypothesis for Black and White children in similar schools,

Table 3. LPM Predicting Child Ever Suspended/Expelled, Age 9: Estimates of "Differential Treatment/Support" by Race for the Same Behavior Problems at School Entry, Additionally Including Main Effects and Interactions Between Race and Family and Child Factors

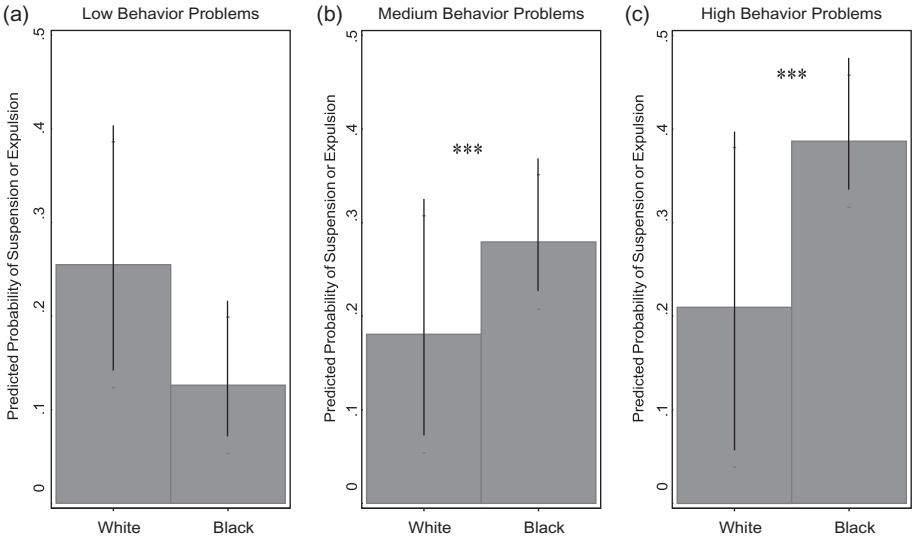
	Full Working Sample (1)	Sub-Sample with Black and White School Clustering (2)	Sub-Sample with Black and White School Clustering (3)
Non-Hispanic Black	0.062 (0.040)	0.134+ (0.081)	0.094 (0.112)
Average of Teacher and Parent Ratings of Child Externalizing Behavior Age 5	0.007*** (0.002)	0.002 (0.005)	−0.000 (0.007)
Black*Average of Teacher and Parent Ratings of Child Externalizing Behavior Age 5	**0.009***** **(0.003)**	**0.014*** **(0.006)**	**0.015*** **(0.007)**
Percent of School Enrollment Black or Hispanic, Start of Elementary	0.127** (0.049)	0.096 (0.133)	
Percent of School Enrollment FRPL, Start of Elementary	−0.036 (0.045)	−0.052 (0.110)	
Black*Percent of School Enrollment Black or Hispanic, Start of Elementary	−0.104 (0.068)	−0.078 (0.160)	
Black*Percent of School Enrollment FRPL, Start of Elementary	0.137+ (0.071)	0.143 (0.144)	
Family Socioeconomic Status and Child Factors (Main Effects) and Interactions with Indicator for Black	X	X	X
School Fixed Effects			X
Observations	2,396	828	828
Adjusted R-squared	0.173	0.158	0.548

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

Notes: Models additionally control for all the family, socioeconomic status, and other child controls shown in Table 1 as well as interactions between each and the indicator for Black.

Source: Fragile Families and Child Well-Being Study, Waves 1-5. Model 1 is restricted to the 2,396 Black and White boys and girls who remained in the study from birth (wave 1) through age 9 (wave 5). Models 2-4 are restricted to the 828 children nested within 354 schools with at least one other FFCWS sample Black or White child. Multiple imputation of 20 datasets is used to handle item-missingness on all but the dependent variable (suspension/expulsion).

Figure 2. Differential Treatment/Support of Black and White Children Who Entered School with Low, Medium, or High Levels of Averaged Teacher and Parent Behavior Problems at Age 5 (Outcome: Predicted Probabilities of Suspension or Expulsion by Age 9).



Notes: Statistically significant differences in pr(suspension/expulsion) between Black and White children at: *p < 0.05, **p < 0.01, ***p < 0.001 (two-sided t-test). Displaying 95 percent confidence intervals around predicted probabilities. School fixed-effects model (N = 828). Model controls for child PPVT, family characteristics and demographics shown in Table 1 as well as interactions of each with child race.

LPMs in Table 3 consider whether there is evidence of differential treatment/support among the roughly 35 percent (828/2,396) of Black and White children in the *same school* as one other sample child of the opposite race (these 828 children are spread across roughly 350 schools). All three models include main effects for race, averaged teacher- and parent-rated student behavior problems, and their interaction, holding constant family and child characteristics and their interactions with race (to avoid a biased estimate of the focal race*behaviors interaction). Without school fixed effects, Models 1-2 additionally include main effects for school racial and socioeconomic composition plus interactions with child race in order to compare Black and White children in similar schools. By way of direct comparison to estimates of the contributions of school composition used in the decomposition analysis (Table 2), Model 1 shows that each unit increase in behavior problems at age 5 is associated with a 0.7 percentage-point increase in suspension among White children and a 1.6 percentage-point increase among Black children [(0.007+0.009)*100] in schools with similar racial and socioeconomic compositions, *ceteris paribus*. This 0.9 percentage-point racial difference is statistically significant.

Because the sub-sample of children with Black-White clustering tend to be poorer and enroll a larger percentage of Black and Hispanic students compared to the overall sample, Model 2 is the same as Model 1 but restricted to the sub-

sample with both Black and White sample children attending the same school. Model 2 shows that each unit increase in behavior problems at school entry is associated with a statistically significant 1.6 percentage-point increase in the probability of suspension among Black children, but a non-statistically significant 0.2 percentage-point increase among White children. The significant coefficient on the interaction term between "Black*behaviors" indicates that the difference in these percentages is statistically significant.

Model 3 displays *within-school* (i.e., fixed effect) estimates for the same subsample with Black-White school clustering, thus controlling for differences in all time-invariant characteristics of the schools attended by Black and White sample students. In these schools, each unit increase in behavior problems is associated with a 1.5 percentage-point increase in suspension among Black children but a 0.0 percentage-point increase among White children, *ceteris paribus*. Compared to the 0.9 percentage-point Black-White difference from Model 1, this 1.5 percentage-point difference suggests that the relative contribution of differential treatment/support may be even larger when comparing Black and White children in the same school.

Figure 2 illustrates the magnitude of the resulting racial gaps in suspension at low, medium, and high levels of behavior problems at school entry, based on tertiles of the averaged teacher and parent ratings of "age 5" externalizing problems. Results indicate that racial differences in the predicted probabilities of suspension are statistically significant only between Black and White children whose comparable school entry behavior problems were in the middle or top (but not bottom) tertiles of the "age 5" externalizing problems distribution. In the middle tertile, Black children are a highly statistically significant 10 percentage-points more likely to be suspended than their White counterparts (predicted probabilities of 29 percent versus 19 percent, respectively). In the top tertile, Black children are a highly statistically significant 19 percentage-points more likely to be suspended (40 percent versus 21 percent, respectively). By contrast, in the bottom tertile, Whites appear to be 12 percentage-points more likely to be suspended, but this difference is not statistically significantly different from 0 (12 percent for Blacks versus 25 percent for Whites).

Robustness Checks

In our sample, Black children's behaviors worsen more than White children's behaviors between ages 5 and 9 (21.6-11.2 = 10.4 points for Black children versus 17.9-10.1 = 7.8 points, for White children). The racial gap in behaviors thus grows from roughly 0.15 SD to 0.32 SD between ages 5 and 9. One explanation for this trend might be that Black children are more likely than White children to experience economic hardship, family structure instability, and/or neighborhood violence, all of which are likely to increase children's behavior problems. Alternatively, the disproportionate worsening of Black children's behaviors may be due to differences in exposure to negative school environments. If Black children are more likely than White children to be suspended, and if suspension leads to an increase in behavior problems as prior research suggests (Jacobsen, Pace and

Ramirez 2018; Okonofua and Eberhardt 2015), we would expect the racial gap in behaviors to increase over time. Finally, between school entry and 4th grade, Black students may be more likely to garner negative reputations in the eyes of teachers, which could account for some of their more negative behavior ratings and greater suspension as they progress through elementary school (Ferguson 2001).

To take account of changes in children's behaviors after they enter school, we re-estimated our decomposition model and included averaged teacher and parent reports of children's behavior at age 9 (see Appendix Table A.7 and Appendix Figure A.1). As expected, estimates of the role of between school sorting (hypothesis 1) remain quite stable (contributing 20 percent) with the inclusion of "age 9" behaviors and the role of behavior differences (hypothesis 2) increases substantially (contributing 24 percent versus 9 percent). Surprisingly, however, the role of differential treatment/support also increases (contributing to 70 percent versus 46 percent of the gap). Thus, 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 children are treated/supported differently (and more harshly) than White children, net of behavior at school entry.

Variation by Gender

Research on intersectionality suggests that the mechanisms described above may operate differently for Black boys and girls. Consistent with prior research, the racial gap in suspension is 50 percent smaller among girls than boys (13.8 percentage-points versus 27.2 percentage-points). In decomposition models stratified by child gender (see Appendix Figure A.2), results for the contributions of differences in school composition (roughly 20 percent) and behaviors (roughly 10 percent) are similar to those in the pooled models. However, whereas for boys, 55 percent of the racial gap is associated with differential treatment/support, for girls the number is only 30 percent. Future research is needed to examine other factors that may help account for suspension disparities between Black and White girls.

Discussion

In this study, we test three hypotheses that have been shown to account for some of the Black-White gap in school suspension and expulsion: between-school sorting, differences in students' behavior, and differential treatment/support of Black and White students who enter school with comparable 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. Using data on a cohort of children born in large U.S. cities at the turn of the 21st century and attending elementary school between 2003 and

2009, we document a roughly 21 percentage-point racial gap in suspension – 28 percent of Black children versus 7 percent of White children—by the 4th grade.

Second, our results lend strong support to the differential treatment/support hypothesis using a national sample of children born in large US cities between 1998 and 2000 and attending elementary school between 2003 and 2009. We find that, *even within the exact same school*, each unit increase in kindergarten behavior problems is, on average, associated with a 1.5 percentage-point larger increase in suspension among Black than White children, holding constant all time-invariant characteristics of schools as well as other child and family characteristics. Although 1.5 percentage-points may appear small at first glance, each unit increase in behavior problems amounts to over 7 percent of the 21 percentage-point Black/White gap in suspension. Importantly, this differential treatment/support is concentrated among children who enter school with middle or high (as opposed to low) levels of behavior problems. Because children with bottom tertile (i.e., low) externalizing problems are considered to be within the "normal" or even "positive" behavioral range (Brame, Nagin and Tremblay 2001), this finding suggests that well-behaved Black and White children are not differentially treated/supported.

Third, we examine the differential treatment/support hypothesis within a broader context that includes other prominent explanations for the racial gap in suspension. Consistent with prior work on between-school sorting, we find that differences in school racial and socioeconomic composition account for a large share (21 percent) of the racial gap in suspension during elementary school (Welch and Payne 2010). We also help clarify prior results by showing that behavior differences account for a relatively small share (9 percent) of the gap when behavior is measured at school entry and prior to suspension. Finally, we find that a large portion (46 percent) of the racial gap is due to the differential treatment/support of Black and White children who attend similar schools and who exhibit similar behaviors at the time they enter school (Ferguson 2001; Rocque 2010; Skiba and Williams 2014).

Our findings are consistent with prior research on interpersonal racial discrimination (Pager and Shepherd 2008). In elementary school in particular, when misbehavior tends to be relatively common and relatively minor, educators exercise high levels of discretion in determining sanctions for inappropriate behavior (Gregory and Weinstein 2008; Smolkowski et al. 2016). 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. Broadly, this finding sheds light on how discrimination can be mutually reinforcing in the context of early racial disparities in school discipline, with cumulative implications for racial inequality.

We should note that a constellation of factors other than interpersonal bias may be subsumed within our finding of "differential treatment/support." For example, relative to White students, Black students may have less access to resources like supportive adults, social emotional learning opportunities, and rigorous and engaging instruction. These factors likely contribute to what

Weinstein (2002) refers to as a "self-fulfilling prophecy" and Okonofua, Walton, and Eberhardt's (2016) call a "vicious cycle." In both senses, these artifacts of the less-resourced environments typical for many minority children may produce increases in behavior problems between the start of schooling and the time of a later suspension (which is consistent with our findings using age 9 behaviors, detailed in the Appendix). As such, it would be overly simplistic to say that policy efforts should focus on any single mechanism.

Fourth, our analyses show how variation in the nature of the sample and the time at which child behaviors are measured can dramatically change results. Our data allow us to assess children's behaviors at two points—when children enter kindergarten and again in 4th grade. Using reports from kindergarten entry, we find that behavior differences play a larger role than what is found in studies that condition on having been referred for punishment. If Black children are more likely than White children to be referred for minor misbehaviors that are less likely to warrant suspension, 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 in studies that measure behaviors later in a child's school career. This happens because Black students are more likely than White students to be suspended and because suspension is likely to negatively affect future behaviors (Dance 2002; Jacobsen, Pace and Ramirez 2018; 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 school, including both stereotyping and harassment or discrimination by school officials (Dance 2002; Ferguson 2001; Rios 2011).

Our study also has limitations. First, the decomposition approach and the use of observational data do not allow us to make 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. Omitted variables bias may result from our inability to link suspension to a particular infraction and thus to determine whether Black and White children are differentially suspended for the same infraction. Related, because students typically have a prior history of referrals to the principal's office or detentions before they are suspended, even kindergarten behavior reports may be endogenous to exposure to these earlier stages in the discipline pipeline. For example, because Black children are more likely to be suspended beginning in preschool, even by age 5, parent and teacher reports of behavior may reflect a transactional process of behavior reinforcement.

Second, our outcome variable cannot distinguish between in-school suspension, out-of-school suspension, and expulsion. However, because estimates by the Government Accountability Office (2018) suggest that only 0.1 percent of children are suspended in elementary school, our measure is likely to be largely capturing in-school and out-of-school suspensions and, in particular, early suspensions. Third, we cannot measure the count of suspensions; repeat suspensions are possible. Our supplementary analyses using "age 9" measures of behavior help compensate for this omission by capturing changes in teacher and parent impressions of a child's behavior following the specific infraction that led a child to be suspended. But, ideally, we would have measures of both overall behaviors and specific infractions. This would allow us to determine if Black children are referred to the principal for more minor infractions.

Fourth, our analysis is based on a sample of children born in large cities and therefore our results may not generalize to children born in suburban and rural settings. Additionally, only 35 percent of our sample includes Black and White children in the same school, and thus our fixed effects estimates generalize only to children in the subset of schools with Black/White clustering. These schools tend to be more disadvantaged than the average school; of the 350 schools with Black/White clustering among sample children, the average school is composed of 77 percent Black or Hispanic students versus 66 percent in the full sample and 67 percent of students receiving free-or-reduced-price lunch versus 61 percent in the full sample. Part of this limitation is a design feature of the data, but a larger portion is a structural reality, reflecting the fact that Black and White children attend very different schools. Because schools that are more disadvantaged also tend to be more punitive toward *all* students (the high correlation between school composition and schools' overall rates of suspension can be seen in Appendix Table A.5), we suspect that our within-school model estimates may in fact be an underestimate of the magnitude of differential treatment/support that exists in schools that are less punitive overall. For example, prior research in social psychology suggests that racial bias is largest in environments where actors have significant discretion in decision-making (Gregory and Weinstein 2008; Smolkowski et al. 2016). In disadvantaged schools where suspension rates are high, on average, teachers may have fewer non-punitive disciplinary options than in more advantaged schools with lower average rates of suspension. Future research should investigate this possibility with respect to punitive school discipline.

Fifth, we cannot disentangle the extent to which differences between parent and teacher reports of behavior reflect real variation in behaviors across contexts (school vs. home) as opposed to differences in the ways identical behaviors are perceived by teachers and parents. Future research would benefit from the use of experimental techniques to investigate this question. Sixth, and related, even among Black and White children whose behaviors are perceived similarly, differential treatment/support 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 children for the challenges they are likely to encounter in the wider society (Gilliam et al. 2016; Howard, Rose and Barbarin 2013). Lacking

knowledge of the specific teacher referring a student to the principal (and by extension, knowledge of the teacher's race/ethnicity), we are unable to examine this question. Future work should investigate how racial similarity/dissimilarity between teachers and students shapes suspension rates for Blacks compared to Whites. Finally, our analysis does not include Latinos, because the Latino-White gap in our sample is minimal – 2 percentage points. However, future research calls for careful sub-group analyses given tremendous ethnic heterogeneity within the Latino population.

At a policy level, our findings suggest that the processes leading to racial gaps in suspension and expulsion, especially the differential treatment and support of Black students, begin much earlier in the life course than previously documented in a population-based sample. Even in elementary school, differential treatment/support accounts for the largest share (46 percent) of the racial gap in exclusionary discipline between Black and White students.

Although the children in our sample attended elementary school between 2003 and 2009, prior to the national dialogue and onset of district and state policy reform around reducing suspensions/expulsions particularly in elementary school (DeVos, Nielsen and Azar 2018; US Department of Education and US Department of Justice 2014), our findings can offer useful insights. First, consistent with policies to ban/reduce early suspensions, the finding that Black children who enter the exact same kindergarten with comparable behaviors as White children experience higher suspension suggests that policy reform should target practices that begin *after* school entry. Second, "differential treatment/support" points to the importance of not only reducing punishment, but also of equipping teachers and schools with positive supports, including those that address children's underlying traumas and strengthen their socioemotional skills. Certainly, one possible policy approach is to rethink the processes governing the assignment of disciplinary sanctions. Another is to provide supports to increase teachers' empathy towards students of color and their ability to provide children with additional resources/services. For example, in a recent evaluation of an empathic mindsets intervention, Okonofua, Paunesku and Walton (2016) found a 50 percent reduction in suspension rates during the school year. In light of our findings about the important contribution of differential treatment/support to the macro-level Black-White gap, districts and states should consider policies to not only reduce sanctions, but also to increase supports, in their efforts to reduce disparities at a macro level.

Supplementary Material

Supplementary material is available at *Social Forces* online.

Footnote

1. It is equally valid to ask the question "what would we expect the racial gap to look like in the counterfactual scenario in which Black children had the same average exposures/means as White children (but their own, Black coefficients), or if Black

children had their own exposures/means but these exposures were linked to suspension in the same way as they are for Whites (i.e., the same coefficients as Whites). We expect and find that the story remains similar regardless of reference group.

About the Authors

Jayanti Owens is the Mary Tefft and John Hazen White, Sr. Assistant Professor of Sociology and International and Public Affairs at Brown University. Her research focuses on social disparities in how people in power evaluate the behaviors of others and the consequences of these evaluations for the reproduction of social inequality. Recent papers appear in *Sociology of Education*, *Social Science Research*, *Ethnic and Racial Studies*, *and Sociology of Religion*.

Sara McLanahan is the William S. Tod Professor of Sociology and Public Affairs at Princeton University. Her research focuses on families and child health and wellbeing. She is Editor of the journal *Future of Children*. Recent papers appear in *Journal of Marriage and Family*, *Demography*, *Proceedings of the National Academy of Sciences*, and *Health Affairs*.

References

Achenbach, Thomas M. 1991. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*: Department of Psychiatry, University of Vermont Burlington, VT.

Achenbach, Thomas M., McConaughy Stephanie H., and Howell Catherine T. 1987. "Child/Adolescent Behavioral and Emotional Problems: Implications of Cross-Informant Correlations for Situational Specificity." *Psychological Bulletin* 101(2):213–32.

Anyon, Yolanda, Gregory Anne, Stone Susan, Farrar Jordan, Jenson Jeffrey M., McQueen Jeanette, Downing Barbara, Greer Eldridge, and Simmons John. 2016. "Restorative Interventions and School Discipline Sanctions in A Large Urban School District." *American Educational Research Journal* 53(6): 1663–97.

Barrett, Nathan, McEachin Andrew, Mills Jonathan N., and Valant Jon. 2017. "Disparities in Student Discipline by Race and Family Income." In *educationresearchallianceNOLA.org*, edited by Education Research Alliance, Pp.1–56. New Orleans, LA: Education Research Alliance for New Orleans.

Bates, John E., Bayles Kathryn, Bennett David S., Ridge Beth, and Brown Melissa M. 1991. "Origins of Externalizing Behavior Problems at Eight Years of Age." In *The Development and Treatment of Childhood Aggression*, edited by Pepler Debra J., and Rubin Kenneth H., Pp.93–120. Hillsadale, NJ: Lawrence Erlbaum Associates, Inc.

Bertrand, Marianne, and Pan Jessica. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5(1):32–64.

Bradshaw, Catherine P., Mitchell Mary M., O'Brennan Lindsey M., and Leaf Philip J. 2010. "Multilevel Exploration of Factors Contributing to the Overrepresentation of Black Students in Office Disciplinary Referrals." *Journal of Educational Psychology* 102(2):508–20.

Brame, Bobby, Nagin Daniel S., and Tremblay Richard E. 2001. "Developmental Trajectories of Physical Aggression from School Entry to Late Adolescence." *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 42(4):503–12.

Brooks-Gunn, Jeanne, and Duncan Greg J. 1997. "The Effects of Poverty on Children." *The Future of Children* 7(2):55–71.

Carter, Prudence L. 2005. *Keepin' It Real: School Success Beyond Black and White*. New York: Oxford University Press.

Dance, Lory Janelle. 2002. *Tough Fronts: The Impact of Street Culture on Schooling*. New York: RoutledgeFarmer.

Deater-Deckard, Kirby, and Dodge Kenneth A. 1997. "Externalizing Behavior Problems and Discipline Revisited: Nonlinear Effects and Variation by Culture, Context, and Gender." *Psychological Inquiry* 8 (3):161–75.

DeVos, Betsy, Nielsen Kirstjen M., and Azar Alex M. 2018. "Final Report of the Federal Commission on School Safety. Presented to the President of the United States." in *US Department of Education*. Washington, D.C.: U.S. Department of Education.

DiPrete, Thomas A., and Eirich Gregory M. 2006. "Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments." *Annual Review of Sociology* 32:271–97.

Egalite, Anna J., and Kisida Brian. 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 Alexander Karl L. 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., Alexander Karl L., and Olson Linda Steffel. 2005. "First Grade and Educational Attainment by Age 22: A New Story." *American Journal Of Sociology* 110(5):1458–502.

Fabelo, Tony, Thompson Michael D., Plotkin Martha, Carmichael Dottie, Marchbanks Miner P. III, and Booth Eric A. 2011. "Breaking Schools' Rules: A Statewide Study of How School Discipline Relates to Students' Success and Juvenile Justice Involvement." In *Council of State Governments Justice Center*, Pp.54–55. New York: Council of State Governments Justice Center.

Farkas, George, Grobe Robert P., Sheehan Daniel, and Shuan Yuan. 1990. "Cultural Resources and School Success: Gender, Ethnicity, and Poverty Groups within an Urban School District." *American Sociological Review* 55(1):127–42.

Ferguson, Ann A. 2001. *Bad Boys: Public Schools in the Making of Black Masculinity*. Ann Arbor: University of Michigan Press.

Gilliam, Walter S., Maupin Angela N., Reyes Chin R., Accavitti Maria, and Shic Frederick. 2016. "Do Early Educators' Implicit Biases Regarding Sex and Race Relate to Behavior Expectations and Recommendations of Preschool Expulsions and Suspensions?" In *Yale Child Study Center*, Pp.991–1013. New Haven, CT: Yale University.

Goff, Phillip A., Jackson Matthew C., Leone Di, Lewis Brooke A., Culotta Carmen M., and DiTomasso Natalie A. 2014. "The Essence of Innocence: Consequences of Dehumanizing Black Children." *Journal of Personality and Social Psychology* 106(4):526.

Government Accountability Office, U.S. 2018. *K-12 Education: Discipline Disparities for Black Students, Boys, and Students with Disabilities*, edited by U.S. Government Accountability Office. Washington, D.C.

Gregory, Anne, Skiba Russell J., and Noguera Pedro A. 2010. "The Achievement Gap and the Discipline Gap Two Sides of the Same Coin?" *Educational Researcher* 39(1):59–68.

Gregory, Anne, and Weinstein Rhona S. 2008. "The Discipline Gap and African Americans: Defiance or Cooperation in the High School Classroom." *Journal of School Psychology* 46(4):455–75.

Hirschfield, Paul J. 2008. "Preparing for Prison? The Criminalization of School Discipline in the USA." *Theoretical Criminology* 12(1):79–101.

Howard, Lionel C., Rose Jason C., and Barbarin Oscar A. 2013. "Raising African American Boys: An Exploration of Gender and Racial Socialization Practices." *American Journal of Orthopsychiatry* 83(2): 218–30.

Jacobsen, Wade C., Pace Garrett T., and Ramirez Nayan G. 2018. "Punishment and Inequality at an Early Age: Exclusionary Discipline in Elementary School." *Social forces* 73(3):973–98.

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, Bracy Nicole, Apple Michael, Hirschfield Paul, Casella Ronnie, Gilliom John, Hope Andrew, Lewis Tyson, Lipman Pauline, and Matthew Richard. 2009. *Schools Under Surveillance: Cultures of Control in Public Education.*. New Brunswick, NJ: Rutgers University Press.

Lareau, Annette, and Horvat Erin McNamara. 1999. "Moments of Social Inclusion and Exclusion: Race, Class, and Cultural Capital in Family-School Relationships." *Sociology of Education* 72(1):37–53.

Lewis, Amanda E., and Diamond John B. 2015. *Despite the Best Intentions: How Racial Inequality Thrives in Good Schools.*. New York, NY: Oxford University Press.

Lindsay, Constance A., and Hart Cassandra M. D. 2017. "Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina." *Educational Evaluation and Policy Analysis* 39(3):485–510.

Losen, Daniel J., Hodson Cheri L., Keith Michael A. II, Morrison Katrina, and Belway Shakti. 2015. "Are We Closing the School Discipline Gap?" In *The Civil Rights Project*, 1–50. Los Angeles, CA: University of California, Los Angeles.

Magnuson, Katherine A., and Waldfogel Jane. 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 Nonnemaker James M. 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. 2018. "A Downward Spiral? Childhood Suspension and the Path to Juvenile Arrest." *Sociology of Education* 91(3):183–204.

Morris, Edward W., and Perry Brea L. 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., Walton Gregory M., and Eberhardt Jennifer L. 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 Eberhardt Jennifer L. 2015. "Two Strikes: Race and the Disciplining of Young Students." *Psychological Science* 26(5):617–24.

Okonofua, Jason A., Paunesku David, and Walton Gregory M. 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.

Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal Of Sociology* 108(5):937–75.

Pager, Devah, and Shepherd Hana. 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 Morris Edward W. 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 (Cambridge, England)* 99:17–33.

Ramey, David M. 2015. "The Social Structure of Criminalized and Medicalized School Discipline." *Sociology of Education* 88(3):181–201.

Rios, Victor M. 2011. *Punished: Policing the Lives of Black and Latino Boys.* New York: NYU Press.

Robinson, Keith, and Harris Angel L. 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.

Shollenberger, Tracey L. 2015. "Racial Disparities in School Suspension and Subsequent Outcomes." In *Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion*, edited by Losen Daniel J., Pp.31–44. New York, NY: Teachers College Press.

Skiba, Russell J., Chung Choong-Geun, Trachok Megan, Baker Timberly L., Sheya Adam, and Hughes Robin L. 2014. "Parsing Disciplinary Disproportionality: Contributions of Infraction, Student, and School Characteristics to Out-of-School Suspension and Expulsion." *American Educational Research Journal* 51(4):640–70.

Skiba, Russell J., Michael Robert S., Nardo Abra Carroll, and Peterson Reece L. 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 Williams Natasha T. 2014. "Are Black Kids Worse? Myths and Facts about Racial Differences in Behavior." In *The Equity Project at Indiana University*. Bloomington, IL: Indiana University.

Smolkowski, Keith, Girvan Erik J., Mcintosh Kent, Nese Rhonda N. T., and Horner Robert H. 2016. "Vulnerable Decision Points for Disproportionate Office Discipline Referrals: Comparisons of Discipline for African American and White Elementary School Students." *Behavioral Disorders* 41(4): 178–95.

US Department of Education, and US Department of Justice. 2014. *Dear Colleague Letter on the Nondiscriminatory Administration of School Discipline*. Washington, D.C.: US Department of Education, Office for Civil Rights and US Department of Justice, Civil Rights Division.

VanderWeele, Tyler J., and Knol Mirjam J. 2011. "Interpretation of Subgroup Analyses in Randomized Trials: Heterogeneity versus Secondary Interventions." *Annals of Internal Medicine* 154(10):680–83.

Vavrus, Frances, and Cole KimMarie. 2002. ""I Didn't Do Nothin'": The Discursive Construction of School Suspension." *The Urban Review* 34(2):87–111.

Verhulst, Frank C., Koot Hans M., and Van der Ende Jan. 1994. "Differential Predictive Value of Parents' and Teachers' Reports of Children's Problem Behaviors: A Longitudinal Study." *Journal of Abnormal Child Psychology* 22(5):531–46.

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 Losen Daniel J. 2003. "Defining and Redirecting a School-to-Prison Pipeline." *New Directions for Youth Development* 99:9–15.

Wallace, John M., Goodkind Sara, Wallace Cynthia M., and Bachman Jerald G. 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.

Weinstein, Rhona S. 2002. *Reaching Higher: The Power of Expectations in Schooling*. Cambridge, MA: Harvard University Press.

Welch, Kelly, and Payne Allison A. 2010. "Racial Threat and Punitive School Discipline." *Social Problems* 57(1):25–48.

Wolf, Kerrin C., and Kupchik Aaron. 2017. "School Suspensions and Adverse Experiences in Adulthood." *Justice Quarterly* 34(3):407–30.

Wright, John Paul, Morgan Mark Alden, Coyne Michelle A., Beaver Kevin M., and Barnes J.C. 2014. "Prior Problem Behavior Accounts for the Racial Gap in School Suspensions." *Journal of Criminal Justice* 42(3):257–66.