ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-168

Categorical Inequality in Black and White: Linking Disproportionality across Multiple Educational Outcomes

Kenneth Shores
Pennsylvania State University

Ha Eun Kim
University of California Irvine

Mela Still
University of Pennsylvania

We characterize the extent to which Black-White gaps for multiple educational outcomes are linked across school districts in the United States. Gaps in disciplinary action, grade-level retention, classification into special education and Gifted and Talented, and Advanced Placement course-taking are large in magnitude and correlated. Racial differences in family income and parent education are strikingly consistent predictors of these gaps, and districts with large gaps in one outcome are likely to have large gaps in another. Socioeconomic and segregation variables explain 1.7 to 3.5 times more variance for achievement relative to non-achievement outcomes. Systemic patterns of racial socioeconomic inequality drive inequalities across multiple educational outcomes; however, discretionary policies at local levels are more influential for non-achievement outcomes.

VERSION: December 2019

Suggested citation: Shores, Kenneth A., Ha Eun Kim, and Mela Still. (2019). Categorical Inequality in Black and White: Linking Disproportionality across Multiple Educational Outcomes. (EdWorkingPaper: 19-168). Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-168

Title:

Categorical Inequality in Black and White: Linking Disproportionality across Multiple Educational Outcomes

Authors:
Kenneth Shores (corresponding author)
Pennsylvania State University
Human Development and Family Studies
HHD Building, Room 234
University Park, PA 16802
Email: kshores@psu.edu
Phone: 650.580.4655

Ha (Grace) Eun Kim
University of California Irvine, School of Education
Email: **hekim4@uci.edu**

Mela Still
University of Pennsylvania, Graduate School of Education
Email: melasti@gse.upenn.edu

Abstract:
We characterize the extent to which Black-White gaps for multiple educational outcomes are linked across school districts in the United States. Gaps in disciplinary action, grade-level retention, classification into special education and Gifted and Talented, and Advanced Placement course-taking are large in magnitude and correlated. Racial differences in family income and parent education are strikingly consistent predictors of these gaps, and districts with large gaps in one outcome are likely to have large gaps in another. Socioeconomic and segregation variables explain 1.7 to 3.5 times more variance for achievement relative to non-achievement outcomes. Systemic patterns of racial socioeconomic inequality drive inequalities across multiple educational outcomes; however, discretionary policies at local levels are more influential for non-achievement outcomes.

Keywords: Black-White inequality, categorical inequality, disproportionality: discipline, special education, Gifted and Talented, Advanced Placement

Forthcoming *American Educational Research Journal*

Introduction

An old question about the U.S. education system concerns its relationship to inequality: does formal schooling exacerbate, reinforce, or improve existing inequalities? Of interest to us is the relationship between formal schooling and racial inequality. Much of the literature on racial inequality and schooling concerns differences in test scores, and focuses on whether extant test score disparities are better explained by differences in school quality or home and neighborhood environments (i.e., family socioeconomic status). Though results across various studies differ somewhat, three general conclusions are drawn: (i) racial disparities in test scores are large before students enter school and therefore cannot be attributed to differences in school quality, (ii) these disparities increase modestly as children progress through, which can partially be attributed to differences in school quality, and (iii) these disparities mostly disappear when comparing among children with similar socioeconomic backgrounds (e.g., Fryer & Levitt, 2004; Quinn, 2015; Rothstein & Wozny, 2011).

Despite the emphasis on test scores, racial disparities occur across a multitude of other educational outcomes. For instance, in the 2015-16 school year, relative to White students, Black students were more than three times as likely to receive an out-of-school suspension, more than three times as likely to be retained a grade, half as likely to be classified as Gifted and Talented, and half as likely to take an Advanced Placement course in high school.[1] Can and should racial inequalities in these outcomes be analyzed in the same way as racial inequalities in test scores?

On the one hand, in the context of statistical regression, racial disparities in students' socioeconomic status, such as parental income and educational attainment, might account for these disparities. That is, a regression model might reveal that Black and White students from similar socioeconomic backgrounds are equally likely to be suspended from school (in fact, this is not the case: as we document below, racial disparities in family socioeconomic status do not fully explain racial disparities in disciplinary outcomes). Yet, persons employed in school systems are engaged in the act of disciplining students, classifying them as "gifted and talented," and assigning them to more academically rigorous classes. There is no mechanical process that requires the translation of racial socioeconomic inequality into racial educational inequality in these domains. Rather, it is those participants in formal schooling who are actively engaged in this process of differentially creating and assigning students into categories that bear this responsibility.

In this paper, we draw on the theory of categorical inequality to quantify the extent to which Black students are disadvantaged in their school experience relative to White students. Categorical inequalities result from two distinct processes. First, social categories are created and assigned to individuals by institutions, such as schools. Second, meaningful differences result when these categorical boundaries regularly yield net advantages (or disadvantages) to people on either side of the boundary (Domina, Penner, & Penner, 2017; Massey, 2007; Tilly, 1998, 2003). Systematically excluding and including students from different types of educational opportunities, at times referred to as racialized tracking (Tyson, 2011, 2013), is a paradigmatic example of the process of categorical inequality formation. The theory of categorical inequality is a useful tool for understanding inequality—and racial inequality specifically—in the context of schooling because it identifies school systems as the causal agent that creates and differentially assigns students to socially meaningful categories.

The categories we identify in this text are (a) school disciplinary policy; (b) grade-level retention; (c) Gifted and Talented classification; (d) special education classification; and (e) AP course-taking. As we detail in our review of the literature below, these specific categorical boundaries are meaningful because assignment to a specific category has both contemporaneous and downstream consequences for students' well-being. For example, the literature describes numerous inequalities that result from categorical assignment on outcomes such as test scores (e.g., Gregory, Skiba, & Noguera, 2010), college attendance (e.g., Chajewski, Mattern, & Shaw, 2011), and incarceration (Eren, Lovenheim, & Mocan, 2018). The inequalities we document are those between Black and White students attending public schools in the United States.

To quantify the extent of racial differences in categorical inequality, we construct a unique district-level dataset from the Office of Civil Rights (OCR) Data Collection (the CRDC) and the Common Core of Data (CCD) for academic years 2011-12, 2013-14, and 2015-16. Our analytic sample consists of 1,887 public school districts, which encompasses 71% of the Black public-school population. With these data, we construct racial differences in (a) school discipline rates, (b) grade retention, classification rates into (c) Gifted and Talented, and (d) special education, and (e) AP course-taking.

Using these data, we present five empirical descriptions about the production of categorical inequalities between Black and White students in U.S. public schools. First, educational gaps for a wide range of outcomes are large and systematically correlated. Second, patterns of segregation and socioeconomic differences are strikingly consistent predictors for each of the educational inequalities we measure. Third, after controlling for racial differences in test scores, family income, parental education, and district segregation, we nevertheless identify large categorical inequalities: Black students are suspended 1.5 times more often than White students, and White students are 1.7 times more likely to be classified as Gifted and Talented and 1.3 times more likely to enroll in AP courses relative to their Black peers. Fourth, many districts have large gaps in multiple outcomes. For example, of the five types of gaps we measure (discipline, retention, Gifted and Talented, special education, and AP uptake), 21% of U.S. public school districts (comprising 12% of the Black public-school population in our sample) have at least three of these gaps in the top quintile, indicative of the cumulative disadvantage facing many Black students. Finally, whereas racial differences in family income, parental education, segregation, and racial composition can explain 71% of variation in test score inequality, these same variables can only explain 24 to 48% of the variation in categorical inequalities, suggesting the importance of local contextual factors in the production of categorical inequality.

The concept of categorical inequality attributes disproportionalities in educational outcomes to an education system that creates categories and differentially assigns students to them. This insight causes us to change focus away from the disproportionality itself to its cause. In this way, existing anti-deficit perspectives, which argue that "achievement gaps" are better characterized as the consequences of inequitable social processes, are directly related to the concept of categorical inequality. Ladson-Billings (2006) suggests the term "education debt", reinterpreting test score disparities as a record of the cumulative number of educational opportunities denied to Black students. Similarly, Carter and Welner (2013) and Milner (2012) suggest the term "opportunity gap," likewise shifting attention away from the outcome to its causes, such as the inequities in the social provision of education.

The concept of a categorical inequality is usefully distinct from existing anti-deficit perspectives in that it focuses attention on the causal role schools play in creating and differentially assigning students to socially meaningful categories. Though anti-deficit perspectives recognize the importance of schooling in generating racial differences in educational outcomes, these approaches also incorporate additional explanatory variables, both contemporary and historical, including race-based differences in access to schooling, housing, health facilities, and democratic engagement. Certainly, any complete explanation for current racial differences in educational outcomes will include both in- and out-of-school factors, but the utility of the concept of categorical inequality is that it gives us something to focus on that we can be confident schools are responsible for generating, thereby putting pressure on the educational system to justify why such disproportionalities should be allowed to exist.

Background Literature

In the following sections, we describe previous literature that explains Black-White gaps in discipline rates, grade-level retention, classification (Gifted and Talented and special education) and AP course-taking. We summarize two aspects of research on these outcomes. First, we provide summary information about these categories and the racial inequalities associated with them. We show that these categories are important for short- and long-term student outcomes, and we describe the known magnitude and prevalence of these inequalities. Second, we highlight three types of explanations that are provided for the different Black-White gaps: systemic factors, discretionary factors, and the combination of systemic and discretionary factors.

Systemic factors include variables such as family socioeconomic status (e.g., parental income and educational attainment) and exposure to neighborhood effects (e.g., neighborhood safety and environmental contaminants). Explanations that involve systemic factors are based on the fact that, on average, there are differences in socioeconomic status and exposure to neighborhood effects among racial groups. Researchers then show that systemic factors are predictive of many academic outcomes, and therefore, racial differences in academic outcomes can be at least partially attributed to racial differences in these factors. In contrast, discretionary factors involve decisions made by school leaders, teachers and education policymakers. Explanations that include discretionary factors identify, for example, racial bias in decision making to explain racial differences in academic outcomes. Finally, though less often, researchers adjudicate between the relative influence of systemic and discretionary factors in the production of racial academic inequality.

Black-White Discipline Gaps

Black-White discipline gaps are one of the many facets of educational inequality, as school discipline policies affect multiple student outcomes. Researchers have consistently found a strong positive relationship between time spent in academic learning and student achievement (Brophy, 1988; Fisher et al., 1981; Greenwood, Horton, & Utley, 2002). Therefore, Black-White discipline gaps are thought to exacerbate existing Black-White achievement gaps (Gregory, Skiba, & Noguera, 2010). Similarly, researchers have found adverse effects of suspension on a student's chance of graduating high school (Balfanz, Byrnes, & Fox, 2015). It is known that Black students are suspended at higher rates than White students. For example, data from the U.S. Department of Education's Civil Rights Data Collection (CRDC) shows that, in the 2015-16 school year, Black students were three times more likely to receive an out-of-school suspension compared to White students (similar results have been identified in Losen et al., 2015; Mendez & Knoff, 2003; Skiba et al., 2011).

A common explanation for the Black-White discipline gap is that there are socioeconomic differences among racial groups, on average, and because socioeconomic variables are highly predictive of being disciplined in school, racial discipline gaps are attributable to racial differences in student socioeconomic status and not race *per se* (e.g., Finn & Servoss, 2013; Mendez, Knoff, & Ferron, 2002; Noltemeyer & Mcloughlin, 2010; Sullivan, Klingbeil & Van Norman, 2013). Though this view is commonplace, it has numerous shortcomings. First, multiple studies using different data emphasize that socioeconomic status cannot fully explain discipline gaps (Wallace, Goodkind, Wallace, & Bachman, 2008). Second, focusing on student background ignores the role that in-school factors, such as teacher race, attendance rates, academic achievement, the amount of spending per student, school climate, and principal perspectives on discipline, have on disciplinary gaps (e.g., Bradshaw, Mitchell, O'Brennan, & Leaf, 2010; Christle, Nelson, & Jolivete, 2004; Kinseler, 2011; Fabelo et al., 2011; Gregory, Cornell, & Fan, 2011; Okonofua & Eberhardt, 2015; Okonofua, Paunesku, & Walton, 2016; Skiba et al., 2014). In particular, teacher bias and public policies are known to explain discipline gaps. Using survey methods, Okonofua and Eberhardt (2015) find that when teachers are randomly assigned names suggestive of a student's race, teachers requested more severe punishments for Black students than for White students. Recent evidence links zero tolerance policies, which mandate severe disciplinary consequences regardless of circumstance, to rates of suspension three times larger for Black students relative to White students (Calton, 2012; Curran, 2016; Skiba, 2014; Welch & Payne, 2012). Thus, though socioeconomic variables are often used to explain variation in racial disciplinary gaps, these variables are not fully explanatory and can misattribute student characteristics as the causal agents of disciplinary disproportionality.

Empirical attempts to separate the respective explanatory power of systemic versus discretionary influence are less common (exceptions include Beck & Muschkin, 2012; Christle, Nelson, & Jolivette, 2004; Fabelo et al., 2011; Gregory, Cornell, & Fan, 2011; Mendez, Knoff, & Ferron, 2002; Skiba, 2014). For example, using administrative data on seventh graders in North Carolina, Beck and Muschkin (2012) find that racial differences in socioeconomic status and demographic characteristics (i.e., students' parental education and free lunch status) explain 26% of the difference in rates of infraction. Even after controlling for school fixed effects, 53% of the difference in rates of infraction remain unexplained. Meaning that within schools, for students of different races but similar SES, most of the differences in disciplinary action remain unexplained.

Black-White Grade Retention Gaps

On average, Black students are retained at grade level more than 1.5 times more frequently than White students (Musu-Gillette et al., 2017). Although the theoretical purpose of grade retention is to provide students additional opportunity to acquire grade-level skills, evidence to support this purpose is inconclusive. Descriptively, students who repeat a grade exhibit more problem behaviors, lower achievement, poorer social adjustment, less frequent attendance, more negative attitudes toward school, and greater rates of dropping out (Jimerson & Kaufman, 2003; Jimerson & Ferguson, 2007; Jimerson et al., 2006; U.S Department of Education, National Center for Education Statistics, 2006).

The causal effects of grade-level retention depend on the age at which retention takes place. Retention at younger grades increases student achievement (Jacob & Lefgren, 2004), though these effects may fade out over time (Schwerdt & West, 2012). For students retained in first grade, teacher-reported hyperactivity and peer-rated sadness and withdrawal decreased; however, most of these effects disappeared three years later (Wu, West, & Hughes, 2010). At later grades, retention is harmful. It reduces student achievement and high school completion (Cockx, Picchio, & Baert, 2018; Jacob & Lefgren, 2009) and increases the risk of being convicted of a crime by age 25 (Eren et al., 2018).

The relatively few papers that explain racial disparities in grade retention emphasize socioeconomic characteristics. Student family income, exposure to poverty, and parental education predict grade-level retention (Child Trends Data Bank, 2015). School-level factors may be more important. Stearns, Moller, Blau, and Potochnick (2007) find that the educational experiences of Black students, including educational pessimism, achievement scores, and misbehaviors, are more predictive of grade-level retention than demographic variables. Tomchin and Impara (1992) find that teacher beliefs about a student will influence whether they are retained or not. And LiCalsi, Ozek, and Figlio (2017) show that greater maternal education reduces the likelihood of grade-level retention, even when the student is below the mandatory cutoff for promotion; this result is explained by teachers emphasizing non-test score criteria for students of well-educated mothers.

Black-White Classification Gaps - Gifted and Talented (GT)

As early as 1998, the underrepresentation of Black students has been documented in GT classes. Ford (1998) found that Black students were underrepresented in gifted education by 41% in comparison to their classmates. For the 2015-16 academic year, the Department of Education's Civil Rights Data Collection shows that Black students are half as likely to be classified as GT. These results are confirmed using nationally representative samples of student-level data (Grissom & Redding, 2016). Access to GT can improve reading, math, and science test scores, with effects largest for minority students (Bui, Craig & Imberman, 2014; Card & Giuliano, 2014).

Morris (2002) suggests that the structural and institutional forces surrounding race and class in the U.S. affect Black representation in gifted education. Robinson (2003) argues that the structural inequalities present in society, particularly socioeconomic disparities, are to blame for the disproportionate representation in GT programs. Though systemic factors likely play a role in the availability of GT programming for Black and White students, much of the empirical research focuses on discretionary factors. A key discretionary factor is teacher practices. For example, before students are even tested for GT classification, teachers are less likely to refer Black students for screening (Elhoweris, 2008; Grantham & Whiting, 2008; McBee, 2006) and assessors give students with Black-sounding names lower scores on intelligence tests for GT placement relative to students with non-Black-sounding names (Fields, 2004). Further evidence of teacher discretionary effects comes from teacher-student race matching; when Black teachers are paired with Black students, their students are three times more likely to be assigned to gifted programs in reading than those with teachers of other races/ethnicities (Grissom & Redding, 2016).

Black-White Classification Gaps - Special Education

Whether Black students are over- or under-represented in special education is an ongoing debate. Descriptively, Black students are over-represented in special education (Ahram, Fergus, & Noguera, 2011; U.S. Department of Education, 2016); however, the debate hinges on whether special education services are fairly allocated according to ability. Proponents of the over-representation hypothesis argue that Black students are over-represented in specific categories of special education known as "judgment" categories, as these depend on clinical judgment rather than biologically verifiable data (Harry & Klingner, 2014). An audit study of teachers finds that White male students are suspected of having exceptionalities when they exhibit academic challenges whereas Black male students are suspected of having exceptionalities when they exhibit behavioral challenges. (Fish, 2017). Further, Black students are more likely to be classified as having "lower-status disabilities," such as intellectual disabilities, in schools with larger concentrations of White students, suggesting the effect that racial isolation has on classification decisions (Fish, 2019). Proponents of the under-representation hypothesis argue that Black and White students should be matched according to socioeconomic status, test scores, and (often teacher-reported) behaviors, as qualification for special education is intended to be needs-based. After controlling for these variables, researchers find that Black students are less likely than White students to be classified in a broad range of special needs categories (Hosp

& Reschly, 2004; Gordon, 2018; Morgan, Farkas, Hillemeier, & Maczuga, 2012; Morgan, Farkas, Hillemeier, & Maczuga, 2017; Morgan et al., 2015).

Whether over- or under-representation of Black students into special education results in a categorical inequality depends on whether assignment into special education connotes advantage or disadvantage. Evidence on the effects of special education on student outcomes is limited. For example, Hanushek, Kain, and Rivkin (2006) find that classification into special education status boosts mathematics achievement, suggesting that if Black students are not receiving special education services they need, they may be harmed by under-representation. Over-representation also carries risks, as Black students that are disproportionately classified into special education classes are less likely to receive a rigorous curriculum, have lower chances for eligibility for postsecondary education, and have decreased employment opportunities in their lifetimes (Fierros & Conroy, 2002; Harry & Klingner, 2006; National Research Council, 2002).

Black-White Advanced Placement Gaps

AP scores are a unique measure of students' preparedness for college coursework and can provide students with the opportunity to receive college course credit based on test scores before they begin their education at a post-secondary institution. Participation in an Advanced Placement course while in high school increases students' exposure to highly qualified and motivated teachers and probability of attending a 4-year university (Chajewski et al., 2011; Santoli, 2002). For the 2015-16 academic year, the Department of Education's Civil Rights Data Collection shows that Black students are under-represented in AP courses, as they are half as likely to take an AP course compared to White students. Disparities in AP course-taking have been documented in many other studies (Barnard-Brak, McGaha-Garnett, & Burley, 2011; Conger, Long, & Iatarola, 2009; Taliaferro & DeCuir-Gunby, 2008; Whiting and Ford, 2009).

Explanations for AP course-taking gaps primarily come from case studies in cities and states, rather than inclusive studies of the entire country (examples of such case studies include Cisneros, Gomez, Powers, Halloway-Libell, & Corley, 2014; Corra, Carter, & Carter, 2011; Klopfenstein, 2004; Godly, Monroe & Castma, 2015). In one of the few national studies, Conger, Long, and Iatarola (2009) find that after controlling for a student's pre-high school characteristics, the Black-White gap in advanced course-taking decreased from 10.2 to 7.2 percentage points. Klopfenstein's (2004) study of access in Texas schools finds that income is a driver of inequities in advanced course participation. Klopfenstein (2004) estimates that replacing Black income with White income changes the AP participation gap from about one-third as likely to over one-half as likely, or a reduction in the gap of about 36%.

Differences in income between Black and White students can be used to explain some of the AP course-taking gap but does not completely explain why White students are enrolled in AP courses more than their Black peers (Hallett & Venegas, 2011). Interviews with teachers and staff members have highlighted the role that schools can play in students' course-taking. Through such interviews, Taliaferro and DeCuir-Gunby (2008) find that Black parents and students are not aware of the process for enrolling in AP classes and therefore do not advocate like their White peers for access to those courses. Corra, Carter, and Carter (2011) find that, despite the open enrollment policy of the district, Black students in North Carolina enrolled in advanced courses at lower rates than their SAT scores indicated they were capable of, suggesting that underrepresentation of Black students in AP courses is not exclusively due to academic preparation and can be addressed through school-level policies.

Deconstructing Categorical Inequality

Here, we highlight alternatives to the production of categorical inequalities available to schools. Attending to alternatives highlights the distinction between an explanation and a justification. An explanation provides an account of how or why an event occurred, while a justification interrogates the merits of the explanatory reasons (see Baier, 1958; Dancy, 2000). For example, the research described above shows that race-based differences in family income are predictive of race-based differences in rates of suspension. A common explanation provided for this relationship is that lower-income Black students are more likely to misbehave in school relative to higher-income White students (e.g., Barrett, McEachin, Mills, & Valant, 2017). Justificatory reasoning challenges this explanation in one of three ways: by questioning the premise (e.g., the classification of behaviors into categories of good and bad is necessarily racially constructed); by providing alternative explanations (e.g., teachers may over-identify misbehaviors among lower-income Black students); or by interrogating the discretionary response (e.g., behavioral infractions need not result in school suspensions). Evidence of school-level policies that ameliorate or deconstruct categorical inequalities therefore militates against the justifiability of explanations linking race-based differences in categorical inequality to race-based differences in student conduct.

Racial differences in school disciplinary outcomes can be reduced by eliminating zero tolerance (Calton, 2012; Curran, 2016; Skiba, 2014; Welch & Payne, 2012). In Los Angeles, for example, suspension rates declined by 53% after zero tolerance was eliminated (Berwick, 2015). Restorative justice policies—broadly defined as "an approach to discipline that engages all parties in a balanced practice that brings together all people impacted by an issue or behavior" (González, 2012, p. 281)—can also reduce the school discipline gaps. A longitudinal study on the impact of restorative justice in Denver Public Schools found that the district's overall suspension rates decreased from 10.6% to 5.6% and rates for Black students decreased by 7.2 percentage points compared to 3.6 percentage points for White students (González, 2015; extensive discussion on methods to reduce disciplinary disproportionality can be found in Losen et al., 2015).

Using objective standards (such as test scores and grade point averages) in place of teacher discretion can reduce disproportionality in special education, GT, and AP course-taking. In special education, teachers mediate the process to recommend students, and in some cases, classification into special education is used as a substitute for instructional support when students are struggling academically (Ahram, Fergus, & Nogeura, 2011). Teachers also influence whether students feel they belong in advanced courses, and when students feel like they do not belong, they are less likely to enroll (Tyson, 2011, 2013). A universal screening program based on student test scores for GT enrollment increased classification rates among Black students by 80% (Card & Giuliano, 2016). Further, in the Fall of 2010, Federal Way Public schools in Washington began automatically enrolling students to AP/IB courses if they scored proficient on the state exam. As a result, the rate of Black students enrolled in advanced courses nearly doubled (Rowe, 2017).

Finally, teacher practices can be changed to reduce categorical inequalities. For example, teachers who received a brief intervention to encourage empathy were half as likely to suspend their students compared to those who did not participate in the intervention, and suspension rates for Black students decreased by nearly half (Okonofua et al., 2016). Likewise, teacher instructional practices can also improve academic outcomes for racial minorities. For example, an ethnic studies program, implemented in San Francisco Unified School District and incorporated aspects of culturally relevant pedagogy, increased attendance, grade point average, and credits earned, with effects largest for racial/ethnic minorities (Dee & Penner, 2017).

Schools have exclusive influence over whether students are suspended, retained, classified into special education or GT, or granted access to advanced courses. These decisions are consequential for student outcomes. Racial disparities in each of these categories are prevalent in public school settings throughout the United States, and because disproportionality in these categories is the direct result of actions taken by schools, it is possible they can be undone. In this paper, we seek to quantify the extent to which these categorical inequalities accumulate in school districts and to evaluate whether these differences are largely the result of discretionary practices. Our analysis therefore answers the following research questions:

RQ1: What is the magnitude and variability of categorical inequality?

RQ2: Are categorical inequalities systematically related?

RQ2a: Are categorical inequalities correlated with each other?

RQ2b: Are categorical inequalities predicted by socioeconomic and demographic variables?

RQ2c: Are there districts where categorical inequalities are cumulative?

RQ3: Are categorical inequalities more attributable to discretionary factors than test score inequality?

Data and Analytic Sample

Data for categorical inequalities are taken from the Office of Civil Rights (OCR) Data Collection (CRDC) public-use file for years 2011-12, 2013-14 and 2015-16. These data are available at the school level for each of the three years. We calculate the proportion or percentage of White (%W) and Black (%B) students in each district that have been suspended, retained, classified as special education and GT, and have taken an AP course. When two outcomes are measured as proportions, it is common to calculate the gap in two ways, as a *relative risk ratio* and an *absolute risk difference*. R*elative risk ratios* are more frequently used (Noordzij, van Diepen, Caskey, & Jager, 2017) and are defined as the natural logarithm of the ratio of these proportions, i.e., the $\ln\left(\frac{\%B}{\%W}\right)$. A*bsolute risk differences* are defined as the difference in these proportions, i.e., $\%B - \%W$. Current standards of reporting when outcomes are measured as proportions (e.g., in medical trials) encourage using both relative and absolute risk (Altman & Moher, 2010), and throughout the remainder of the manuscript, we report categorical inequalities in both the relative and absolute risk metrics. In each case, the gap is signed so that larger positive values indicate a gap that is disadvantageous to Black students. Therefore, for disciplinary outcomes, grade level retention, and special education, gaps are generated as the percentage of Black students in the district who have either been disciplined, retained a grade, or classified as special needs relative to the percentage of White students in the district that have been assigned to these categories. For GT classification and AP course-taking, gaps are generated as the proportion of White students relative to the proportion of Black students in the district assigned to the category.[2] For convenience, we frequently use the word "gap" to describe racial disparities in these outcomes, as the word "difference" can be confusing, since only absolute risks are strictly differences. A "gap" is therefore a convenient shorthand for specifying both "relative risk ratios" and "absolute risk differences."

Categorical inequalities in the metrics of relative and absolute risk will occasionally provide different inferences about the level of inequality in a district. Though current guidelines encourage reporting both metrics, there is no standard for which metric is to be emphasized when inferences diverge. To help with interpretation, we provide descriptive information about how these metrics correlate with other district characteristics. Specifically, we compare partial correlations between relative and absolute risk for each of the categorical inequalities we measure. In addition, we compare partial correlations between categorical inequalities (in both metrics) and the mean incidence rate of the category (e.g., the number of White and Black students enrolled in GT divided by the number of White and Black students in the district), the mean level of achievement, and the racial test score gap. Our thought for these latter correlations is that a categorical inequality will be more informative when it tracks an underlying inequality, such as a test score gap, and less informative when it tracks an average, such as the level of achievement or the mean incidence rate. Ultimately, however, we are unable to make definitive statements about a district's level of categorical inequality when, for example, it is ranked higher in the metric of relative risk but lower in the metric of absolute risk. We offer this information as suggestive guidelines for interpretation.

We make three observations. First, categorical inequalities in the metrics of absolute and relative risk are weakly correlated for disciplinary outcomes, moderately correlated for grade retention, classification into GT and AP course-taking, and strongly correlated for classification into special education. Second, differences in the metric of absolute risk are strongly correlated with mean incidence rates for disciplinary outcomes, classification into GT and AP uptake, whereas differences in the metric of relative risk are negatively correlated with mean incidence rates among all categories. Third, test score inequality more closely tracks relative risk among disciplinary outcomes, whereas test score inequality more closely tracks absolute risk for GT classification and AP uptake. We therefore report disproportionality according to both metrics noting that these metrics are poorly aligned, that absolute risk tends to track mean incidence rates, and that the correlation between test score inequality and absolute and relative risk varies by category.

<Insert Table 1 Here>

Table 2 provides descriptive characteristics for predictor variables that will be used in subsequent analysis. These predictor variables are gathered from three additional sources to provide district-by-year covariates. Average district test scores and test score differences between Black and White students come from the Stanford Education Data Archive (SEDA). The SEDA data include average district achievement based on over 200 million standardized achievement test scores for approximately 40 million public school Black and White students in grades 3 through 8 during the 2008-09 through 2014-15 school years for subjects English language arts (ELA) and math. Test score gaps are calculated as the standardized difference in achievement between White and Black students.[3] We gather socio-economic descriptors of the school district from the American Community Survey (ACS) Education Demographic and Geographic Estimates (EDGE) database.[4] The variables we collect are log median family income and the proportion of the adult population with a Bachelor's degree or higher, taken from the supplemental tables total population file. These data are available as district-year levels of income and educational attainment, as well as district-year racial differences in income and educational attainment, for years 2011-12, 2013-14, and 2015-16. Finally, enrollment data are taken from the Common Core of Data (CCD) for years 2011-12, 2013-14, and 2015-16. From the CCD, we calculate the proportions of the school (K-12) population that are Black and White. We also generate between-school measures of Black-White and poor-nonpoor (free/reduced lunch to non-free/reduced lunch) segregation. Additional details are available in the Data Appendix.

Because log ratios are undefined when either the proportions of White or Black students assigned to the category are equal to zero, our dataset consists of two analytic samples. The first includes non-missing relative risk data among each of the categorical inequalities, as well as non-missing data from the predictor variables. This analytic sample includes 4,371 observations, 1,887 public school districts, and 71% of the Black student population in public schools. The second analytic sample includes non-missing data for the absolute risk categorical inequalities and predictor variables and includes 5,523 observations, 2,102 public school districts, and 76% of the Black student population in public schools.[5]

Table 2 therefore includes information on three samples: the full population sample (all available data), the relative risk analytic sample (non-missing data among all variables in years 2011-12, 2013-14, and 2015-16), and the relative risk analytic sample for years 2015-16 (the most recent year data are available). Samples limited to non-missing absolute risk metrics of categorical inequality will be slightly larger, as described above, but descriptive statistics using this sample are nearly identical. Values are similar for both the population and analytic samples; racial family income and parental educational inequality and segregation are slightly higher in the analytic sample. Racial differences in family income and parent educational attainment are substantial in all cases. The average difference in log family income is 0.45 (i.e., 1.6 times higher for White families). The log ratio of adult education for Whites and Blacks is 0.55 (i.e., 1.73 times more White parents with Bachelor's degrees compared to Black parents). Test score gaps are, on average, 0.68 standard deviations, which is nearly identical to the estimates from Reardon, Kalogrides, and Shores (2019), using a slightly different sample of data.

<Insert Table 2 Here>

Methods and Results

Results are presented in three stages corresponding with RQ 1 through RQ 3. Information pertaining to RQ 1 consists of means and standard deviations for the categorical inequalities described previously, as well as estimates of these inequalities, conditional on observed district characteristics. We assess the relatedness of categorical inequality (RQ 2) in three ways. First, we show that these variables are correlated with each other. Second, we estimate a series of bivariate regressions that test whether a common set of variables (family income and parental educational attainment, between-school segregation, and racial demographic composition) consistently predict (in sign and magnitude) different forms of categorical inequality. Third, we identify specific districts that have large disparities in multiple outcomes. Finally (RQ3), we show that racial inequality at the district level is much less predictive of categorical inequality than it is of test score inequality. This benchmarking exercise is intended to demonstrate that local discretionary choices—either through teacher or administrator action or school and district policy—are the foundation upon which categorical inequality is laid.

RQ 1: Magnitudes and Conditional Estimates of Categorical Inequalities

Table 3 presents means and standard deviations (SD) for each of the categorical inequality variables for three samples: the full population (all available data, for years 2011-12, 2013-14, 2015-16), the analytic sample (non-missing data among all variables, respectively for the relative and absolute risk metrics, for years 2011-12, 2013-14, 2015-16), and the analytic sample for 2015-16 (non-missing data for most recent year available). Looking at the analytic sample, discipline rate gaps are also very large, with logged ratios ranging from 1.00 to 1.43 (i.e., Black students are given some form of suspension 2.72 to 4.18 times more often than White students) and absolute differences from 3% to 6%.[6]

Gaps in grade-level retention, GT, special education classification, and AP course-taking are smaller but pronounced, with values at 0.69, 1.05, 0.11, and 0.78, respectively (i.e., Black students are retained 2 times more often than White students; White students are classified as GT 2.9 times more often than Black students; Black students are more likely to be classified as special needs 1.1 times more than White students, and White students take AP classes 2.2 times more often than Black students). In absolute terms, differences range from 2% to 12% and are largest for GT and AP course-taking (at 9% and 12%). Compared to test score gaps, there is between 1.9 to 2.6 times more variation among districts for categorical inequalities (the test gap SD is 0.22, whereas the SD for categorical inequalities range from 0.42 to 0.58). Values from the analytic sample for 2015-16 are slightly higher among disciplinary outcomes and nearly identical for the other outcomes. Finally, if we apply the smaller analytic sample from Panel A (i.e., in the metric of relative risk), the summary statistics for categorical inequalities in the metric of absolute risk are nearly identical (not shown).

<Insert Table 3 Here>

Conditional mean estimates of the gaps for these categorical inequalities are presented in Table 4. We present five types of estimates. The baseline estimate is the unconditional average (values are identical to those found in Table 3); we then provide conditional estimates including control variables for (a) racial socioeconomic gaps, (b) racial socioeconomic gaps and segregation; (c) racial socioeconomic gaps, segregation, and racial differences in test scores; and (d) racial socioeconomic gaps, segregation, racial differences in test scores, and state fixed effects, which test whether disparities are attributable to between-state differences. The generic model is of the form:

(1) $CatIneq_{dy} = \alpha + \boldsymbol{X_{dy}\beta} + \lambda_y + \varepsilon_{dy}$

$CatIneq_{dy}$ is one of seven categorical inequalities variables in district *d* and year *y*; $\lambda_y$are year fixed effects (controlling for average differences over time) and $\boldsymbol{X}$ is a vector of control variables described above. Because each of the control variables has a natural interpretation at zero (namely, that there is zero inequality or segregation in the district), the constant α can be interpreted as the average gap in *Y* when inequality is zero. In Table 4 Panel A, we present estimates of α in the relative risk metric exponentiated, so they can be interpreted as ratios, and, in Panel B, we present estimates of α in the absolute risk metric as percentages.

<Insert Table 4 Here>

Controlling for racial differences in family income and parental educational attainment (Model 2) shrinks the gaps in all cases (relative to unconditional gaps in Model 1 and from Table 3), but only for special education does the gap shrink to zero (in both relative and absolute metrics). Black students are still over two times more likely than White students to face a form of suspension; 1.8 times more likely to be retained a grade; White students are 2.36 times more likely than Black students to be classified as GT and 1.65 times more likely to take an AP class. Including segregation control variables (Model 3) does little to shrink the estimates. Controlling for test score differences (Model 4) has the largest effect. The special education gap is reversed in both relative and absolute metrics, meaning that we can recover estimates comparable to those of Morgan, Farkas, Hillemeier, and Maczuga (2012) using aggregate data. In absolute metrics, differences in GT and AP uptake are also reversed, though this is not replicated in the relative metric. Controlling for state fixed effects (Model 5) has almost no effect on the conditional gap estimate. Thus, in the metric of relative risk, after controlling for racial socioeconomic differences, segregation, test score differences, and between-state differences, Black students are still approximately 1.5 times more likely than Whites to be suspended in some form; White students are still 1.6 times more likely than Black students to be classified as GT and 1.25 times more likely to take an AP class.

These results underscore the prevalence of categorical inequalities. We reiterate here the explanation/justification distinction: to the extent these variables *explain* racial differences in categorical inequality (supposing an accurate causal model can be supplied), they do not *justify* these differences. We now turn to the question of whether these inequalities are systematically related.

RQ 2a: Correlation among Categorical Inequalities

Correlations among the categorical inequalities and test score gaps, in both relative and absolute risk metrics, are shown in Table 5. In the metric of relative risk, the correlations between discipline and test score gaps are between 0.44 and 0.5. Intra-disciplinary correlations are approximately 0.5, meaning that discipline gaps are as internally correlated as they are correlated with test score gaps. Finally, the correlations between disparities in grade retention, classification and AP course-taking and test score differences is nearly 0.4. These non-disciplinary gaps are also internally correlated, with values ranging between 0.08 to 0.37.

In the metric of absolute risk, the correlations between discipline and test score gaps are less than 0.2, though these disciplinary gaps are internally correlated with values ranging between 0.31 to 0.39. In contrast, the correlations between gaps in classification and AP course-taking and test score differences are between 0.36 and 0.51, more than twice as large as the correlations between disciplinary and test score gaps. Thus, in the metric of relative risk, discipline gaps are more strongly correlated with test score gaps, whereas in the metric of absolute risk, gaps in classification and AP course-taking are more strongly correlated with test score gaps.

<Insert Table 5 Here>

RQ 2b: Predictors of Categorical Inequalities

We now show that, on average, gaps in discipline, retention, classification, and AP course-taking are consistently predictable. That is, the district level variables used here consistently predict each of the different categorical inequalities, in terms of sign and of magnitude. Our regression models are of the same form in Equation (1) with the inclusion of state fixed effects. In this case, however, we iteratively include each variable $\boldsymbol{X}$ to estimate a sequence of bivariate regressions; here we are interested in the coefficient β. To facilitate comparability, all dependent and independent variables are standardized based on the analytic samples of the relative and absolute metrics to be mean zero with standard deviation of one. Thus, the values on $\hat{\beta}$ are interpreted as the standard deviation change in $Ineq_{dy}$ for a one-standard deviation change in $\boldsymbol{X}$, net of state and year differences.

We present results in Figure 1. The figure is organized as follows: columns A and B show results for relative and absolute risk metrics, respectively; panels (entitled and enumerated as 01 through 06) correspond to a categorical inequality*;* along the y-axis of each panel are listed the predictor variables; along the x-axis of each panel are the values of $\hat{\beta}$. For each predictor, we plot two estimated $\hat{\beta}$ (along with the 95% confidence interval, with heteroskedasticity robust standard errors): one using the categorical inequality as the outcome variable, and the other using the test score gap as the outcome variable. Estimates for categorical inequalities are shown as gray bars; estimates for test score differences are shown as Black markers. Note that we exclude single suspension rate gaps, as results are nearly identical to multiple suspension rate gaps.

<Insert Figure 1 Here>

As we observed from the correlation coefficients shown in Tables 1 and 5, mean incidence rates strongly predict categorical inequalities in the metric of absolute differences (and especially for disciplinary outcomes), and test score inequality strongly predicts disciplinary gaps in the metric of relative risk and classification and AP course-taking gaps in the metric of absolute risk. Further, in the relative risk metric, discipline gaps are better predicted by test score inequality than the other more academically oriented categorical inequalities, whereas, in the metric of absolute risk, the academically oriented categorical inequalities are better predicted by test score inequality.

In all metrics of risk and for nearly all categorical inequalities, district economic and demographic characteristics are less predictive of categorical inequalities than test score gaps. This result can be seen, for example, by looking at rows 03 and 04 for each of the outcomes. Coefficients using family income and parental educational attainment gaps as predictors are always lower than those for test score gaps. In the metric of relative risk, a striking feature of these data is the positive association between a district's overall level of parental educational attainment (row 08) and the magnitude of its discipline gap. Reardon, et al. (2019) found a similarly strong relationship between district parental educational attainment and test score inequality; here, we see that the same variables predict discipline gaps. At the same time, average parental educational attainment negatively predicts disciplinary gaps in the metric of absolute risk, which is explained by the strong positive correlation between disciplinary gaps and mean incidence rates of disciplinary outcomes: districts with more educated parents, on average, have a lower mean incidence rate of disciplinary infractions, meaning that disciplinary inequality in the metric of absolute risk will also be lower in those districts.

RQ 2c: Cumulative Categorical Inequalities

The preceding results indicate that disparities in disciplinary rates, grade-level retention, classification and AP course-taking are correlated; moreover, these categorical inequalities are predicted by a common set of variables. Therefore, we would expect some districts that are ranked highly in one outcome to be ranked highly in others. Here we present evidence in confirmation of that hypothesis. Tables 6 and 7 display districts that are ranked in the top 10 of the respective categorical inequalities in the metrics of relative and absolute risk, respectively, for the year 2015-16. Districts are sorted by rank-order and highlighted if they are observed more than once among the multiple outcomes; the district's gap (non-standardized) is shown in parentheses.

<Insert Table 6 Here>

<Insert Table 7 Here>

In the metric of relative risk, nine of ten districts with the largest test score gap have at least one additional categorical inequality in the top ten. Berkeley Unified, CA has a special education gap in the top ten. Atlanta Public School District, GA has single and multiple suspension gaps in the top ten. Chapel Hill-Carrboro City Schools, NC has a special education gap in the top ten. Minneapolis Public School District, MN has a grade retention gap in the top ten. Oakland Unified, CA has a multiple suspension gap in the top ten. Asheville City Schools, NC has a special education gap in the top ten. Cleveland Heights-University Heights City, OH has a multiple suspension rate gap in the top ten. Charlottesville City Public Schools, VA has a special education gap in the top ten.

Other districts are repeated multiple times but do not have test score gaps in the top ten. Newark Public Schools, NJ is a top ten disciplinary gap for in-school suspension, single suspensions and multiple suspensions. Parma City, OH is a top ten gap for in-school suspensions and AP course-taking. Osseo Public School District, MN is repeated for in-school and single suspension rate gaps. Thomasville City, GA is repeated for in-school and multiple suspension gaps, as well as grade retention gaps. Elmhurst School District 205, IL is repeated for single and multiple suspension gaps, as well as grade retention gaps. Interestingly, there are no districts in the top ten for Gifted and Talented present elsewhere.

In the metric of absolute risk, the list of repeating districts is smaller, which reflects the relatively lower correlation among the categorical inequalities (see the un-highlighted portion of Table 5; note the list of top-10 test score differences is different because of the different sample sizes between the risk metrics). Nevertheless, the following districts have a test score gap and another categorical inequality in the top ten: Chapel Hill-Carrboro City Schools, NC (AP uptake), Asheville City Schools, NC (special education), and Shaker Heights City, OH (AP uptake gap). In addition, Brownsville School District, PA, England School District, AR and Merrillville Community Schools, IN have top-ten gaps in at least two disciplinary outcomes. Finally, Groton School District, CT has a top-ten gap in grade retention and special education, and Ogden City District, UT has a top-ten gap in grade retention and Gifted and Talented.

We can expand what constitutes large categorical inequalities by identifying districts that have gaps in the top quintile. Because discipline gaps are highly correlated internally, we count top quintile gaps across unique outcomes. Specifically, we count the number of gaps in the top quintile that a district has in (1) in-school suspensions or single- or multiple- out-of-school suspension, (2) grade level retention, (3) GT, (4) special education, or (5) AP course-taking, for a total of up to five gaps a district can have in the top quintile. In Table 8, we present the 16 and 12 districts in the relative and absolute risk metrics, respectively, that have five categorical inequalities in the top quintile. Next to the district names, we display the test score gap (not included in the ranking) and each of the indicators of categorical inequality. These values are standardized so that values greater than zero are above the sample mean; each unit-increase is a standard deviation increase above the sample mean.

<Insert Table 8 Here>

In the relative risk metric, five of the 16 districts (highlighted in blue) were also present in Table 6 (Asheville City, NC; Batavia Unified School District, IL; Charlottesville Public School District, VA; Edenton-Chowan Schools, NC; Thomasville City, GA), meaning that these districts not only have five categorical inequalities in the top quintile, but also have at least one indicator of categorical inequality in the top ten. Four of these 16 districts are also in the state of North Carolina. In the metric of absolute risk, three of the 12 districts (highlighted in blue) also have at least one indicator of categorical inequality in the top ten (Asheville City, NC; Shaker Heights City, OH; Thomasville City, GA). Finally, four districts have five unique categorical inequalities in the top quintile in both the relative and absolute metrics (indicated with an asterisk): Asheville City, NC; Huntington Union Free School District, NY; Oxford School District, MS; Shaker Heights City, OH; Thomasville, GA.

To what extent do racial differences in test scores and socioeconomic status (operationalized as family income and parent educational attainment) predict which districts have multiple categorical inequalities in the top quintile? To answer this question, we generate a dependent variable called $MultGaps_{dy}$ that is equal to one of four values $\in \{0,1,2,3\}$ if district *d* has 0, 1, 2, or 3 or more unique gaps in the top quintile. We top-code gaps greater than or equal to three because only about 6% of districts have four or more outcomes in the top quintile. In our data, for years 2015-16, in the respective metrics of relative and absolute risk, 37% and 30% of districts have zero gaps in the top quintile; 31% and 36% have one gap in the top quintile; 19% and 21% have two gaps in the top quintile; and 13% and 13% have three or more gaps in the top quintile. These 13% of districts with three or more gaps in the top quintile enroll 12% and 21% of the Black student population in the sample, respectively. For years 2011-12 and 2013-14, the distribution is very similar.

Our model is comparable to Equation (1), except that we replace the dependent variable $Ineq_{dy}$ with $MultGaps_{dy}$. Estimation takes the form of a multinomial logistic regression:

$$(2)\ MultGaps_{d} = \alpha + \beta_1\mathbf{X} + \lambda_y + \varepsilon_{dy}$$

The multinomial logit model expresses the log-odds of having one to three educational gaps in the top quintile relative to having zero gaps in the top quintile as a linear function of the included predictor variables. Specifically, with four possible outcomes, and setting the base outcome to be zero gaps in the top quintile, the model is equivalent to the three conditional logistic regressions:

$$(2a)\ Ln\left(\frac{MultGaps = 1}{MultGaps = 0}\right) = \alpha + \beta_1\mathbf{X} + \lambda_y + \varepsilon_{dy}$$

$$(2b)\ Ln\left(\frac{MultGaps = 2}{MultGaps = 0}\right) = \alpha + \beta_1\mathbf{X} + \lambda_y + \varepsilon_{dy}$$

$$(2c)\ Ln\left(\frac{MultGaps = 3}{MultGaps = 0}\right) = \alpha + \beta_1\mathbf{X} + \lambda_y + \varepsilon_{dy}$$

We estimate two versions of these models for each of the risk metrics, replacing $\boldsymbol{X}$ with test score gaps in the first model and family income and parental educational attainment gaps in the second model. See note for an explanation of why the ordered logistic regression is not appropriate for this research question.[7]

To facilitate understanding and to identify nonlinear effects in the predictor variables, we convert the log-odds coefficients into average marginal effects for different values of $\boldsymbol{X\beta}$—i.e., we calculate the average change in probability of selecting $MultGaps_d \in \{0,1,2,3\}$ for different values of the predictor variables. The range of values for the predictor variables is based on percentiles of $\boldsymbol{X}$, specifically the percentiles at the 5th to 95th percentiles in increments of five (e.g., 5th, 10th, 15th, …, 95th). We calculate average marginal effects for test score gaps and socioeconomic (family income and parental educational attainment) gaps. Results are displayed in Figure 2.

<Insert Figure 2 Here>

The average marginal effects based on variation in the percentiles of test score and socioeconomic gaps are displayed in two columns, corresponding to the variable groupings described above: the left panel displays racial test score differences; the right panel displays racial socioeconomic differences. Results for the metrics of relative and absolute risk are shown in the first and second rows, respectively. Probabilities for the different outcomes of $MultGaps_d$ are displayed as separate lines, with 95% confidence intervals (adjusted for heteroskedasticity) as range areas.

One of the most compelling results we see is that the probabilities of having zero or multiple categorical inequalities as a function of racial test score or socioeconomic inequality is similar for both the relative and absolute risk metrics. Thus, despite the previous differences we have observed between these two metrics, cumulative disadvantage experienced by Black students is similarly predicted by both racial test score and socioeconomic inequality. For both metrics of risk, the probability that a district will have zero educational gaps in the top quintile in either metric steeply declines as a function of racial test score differences. Between the 5th and 10th percentiles of racial test score inequality, the probability of having zero gaps in the top quintile is roughly 50%. Between the 80th to 95th percentiles of test score inequality, the probability of having zero gaps in the top quintile is between 12% and 17.5% for absolute and relative risk metrics, respectively. In contrast, the probability of having two or three or more gaps of categorical inequality in the top quintile increase dramatically as test score gaps increase. When test score gaps are below the 10th percentile, the probability of having two or three or more gaps is less than 5%. When test score gaps are in the 80th percentile or higher, the probability of having multiple gaps in the top quintile is between 18.5 and 26% for absolute and relative risk metrics, respectively.

Having multiple categorical inequalities in both metrics of risk is also predicted by racial soci-oeconomic differences. Between the 5th and 10th percentiles of racial socioeconomic inequality, the probability of having zero gaps in the top quintile is greater than 50%. Between the 80th to 95th per-centiles of racial socioeconomic inequality, the probability of a district having zero gaps in the top quintile is between 27.6% and 13.4% in the relative and absolute risk metrics, respectively. In both the relative and absolute risk metrics, at the 90th percentile of racial socioeconomic inequality, the proba-bility of having one, two or three or more gaps is effectively the same. This last result indicates the great deal of variability in categorical inequality in districts with the most racial socioeconomic ine-quality, which we have characterized as school discretionary factors.

RQ 3: Variance Decomposition of Categorical Inequalities

An important feature of categorical inequalities is that schools both create and sort students into them, meaning that categorical inequalities are shaped by discretionary processes taking place in schools and classrooms. We demonstrate the relevance of local discretionary processes by comparing how predictive a vector of district covariates (indicators of racial socioeconomic inequality, segregation, and racial composition shown in Table 2) are for test score gaps relative to the measures of categorical inequality. We reason that, while test score differences are produced by both in- and out-of-school processes, Reardon, et al. (2019) show that out-of-school processes are much more predictive of test score differences. While we observed previously (Figure 1) that district characteristics are predictive of categorical inequality, categorical inequalities are still produced in schools. Thus, we would expect the district characteristics to be less explanatory for categorical inequalities than they are for test score gaps. We therefore estimate a series of regression as follows:

(3) $$Ineq_{dy} = \alpha + \beta_1 \Delta Inc + \beta_2 \Delta Educ + \beta_3 SegPoor + \beta_4 SegRace + \beta_5 Inc + \beta_6 Educ + \beta_7 Blk + \beta_8 Wht + \delta_{s[d]} + \lambda_y + \varepsilon_{dy}$$

$Ineq_{dy}$ represents one of eight outcomes, including gaps in test scores, discipline rates, grade retention, classification, and AP course-taking. We include predictor variables from Table 2, including gaps in household income and parental educational attainment, free/reduced lunch segregation (FLE), racial segregation, average household income and parent education, and racial composition. As before, we include state and year fixed effects. Finally, for each of the indicators of categorical inequality, we estimate three complementary models that include the same listed covariates plus (a) test score gaps, (b) the mean incidence rate of the categorical inequality, and (c) test score gaps and the mean incidence rate together. Results are shown in Table 9.

<Insert Table 9 Here>

Our operating hypothesis is that these socioeconomic and demographic variables will be more explanatory of test score inequality than they will be of categorical inequality. The adjusted R^2 statistic tells us how much of the variance in either test score gaps or categorical inequalities these variables explain, and it is therefore useful to compare across models. Overall, racial differences in socioeconomic status, racial and socioeconomic segregation, and racial composition variables explain 1.65 to 6.9 times more variation for test score gaps relative to categorical inequalities. To see this, first note that 71% and 69% of variation in test score gaps in the metrics of relative and absolute risk, respectively, (the difference is entirely attributable to sample size differences, since the outcome is the test score gap) is explained using these variables. For suspension gaps in the relative risk metric, we can explain 32% of the variance for in-school suspension rates, 38% for single suspension rates, and 43% for multiple suspension rates; in absolute risk, we can explain 26% of the variance for in school suspension rates, 17% for single suspension rates, and 27% for multiple suspension rates (M1 in Panels A and B). For grade retention gaps, we can explain 32% and 10% of the variance in both risk metrics; for GT, we can explain 30% and 35%; for special education, we can explain 20% and 26%; and for AP uptake, we can explain 24% and 37%. Thus, it appears that local discretionary processes are more likely to explain variation in categorical inequality than variables indicative of home and neighborhood context. Even when we include test score gaps as right-hand side variables, the R^2 ranges from 12% to 47%, still much less than what is possible for predicting test score inequality. These last results support the hypothesis that education policy and local decision makers are primarily responsible for disparities in these types of outcomes.

Finally, further studies may benefit from considering the interactions between the overall incidence rate, the risk metric, and the outcome considered. Here we comment on the effect of including the mean incidence rate as a predictor on the model's R^2. First, including mean incidence rates has almost no effect on the R^2 for categorical inequalities in the metric of relative risk; this result is not surprising, as mean incidence was not correlated with categorical inequalities in this metric (see Table 1 and Figure 1). However, its influence on disciplinary disproportionality in the metric of absolute risk is profound, elevating the R^2 for those outcomes above that of models in which test score gaps are the dependent variable. Hence, we conclude that absolute risk, at least with respect to disciplinary outcomes, is a construct better represented by the overall incidence rate than by the level of racial inequality in the district.

Implications for Research, Policy, and Practice

We have argued that conventional research methods that explain racial disparities in categorical inequalities as a function of in- and out-of-school factors are inappropriate, as categorical inequalities are the consequence of decisions made by school personnel. Researchers, therefore, should limit their explanations to in-school factors, or at least provide justification for the use of out-of-school factors in these explanatory models. Policymakers and practitioners can also conclude that the onus is on them to remedy these disparities. We have highlighted some policies and practices that can reduce categorical inequalities. These include eliminating zero tolerance policies, implementing uniform (or blinded) selection criteria for assigning students into advanced courses, and changing teacher practices (for example, by providing empathy-promoting interventions or offering culturally relevant pedagogy). Though widespread implementation of these policies and practices is likely to reduce the magnitudes of categorical inequalities among Black and White students, a larger question raised by the analysis is the justification for creating categories that enable stratification in the first place. If schools are to be an equalizer, then a necessary condition is for them to assign students equitably to categories. Yet, the mere existence of socially relevant categories will create constant pressure to leverage those categories for stratification. Therefore, any proposed benefit of category creation should be evaluated against corresponding threats to equal opportunity brought about by them.

Discussion and Conclusion

Explanations for categorical inequalities in disciplinary policy, grade-level retention, Gifted and Talented classification, special education classification, and Advanced Placement course-taking often emphasize racial differences in test scores and socioeconomic status. Our results generally conform to prior knowledge in this area, as we show that Black-White disparities in various school-sanctioned policies are large and persistent, even after controlling for student characteristics.

Our research diverges from prior efforts by highlighting the central role school systems play compounding disadvantages experienced by Black students. As evidence, consider the following: first, we find that racial differences in disciplinary rates, grade-level retention, classification into Gifted and Talented and special education, and AP course-taking are large in magnitude, even after controlling for racial differences in socioeconomic status and neighborhood contexts. Thus, the over- (or under-) representation of Black students in salient categories of disadvantage (or advantage) is a common feature of schooling in the United States.

Second, districts with greater racial socioeconomic differences, levels of segregation, and socioeconomic status also have larger achievement and disciplinary gaps. While the association between family resources and student achievement is a well-documented explanation for achievement gaps, we have asked whether we should expect the same pattern for other outcomes. Even if racial socioeconomic inequality is predictive of Black-White differences in behavioral problems, these differences need not consequentially result in disproportionate disciplinary actions, such as suspending students, which exacerbate disadvantages experienced by Black students.

Third, educational gaps are commonly found clustered together in districts. Thirteen percent of school districts with up to 21% of the Black student population in our sample have three or more unique gaps in the top quintile. Moreover, the probability of having multiple gaps in the top quintile steeply increases as test scores and racial socioeconomic inequality increases. Thus, in districts where Black students are the worst off socioeconomically, school districts increase the educational disadvantages Black students face, as they are more likely to be disciplined, retained, classified into special education, or excluded from Gifted and Talented or AP courses.

Finally, we find that, in addition to having larger test score gaps, Black students are more likely to be disciplined and have multiple large educational gaps in districts with the highest concentration of college educated parents. The data do not reveal why this pattern exists. Still, it suggests that highly educated parents may be able to procure educational advantages for their children in the form of achievement and access to advanced courses and, at the same time, insulate them from the detrimental aspects of schooling, such as disciplinary policy and grade-level retention.

The study of categorical inequality yields the following key insight: racial disparities in outcomes like disciplinary policy, classification and access to advanced courses are strictly the result of decisions made by school personnel, and because schools create these categories, they can also eliminate them. As evidence, we note that while our primary results indicate that many districts are replicating Black-White inequality, district socioeconomic status and demography explain 1.65 to 6.9 times more variation for test score gaps relative to categorical outcomes. Thus, there are many districts where discipline, retention, classification, and AP gaps are smaller than would be expected based on observed data, making it possible to identify them and learn from their examples.

References

Ahram, R., Fergus, E., & Noguera, P. (2011). Addressing racial/ethnic disproportionality in special education: Case studies of suburban school districts. *Teachers College Record*, *113*(10), 2233-2266.

Baier, K. (1958). *The moral point of view.* Ithaca, NY: Cornell University Press.

Balfanz, J., Byrnes, V., & Fox, J. (2015). Sent home and put off track: The antecedents, disproportionalities, and consequences of being suspended in the 9th grade. In D. J. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 17– 30). New York, NY: Teachers College Press.

Barnard-Brak, L., McGaha-Garnett, V. & Burley, H. (2011). Advanced Placement course enrollment and school-level characteristics. *NASSP Bulletin, 93*(3), 165-174.

Barrett, N., McEachin, A., Mills, J., & Valant, J. (2017). *Technical report: Disparities in student discipline by race and family income.* New Orleans, LA: Education Research Alliance for New Orleans.

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

Berwick, C. (2015, March 17). Zeroing out Zero Tolerance. *The Atlantic*. Retrieved from https://www.theatlantic.com/education/archive/2015/03/zeroing-out-zero-tolerance/388003/

Bradshaw, C. P., Mitchell, M. M., O'Brennan, L. M., & Leaf, P. J. (2010). Multilevel explorations of factors contributing to the overrepresentation of Black students in office discipline referrals. *Journal of Educational Psychology, 102*(2), 508 –520.

Brophy, J. (1988). Classroom management as socializing students into clearly articulated roles. *Journal of Classroom Interaction*, *33*(1), 1–4.

Bui, S. A, Craig, S. G., & Imberman, S. A. (2014). Is gifted education a bright idea? Assessing the impact of gifted and talented programs on students. *American Economic Journal: Economic Policy, 6*(3), 30-62.

Card, D. & Giuliano, L. (2014). Does gifted education work? For which students? No. w20453. National Bureau of economic research.

Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *PNAS, 113*(48), 13678-13683.

Calton, M. T. (2012). Black male perspectives on their educational experiences in high school. *Urban Education, 47*(6), 1055-1085.

Carter, P. L., & Welner, K. G. (Eds.). (2013). *Closing the opportunity gap: What America must do to give every child an even chance*. Oxford University Press.

Chajewski, M., Mattern, K. D., & Shaw, E. J. (2011). Examining the role of Advanced Placement Exam participation in 4-year college enrollment. *Educational Measurement: Issues and Practice, 30*(4), 16-27.

Child Trends Data Bank. (2015). *Children who repeated a grade*. Bethesda, MD: Child Trends.

Christle, C., Nelson, C. M., & Jolivette, K. (2004). School Characteristics Related to the Use of Suspension. *Education and Treatment of Children, 27*(4), 509-526.

Cisneros, J., Gomez, L. M, Powers, J. M, Halloway-Libell, J. & Corley, K. M. (2014). The Advanced Placement opportunity gap in Arizona: Access, participation, and success. *AASA Journal of Scholarship and Practice, 11*(2), 20-32.

Cockx, B., Picchio, M., & Baert, S. (2018). Modeling the effects of grade retention in high school. *Journal of Applied Econometrics, 34*(3), 403-424.

Conger, D., Long, M.C., & Iatarola, P. (2009). Explaining race, poverty, and gender disparities in advanced course-taking. *Journal of Policy Analysis and Management, 28*(4), 555-576.

Corra, M., Carter, J. S., & Carter, S. K. (2011). The interactive impact of race and gender on high school advanced course enrollment. *The Journal of Negro Education, 80*(1), 33-46.

Curran, C. F. (2016). Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student behavior. *Educational Evaluation and Policy Analysis 38*(4), 647-668.

Dancy, J. (2000). *Practical Reality*. New York, NY: Oxford University Press.

Dee, T. S., & Penner, E. K. (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal*, *54*(1), 127-166.

Domina, T., Penner, A., & Penner, E. (2017). Categorical Inequality: Schools As Sorting Machines. *Annual Review of Sociology*, *43*, 311-330.Eide, E. R., & Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review, 20*(6), 563-576.

Elhoweris, H. (2008). Teacher judgment in identifying gifted/talented students. *Multicultural Education*, *15*(3), 35-38.

Eren, O., Lovenheim, M. F., & Mocan, N. H. (2018). *The Effect of Grade Retention on Adult Crime: Evidence from a Test-Based Promotion Policy* (No. w25384). National Bureau of Economic Research.

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

Fields, S. A. (2004). Assessor effects on the evaluation of the WISC-III. *Unpublished manuscript*.

Fierros, E. G., & Conroy, J. W. (2002). Double jeopardy: An exploration of restrictiveness and race in special education. In D. J. Losen & G. Orfield (Eds.), *Racial inequity in special Education* (pp. 39–70). Cambridge, MA: Harvard Education Press.

Finn, J. D., & Servoss, T. J. (2014). Misbehavior, suspensions, and security measures in high school: Racial/ethnic and gender differences. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, *5*(2), 11.

Fish, R. E. (2017). The racialized construction of exceptionality: Experimental evidence of race/ethnicity effects on teachers' interventions. *Social Science Research, 62*, 317-334.

Fish, R. E. (2019). Standing Out and Sorting In: Exploring the Role of Racial Composition in Racial Disparities in Special Education. *American Educational Research Journal*, 1-36.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1981). Teaching behaviors, academic learning time, and student achievement: An overview. *Journal of Classroom Interaction*, *17*(1), 2–15.

Ford, D. Y. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *The Journal of Special Education*, *32*(1), 4-14.

Fryer Jr., R. G., & Levitt, S. D. (2004). Understanding the Black-White test score gap in the first two years of school. *Review of Economics and Statistics*, *86*(2), 447-464.

Godley, A., Monroe, T. & Castma, J. (2015). Increasing access to and success in Advanced Placement English in Pittsburgh public schools. *English Journal, 105*(1), 28-34.

González, T. (2012). Keeping kids in schools: Restorative justice, punitive discipline, and the School to Prison Pipeline. *Journal of Law & Education, 41*(2), 281-335.

González, T. (2015). Socializing schools: Addressing racial disparities in discipline through restorative justice. In D. J. Losen (Ed.), *Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion*. New York, NY: Teachers College Press.

Gordon, N. (2018). Who is in Special Education and who has access to related services? New evidence from the National Survey of Children's Health. *The Brookings Institution Evidence Speaks Reports*, *2*(46).

Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review, 31*(3), 328-349.

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

Gregory, A., Cornell, D., & Fan, X. (2011). The relationship of school structure and support to suspension rates for Black and White high school students. *American Educational Research Journal, 48*(4), 904-934.

Grissom, J. A. & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, *2*(1), 1-25.

Hallett, R. E, & Venegas, K. M. (2011). Is increased access enough? Advanced Placement courses, quality, and success in low-income urban schools. *Journal for the Education of the Gifted, 34*(3), 468-487.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). Inferring program effects for special populations: Does special education raise achievement for students with disabilities?. *Review of Economics and Statistics, 84*(4), 584-599.

Harry, B., & Klingner, J. (2014). *Why are so many minority students in special education?*. New York, NY: Teachers College Press.

Hosp, J. L., & Reschly, D. J. (2004). Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. *Exceptional Children*, *70*(2), 185-199.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of economics and statistics*, *86*(1), 226-244.

Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics, 1*(3), 33-58.

Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: Academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly, 22*(3), 314-339.

Jimerson, S. R., & Kaufman, A. M. (2003). Reading, writing, and retention: A primer on grade retention research. *The Reading Teacher, 56*(7), 622-635.

Jimerson, S. R., Pletcher, S. M. W., Graydon, K., Schnurr, B. L, Nickerson, A. B, & Kundert, D. K. (2006). Beyond grade retention and social promotion: Promoting the social and academic competence of students. *Psychology in the Schools, 43*: 85–97.

Klopfenstein, K. (2004). Advanced Placement: Do minorities have equal opportunity? *Economics of Education Review, 23*(2), 115-131.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational researcher*, *35*(7), 3-12.

LiCalsi, C., Ozek, U., & Figlio, D. (2017). The uneven implementation of universal school policies: Maternal education and Florida's mandatory grade retention policy. *Education Finance and Policy*, (Just Accepted), 1-53.

Losen, D., Hodson, C. I., Keith, I. I., Michael, A., Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap? K-12 racial disparities in school discipline*. Los Angeles, CA: University of California, Los Angeles.

Massey, D. (2007). *Categorically Unequal: The American Stratification System*. New York, NY: Russell Sage Foundation.

McBee, M. T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *The Journal of Secondary Gifted Education, 17*(2), 103-111.

Mendez, L. M. R., & Knoff, H. M. (2003). Who gets suspended from school and why: A demographic analysis of schools and disciplinary infractions in a large school district. *Education and Treatment of Children, 26*(1), 30-51.

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

Milner IV, H. R. (2012). Rethinking achievement gap talk in urban education. *Urban Education, 48(1), 3-8*.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2012). Are minority children disproportionately represented in early intervention and early childhood special education?. *Educational Researcher*, *41*(9), 339-351.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2017). Replicated evidence of racial and ethnic disparities in disability identification in U.S. schools. *Educational Researcher*, *46*(6), 305-322.

Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher, 44*(5), 278–292.

Morris, J. E. (2002). African American students and gifted education: The politics of race and culture. *Roeper Review, 24*(2), 59-62.

Musu-Gillette, L., de Brey, C., McFarland, J., Hussar, W., Sonnenberg, W., and Wilkinson-Flicker, S. (2017). *Status and trends in the education of racial and ethnic Groups 2017* (NCES 2017-051). U.S.

Department of Education, National Center for Education Statistics. Washington, DC. Retrieved 05/2018 from http://nces.ed.gov/pubsearch

National Research Council. 2002. Minority Students in Special and Gifted Education. Washington, DC: The National Academies Press. https://doi.org/10.17226/10128.

Noltemeyer, A., & Mcloughlin, C. S. (2010). Patterns of exclusionary discipline by school typology, ethnicity, and their interaction. *Penn GSE Perspectives on Urban Education, 7*(1), 27-40.

Noordzij, M., van Diepen, M., Caskey, F. C., & Jager, K. J. (2017). Relative risk versus absolute risk: one cannot be interpreted without the other. *Nephrology Dialysis Transplantation*, *32*(suppl_2), ii13-ii18.

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

Okonofua, J. A., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences, 113*(19), 5221-5226.

Quinn, D. M. (2015). Kindergarten Black–White test score gaps: Re-examining the roles of socioeconomic status and school quality with new data. *Sociology of Education*, *88*(2), 120-139.

Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The geography of racial/ethnic test score gaps. American Journal of Sociology, 124(4), 1164-1221.

Reardon, S. F., Shear, B. R., Castellano, K. E., & Ho, A. D. (2017). Using heteroskedastic ordered probit models to recover moments of continuous test score distributions from coarsened data. *Journal of Educational and Behavioral Statistics*, *42*(1), 3-45.

Robinson, N. M. (2003). Two wrongs do not make a right: Sacrificing the needs of gifted students does not solve society's unresolved problems. *Journal for the Education of the Gifted, 26*(4), 251-273.

Rothstein, J., & Wozny, N. (2013). Permanent income and the Black-White test score gap. *Journal of Human Resources*, *48*(3), 510-544.

Rowe, C. (2017, April 2). Gifted programs across Washington leave out Black and Latino students — but Federal Way is one model for change. *The Seattle Times*. Retrieved from https://www.seattletimes.com/education-lab/gifted-programs-across-washington-leave-out-Black-and-latino-students-except-in-federal-way/

Santoli, S. P. (2002). Is there an Advanced Placement advantage? *American Secondary Education, 30*(3), 23-35.

Schwerdt, G., & West, M. R. (2012). The effects of early grade retention on student outcomes over time: Regression discontinuity evidence from Florida. Program on Education Policy and Governance Working Papers Series. PEPG 12-09. *Program on Education Policy and Governance, Harvard University*.

Skiba, R. J. (2014). The failure of zero tolerance. *Reclaiming children and youth, 22*(4), 27-33.

Skiba, R. J., Chung, C. G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. A*merican Educational Research Journal, 51*(4), 640-670.

Skiba, R. J., Horner, R. H., Chung, C. G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review, 40*(1), 85-107.

Stearns, E., Moller, S., Blau, J., & Potochnick, S. (2007). Staying back and dropping out: The relationship between grade retention and school dropout. *Sociology of Education, 80*(3), 210-240.

Sullivan, A. L., Klingbeil, D. A., & Van Norman, E. R. (2013). Beyond behavior: Multilevel analysis of the influence of sociodemographics and school characteristics on students' risk of suspension. *School Psychology Review, 42*(1), 99-113.

Taliaferro, J. D., & DeCuir-Gunby, J. T. (2008). African American educators' perspectives on the Advanced Placement opportunity gap. *The Urban Review, 40*(2), 164-185.

Tilly, C. (1998). *Durable inequality*. Berkeley, CA: University of California Press.

Tilly, C. (2003). Inequality, democratization, and de-democratization. *Sociological Theory, 21*(1), 36-43.

Tomchin, E. M., & Impara, J. C. (1992). Unraveling teachers' beliefs about grade retention. *American Educational Research Journal, 29*(1), 199-223.

Tyson, K. (2013). Tracking segregation, and the opportunity gap. In P.L. Carter & K.G. Welner (Eds.) *Closing the opportunity gap: What America must do to give every child an even chance* (169-180). Oxford Scholarship Online.

Tyson, K., ed. (2011). *Integration interrupted: Tracking, Black students, and acting White after Brown*. New York, NY: Oxford University Press.U.S. Department of Education, National Center for Education Statistics. (2006). *The Condition of Education 2006* (NCES 2006-071). Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs. (2016). *38th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2016*. Washington, D.C.

Wallace Jr., J. M., Goodkind, S., Wallace, C. M., & Bachman, J. 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.

Welch, K., & Payne, A. A. (2012). Exclusionary school punishment: The effect of racial threat on expulsion and suspension. *Youth Violence and Juvenile Justice, 10*(2), 155-171.

Whiting, G. W., & Ford, D. Y. (2009). Multicultural issues: Black students and Advanced Placement classes: Summary, concerns, and recommendations. *Gifted Child Today*, *32*(1), 23-26.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal*, *6*(1), 58.

Williams, R. (2009). Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods & Research*, *37*(4), 531-559.

Williams, R. (2010). Fitting heterogeneous choice models with oglm. *Stata Journal*, *10*(4), 540.

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology, 102*(1), 135-152.

NOTES

[1] Authors' calculations taken from the 2015-16 Office of Civil Rights Data Collection.
[2] Data can be downloaded from the Office of Civil Rights website (https://www2.ed.gov/about/offices/list/ocr/data.html).
[3] Data can be downloaded from the Stanford Education Data Archive website (https://cepa.stanford.edu/seda/overview). The data file we use is called SEDA_geodist_pool_CS_v20.dta.
[4] Data can be downloaded from the American Community Survey website (https://nces.ed.gov/programs/edge/Demographic/ACS).
[5] The estimated is 71% and 76% is based on the total number of Black students in the analytic sample divided by the total number of Black students in the full dataset. Counts of Black students are taken from the CCD.
[6] The average absolute differences for disciplinary gaps reported in Table 3 are smaller than those reported from the Office of Civil Rights (OCR). For example, the OCR reports absolute average differences for all types of suspension rates using 2011-12 data as 11% (Civil Rights Data Collection Data Snapshot: School Discipline. Issue Brief No. 1, March 2014). Part of the explanation can be attributed to differences in sample composition. However, the fuller explanation is that the population absolute difference is not equal to the weighted (or unweighted) average absolute difference (i.e., $\frac{\sum Sus_i^b}{\sum Tot_i^b} - \frac{\sum Sus_i^w}{\sum Tot_i^w} \neq \frac{1}{n}\sum\left[\left(\frac{Sus_i^b}{Tot_i^b} - \frac{Sus_i^w}{Tot_i^w}\right) * \left(\frac{Tot_i^b + Tot_i^w}{Tot^b + Tot^w}\right)\right]$). Population absolute differences reported in the OCR are derived from the former expression and average absolute differences reported here are derived from the latter. Therefore, we should not expect the disproportionalities to be identical.
[7] Note that because our outcome variable is ordered, the ordered logistic regression might seem preferable. However, ordinal logistic regressions assume parallel lines for the coefficients, i.e., that the coefficients are the same for all cases of $MultGaps_d$. Given the research question is to identify whether any variables *differently* predict changes in the number of large gaps, this assumption makes the results uninformative. Alternative models to the multinomial logistic regression exist, such as the generalized ordered logistic regression and ordinal generalized linear models (Williams, 2006; 2009; 2010). These models are more efficient than multinomial logistic regression but provide comparable results (Williams, 2006; 2010). Because individuals are more acquainted with multinomial logistic, we opt for this modeling approach. We replicate estimation using generalized ordered logistic; results are nearly identical and are available upon request.

Tables and Figures

Table 1: Categorical Inequalities, Relative Risk Ratios and Absolute Risk Differences

	Relative Risk *LN(%B/%W)*		Absolute Risk *%B - %W*		Relative Risk *LN(%B/%W)*		Absolute Risk *%B - %W*
ISS		0.25		**Gifted/Talented**		0.47	
Mean Incidence	-0.15		0.75	Mean Incidence	-0.15		0.41
Test score level	0.21		-0.05	Test score level	-0.02		-0.09
Test score gap	0.46		0.01	Test score gap	0.36		0.51
Single suspension		0.40		**IEP/504**		0.82	
Mean Incidence	-0.20		0.60	Mean Incidence	-0.08		-0.06
Test score level	0.29		-0.25	Test score level	0.16		0.26
Test score gap	0.53		0.26	Test score gap	0.27		0.30
Multiple suspension		0.15		**AP uptake**		0.57	
Mean Incidence	-0.12		0.84	Mean Incidence	-0.02		0.58
Test score level	0.23		-0.31	Test score level	-0.05		0.10
Test score gap	0.57		0.14	Test score gap	0.37		0.61
Grade retention		0.53					
Mean Incidence	-0.15		0.10				
Test score level	0.41		0.08				
Test score gap	0.36		0.24				

Note: This table provides partial correlation coefficients between the categorical inequalities in the metrics of relative and absolute risk shown in Table 2 and three predictor variables: (a) the average incidence rate of the categorical inequality (e.g., the number of Black and White students enrolled in Gifted and Talented in the district-year divided by the number of Black and White students in the district-year), (b) the average level of achievement in the district, and (c) the racial test score gap in the district. All correlations are based on the residuals of variables from models that include state and year fixed effects and are weighted by Black and White district enrollment. All outcomes are taken from the 2011/12, 2013/14, and 2015/16 Office of Civil Rights Data Collection. The analytic sample is restricted to non-missing data for both predictors and inequality statistics in the metric of relative risk included in Tables 2 and 3. The analytic sample includes 4,371 observations, 1,887 public school districts, and 71% of Black students in the K-12 public school population.

Table 2: Summary Statistics, Predictor Variables

	Full Sample			Analytic Sample		
	Mean	Std. Deviation	N	Mean	Std. Deviation	N
Income gap	0.40	0.34	15373	0.45	0.24	4371
Parent education gap	0.54	0.63	15373	0.55	0.45	4371
FRL segregation	0.09	0.10	39498	0.14	0.10	4371
Racial segregation	0.11	0.14	35671	0.16	0.14	4371
Income level	10.92	0.33	39516	10.90	0.30	4371
Education level	0.29	0.14	39516	0.30	0.12	4371
Percent Black	0.16	0.20	41877	0.24	0.19	4371
Percent White	0.58	0.27	41877	0.46	0.23	4371
Mean Achievement	0.06	0.33	34448	0.00	0.30	4371
Test score gap	0.66	0.22	7786	0.68	0.22	4371

Note: This table provides means and standard deviations for district characteristics. Income and parent education are taken from the 2011/12, 2013/14, and 2015/16 ACS Census waves. FRL (free/reduced lunch), racial segregation, and proportions Black and White are estimated using data from the Common Core of Data for years 2011/12, 2013/14, and 2015/16. District test score gaps and levels are taken from the Stanford Education Data Archive and are averaged over the sample period 2008/09 to 2014/15. Income and income gaps are expressed as log ratios. Test scores are reported in standard deviation units. All means and standard deviations are weighted by Black and White district enrollment. The full sample includes all non-missing data for the variables. The analytic sample is restricted to non-missing data for both predictors and inequality statistics in the metric of relative risk ratios included in Table 2. The analytic sample includes 4,371 observations (1,887 public school districts) in the three years and 71% of Black students in the K-12 public school population.

Table 3: Summary Statistics, Categorical Inequalities

	Full Sample			**Analytic Sample**			**Analytic Sample for 2015/16**		
	Mean	Std. Deviation	N	Mean	Std. Deviation	N	Mean	Std. Deviation	N
Panel A: Relative Risk Ratios (LN(%B/%W))									
In-school suspension (ISS) gap	1.08	0.59	18040	1.00	0.42	4371	1.05	0.42	1388
Single suspension gap	1.17	0.61	16863	1.07	0.42	4371	1.13	0.39	1388
Multiple suspension gap	1.48	0.70	13184	1.43	0.58	4371	1.50	0.56	1388
Grade retention gap	0.81	0.69	11924	0.69	0.49	4371	0.67	0.48	1388
Gifted/Talented gap	0.86	0.69	12461	1.05	0.50	4371	1.05	0.48	1388
IEP/504 gap	0.08	0.46	18553	0.11	0.30	4371	0.09	0.23	1388
AP uptake gap	0.60	0.72	13013	0.78	0.47	4371	0.77	0.41	1388
Panel B: Absolute Risk Differences (%B - %W)									
In-school suspension (ISS) gap	6%	10%	35574	6%	5%	5523	6%	4%	1737
Single suspension gap	4%	7%	35627	4%	2%	5523	4%	2%	1737
Multiple suspension gap	3%	6%	35627	3%	3%	5523	3%	2%	1737
Grade retention gap	2%	8%	29266	2%	2%	5523	2%	3%	1737
Gifted/Talented gap	7%	11%	22618	9%	7%	5523	9%	8%	1737
IEP/504 gap	0%	11%	35633	2%	4%	5523	2%	4%	1737
AP uptake gap	9%	14%	20463	12%	9%	5523	13%	8%	1737

Note: This table provides means and standard deviations for district characteristics. Gaps in *Panel A* are reported as log ratios (i.e., relative risk ratios). Gaps in *Panel B* are reported differences in percentages (i.e., absolute risk differences). For *Panels A* and *B,* values greater than zero imply disproportionality unfavorable to Black students. For discipline, grade retention and IEP/504 (special education) gaps, gaps are based on the percentage of the Black student population in a district relative to the percentage of the White student population in a district. For Gifted and Talented and AP uptake, gaps are based on the percentage White relative to the percentage Black. All outcomes are taken from the 2011/12, 2013/14, and 2015/16 Office of Civil Rights Data Collection. All means and standard deviations are weighted by Black and White district enrollment. The full sample includes all non-missing data for the variable. The analytic sample is restricted to non-missing data for both predictors and inequality statistics included in Tables 1 and 2. The analytic samples in *Panels A* and *B* include 4,371 and 5,523 observations in the three years and 71% and 76% of Black students in the K-12 public school population, respectively. The 2015/16 analytic samples in *Panels A* and *B* include 1,388 and 1,737 districts and 71% and 76% of Black students in the K-12 public school population for that year, respectively.

Table 4: Unconditional and Conditional Gap Estimates

	Model 1	Model 2	Model 3	Model 4	Model 5
Panel A: Relative Risk Ratios (LN(%B/%W))					
In-School Suspension (ISS) gap	2.71	2.41	2.32	1.45	1.46
Single suspension gap	2.92	2.53	2.37	1.49	1.42
Multiple suspension gap	4.16	3.07	2.74	1.50	1.48
Grade retention gap	1.99	1.80	1.74	1.09	1.03
Gifted/Talented gap	2.85	2.26	2.25	1.70	1.60
IEP/504 gap	1.12	1.01	1.03	0.85	0.83
AP uptake gap	2.19	1.65	1.66	1.27	1.25
Panel B: Absolute Risk Differences (%B - %W)					
In-School Suspension (ISS) gap	7%	6%	6%	6%	6%
Single suspension gap	4%	3%	3%	2%	2%
Multiple suspension gap	3%	2%	2%	2%	2%
Grade retention gap	2%	1%	1%	0%	0%
Gifted/Talented gap	10%	6%	5%	-1%	-2%
IEP/504 gap	2%	0%	1%	-3%	-3%
AP uptake gap	13%	8%	7%	-2%	-4%
Racial Socioeconomic Gap		X	X	X	X
Segregation			X	X	X
Test Score Gap				X	X
State Fixed Effects					X

Note: This table provides estimated unconditional and conditional gaps for each of the inequality measures from the 2011/12, 2013/14, and 2015/16 Office of Civil Rights Data Collection. The statistics reported here are the constant of the model $Y = \alpha + \boldsymbol{X\beta} + \lambda_y + \varepsilon$, where λ_y are year fixed effects, and $\boldsymbol{X}$ is a vector of control variables described below and included in Table 1; each regression is weighted by Black and White enrollment. Values in *Panel A* are exponentiated so to be interpreted as a simple ratio; values in *Panel B* correspond to the unadjusted constant. Model 1 is a summary statistic of the data, equivalent to values in Table 2, net of year fixed effects. Model 2 is estimated controlling for racial socioeconomic differences. Model 3 is estimated controlling for Model 2 covariates plus segregation. Model 4 is estimated controlling for Model 3 covariates plus test score gaps. Model 5 includes all covariates in Model 4 plus state fixed effects. Because each of the control variables indicate no racial inequality or segregation at zero, the constant in the regression model is interpretable as the conditional gap. All outcomes are taken from the 2011/12, 2013/14, and 2015/16 Office of Civil Rights Data Collection. The analytic samples in *Panels A* and *B* include 4,371 and 5,523 observations in the three years and 71% and 76% of Black students in the K-12 public school population, respectively.

Table 5: Partial Correlations, Inequality Variables

	Test	ISS	Single	Multiple	Retention	G/T	IEP/504	AP
Panel A: Relative Risk Ratios (LN(%B/%W))								
Test score gap	1.00							
ISS gap	0.44***	1.00						
Single suspension gap	0.46***	0.50***	1.00					
Multiple suspension gap	0.50***	0.51***	0.54***	1.00				
Grade retention gap	0.37***	0.32***	0.38***	0.38***	1.00			
Gifted/Talented gap	0.37***	0.12***	0.11***	0.13***	0.08***	1.00		
IEP/504 gap	0.37***	0.22***	0.24***	0.25***	0.27***	0.14***	1.00	
AP uptake gap	0.38***	0.15***	0.12***	0.15***	0.09***	0.37***	0.14***	1.00
Panel B: Absolute Risk Differences (%B - %W)								
Test score gap	1.00							
ISS gap	0.10***	1.00						
Single suspension gap	0.18***	0.31***	1.00					
Multiple suspension gap	0.11***	0.36***	0.39***	1.00				
Grade retention gap	0.17***	0.06***	0.11***	0.07***	1.00			
Gifted/Talented gap	0.43***	0.05***	0.11***	0.09***	0.06***	1.00		
IEP/504 gap	0.36***	0.07***	0.06***	-0.01	0.18***	0.07***	1.00	
AP uptake gap	0.51***	0.04**	0.08***	0.03*	0.03*	0.37***	0.06***	1.00

Note: This table provides partial correlation coefficients for the categorical inequalities shown in Table 2. The correlations are based on the residuals from models that include state and year fixed effects and are weighted by Black and White district enrollment. Highlighting indicates correlations emphasized in the text. All outcomes are taken from the 2011/12, 2013/14, and 2015/16 Office of Civil Rights Data Collection. The analytic sample is restricted to non-missing data for both predictors and inequality statistics included in Tables 1 and 2. The analytic samples in *Panels A* and *B* include 4,371 and 5,523 observations in the three years and 71% and 76% of Black students in the K-12 public school population, respectively. * p<0.05; ** p<0.01; ***p<0.001

Table 6: Top 10 Districts Observed Black-White Categorical Inequality, Relative Risk Ratios, by Outcome (Sorted in Rank Order), 2015-16

Test Score Gap	In-School Suspension Gap	Single Suspension Gap	Multiple Suspension Gap
BERKELEY UNIFIED, CA (1.60)	NEWARK PUBLIC SD, NJ (3.89)	FAIRFIELD SD, CT (3.59)	ATLANTA PUBLIC SCHOOLS, GA (5.22)
ATLANTA PUBLIC SCHOOLS, GA (1.53)	GREENSBURG SALEM SD, PA (3.23)	ATLANTA PUBLIC SCHOOLS, GA (3.06)	MASON CITY, OH (4.31)
CHAPEL HILL, NC (1.45)	OXFORD CITY, AL (2.89)	NEBO DISTRICT, UT (2.85)	ELMHURST SD 205, IL (3.64)
MINNEAPOLIS PUBLIC SD., MN (1.44)	PARMA CITY, OH (2.82)	NEWARK PUBLIC SD, NJ (2.41)	FULTON COUNTY, GA (3.54)
ELMHURST SD 205, IL (1.43)	TURLOCK UNIFIED, CA (2.77)	SOUTH WINDSOR SD, CT (2.36)	NEWARK PUBLIC SD, NJ (3.46)
OAKLAND UNIFIED, CA (1.42)	WEST SHORE SD, PA (2.77)	STRATFORD SD, CT (2.35)	OAKLAND UNIFIED, CA (3.35)
CLEVELAND HEIGHTS, OH (1.36)	OSSEO PUBLIC SD, MN (2.56)	OSSEO PUBLIC SD, MN (2.17)	CLEVELAND HEIGHTS, OH (3.29)
ASHEVILLE CITY SCHOOLS, NC (1.36)	DAVIS DISTRICT, UT (2.47)	COLLEGE STATION ISD, TX (2.17)	DOWNINGTOWN AREA SD, PA (3.16)
ORLEANS PARISH, LA (1.34)	WESTERN PLACER UNIFIED, CA (2.40)	ELMHURST SD 205, IL (2.15)	THOMASVILLE CITY, GA (3.15)
CHARLOTTESVILLE , VA (1.27)	THOMASVILLE CITY, GA (2.34)	PHOENIXVILLE AREA SD, PA (2.10)	ORANGE CITY, OH (3.11)
Grade Retention Gap	**Gifted and Talented Gap**	**IEP/504 Gap**	**AP Uptake Gap**
DELRAN TOWNSHIP SD, NJ (3.25)	SIKESTON R-6, MO (3.11)	NORTHVILLE PUBLIC SCHOOLS, MI (1.12)	EUCLID CITY, OH (2.97)
MINNEAPOLIS PUBLIC SD., MN (3.06)	MORGAN COUNTY, GA (2.84)	BERKELEY UNIFIED, CA (1.03)	CARTHAGE ISD, TX (2.83)
BUFORD CITY, GA (2.97)	EDENTON-CHOWAN SCHOOLS, NC (2.79)	CRESTWOOD SD, MI (0.99)	LAFAYETTE CO SD, MS (2.71)
WESTERN PLACER UNIFIED, CA (2.90)	WASHINGTON LOCAL, OH (2.70)	VIDALIA CITY, GA (0.93)	BRENHAM ISD, TX (2.65)
CANYONS DISTRICT, UT (2.75)	WESTSIDE COMM SCH, NE (2.66)	MONTGOMERY ISD, TX (0.92)	INDIANAPOLIS PUBLIC SCHOOLS, IN (2.56)
GREENSBURG SALEM SD, PA (2.68)	GRIFFITH PUBLIC SCHOOLS, IN (2.53)	ASHEVILLE CITY SCHOOLS, NC (0.91)	LAURENS 56, SC (2.49)
THOMASVILLE CITY, GA (2.61)	MANKATO PUBLIC SD, MN (2.53)	BATAVIA USD 101, IL (0.90)	WEST JASPER CONSOLIDATED, MS (2.43)
ELMHURST SD 205, IL (2.50)	CARLISLE AREA SD, PA (2.38)	CHAPEL HILL, NC (0.88)	SPARTANBURG 07, SC (2.25)
BARNEGAT TOWNSHIP SD, NJ (2.50)	SEATTLE PUBLIC SCHOOLS, WA (2.38)	CHARLOTTESVILLE , VA (0.84)	PARMA CITY, OH (2.20)
AUSTIN PUBLIC SD, MN (2.46)	JEFF DAVIS COUNTY, GA (2.38)	CEDAR FALLS COMM SD, IA (0.84)	NORTH PIKE SD, MS (2.18)

Note: The districts identified here have the largest 10 gaps in the metric of *relative risk ratios* for 2015-16 in either test scores, in school suspension rates, single suspension rates, multiple suspension rates, grade retention rates, Gifted and Talented, special education (IEP/504), or AP uptake. Top 10 identified from 2015/16 analytic sample, i.e. 1,388 districts and 71% of Black students in the K-12 public school population for that year. Blue highlighting indicates a district that has a top 10 gap in at least two outcomes.

Table 7: Top 10 Districts Observed Black-White Categorical Inequality, Absolute Risk Differences, by Outcome (Sorted in Rank Order), 2015-16

Test Score Gap	In-School Suspension Gap	Single Suspension Gap	Multiple Suspension Gap
BERKELEY UNIFIED, CA (1.60)	JACKSON CO SD, MS (32%)	CALHOUN 01, SC (20%)	FORREST COUNTY SD, MS (18%)
ATLANTA PUBLIC SCHOOLS, GA (1.53)	CARUTHERSVILLE 18, MO (29%)	WARREN SD, AR (15%)	SHARON CITY SD, PA (17%)
CHAPEL HILL, NC (1.45)	N. LITTLE ROCK SD, AR (25%)	MERRILLVILLE COMM SCH, IN (13%)	WEST MIFFLIN AREA SD, PA (16%)
MINNEAPOLIS PUBLIC SD., MN (1.44)	BROWNSVILLE AREA SD, PA (25%)	GREENWOOD 52, SC (13%)	MARION 10, SC (15%)
ELMHURST SD 205, IL (1.43)	SUMTER COUNTY, GA (24%)	BROWNSVILLE AREA SD, PA (12%)	MERRILLVILLE COMM SCH, IN (14%)
OAKLAND UNIFIED, CA (1.42)	PADUCAH INDEPENDENT, KY (23%)	WESTERN PLACER UNIFIED, CA (12%)	ROSEVILLE COMM SCH, MI (13%)
CLEVELAND HEIGHTS, OH (1.36)	BREWTON CITY, AL (23%)	VICKSBURG WARREN SD, MS (11%)	ORANGEBURG 03, SC (12%)
ASHEVILLE CITY SCHOOLS, NC (1.36)	DELMAR SD, DE (23%)	BENTON SD, AR (11%)	ENGLAND SD, AR (12%)
SHAKER HEIGHTS CITY, OH (1.35)	ENGLAND SD, AR (21%)	WINDHAM SD, CT (11%)	KALAMAZOO PUBLIC SCHOOLS, MI (12%)
UNIVERSITY CITY, MO (1.34)	TROY CITY, AL (21%)	MARYSVILLE JOINT UNIFIED, CA (11%)	WESTERN HEIGHTS, OK (11%)
Grade Retention Gap	**Gifted and Talented Gap**	**IEP/504 Gap**	**AP Uptake Gap**
W ST PAUL-MENDOTA HTS-EAGAN, MN (25%)	PATERSON PUBLIC SD, NJ (79%)	BATAVIA USD 101, IL (23%)	CHARLOTTESVILLE , VA (51%)
ROCHESTER PUBLIC SD, MN (18%)	SAGINAW SD OF THE CITY OF, MI (50%)	NORTHVILLE PUBLIC SCHOOLS, MI (21%)	TUPELO PUBLIC SD, MS (49%)
MCLEAN COUNTY USD 5, IL (17%)	WAUKEGAN CUSD 60, IL (48%)	WEST DEPTFORD TOWNSHIP SD, NJ (19%)	CLARKE COUNTY, GA (47%)
SHAWNEE HEIGHTS, KS (13%)	PLYMOUTH-CANTON COMM SCH, MI (45%)	GROTON SD, CT (18%)	THOMASVILLE CITY, GA (45%)
BATAVIA CITY SD, NY (13%)	OGDEN CITY DISTRICT, UT (43%)	PONCA CITY, OK (18%)	M S D WASHINGTON TOWNSHIP, IN (42%)
CONEJO VALLEY UNIFIED, CA (12%)	NEWBURGH CITY SD, NY (42%)	STILLWATER, OK (18%)	M S D PIKE TOWNSHIP, IN (42%)
GROTON SD, CT (11%)	SPRINGFIELD SD 186, IL (42%)	BIBB COUNTY, AL (17%)	CHAPEL HILL, NC (38%)
POUDRE SD R-1, CO (11%)	BURLINGTON CITY PUBLIC SD, NJ (40%)	LONG BEACH CITY SD, NY (17%)	DEKALB COUNTY, GA (37%)
BALDWIN-WHITEHALL SD, PA (11%)	ROCKFORD SD 205, IL (39%)	CEDAR FALLS COMM SD, IA (16%)	RICHLAND 01, SC (37%)
OGDEN CITY DISTRICT, UT (10%)	SAGINAW TOWNSHIP COMM SCH, MI (39%)	ASHEVILLE CITY SCHOOLS, NC (15%)	SHAKER HEIGHTS CITY, OH (36%)

Note: The districts identified here have the largest 10 gaps in the metric of *absolute risk differences* for 2015-16 in either test scores, in school suspension rates, single suspension rates, multiple suspension rates, grade retention rates, Gifted and Talented, special education (IEP/504), or AP uptake. Top 10 identified from 2015/16 analytic sample, i.e. 1,737 districts and 75% of Black students in the K-12 public school population for that year. Blue highlighting indicates a district that has a top 10 gap in at least two outcomes.

Table 8: Districts with Five Unique Black-White Gaps in the Top Quintile, 2015-16

Panel	District Name	Ach Gap	ISS	1 Sus	2+ Sus	Retention	GT	IEP/504	AP
Panel A: Relative Risk Ratios (LN(%B/%W))	ASHEBORO CITY SCHOOLS, NC	0.25	0.15	0.59	-0.11	1.04	1.84	1.41	1.26
	ASHEVILLE CITY SCHOOLS, NC*	3.12	1.31	0.53	1.38	2.77	1.21	3.49	1.85
	AUBURN WASHBURN, KS	-0.42	1.03	1.41	1.90	2.21	0.95	2.23	1.41
	BATAVIA USD 101, IL	1.71	1.26	2.37	0.67	3.34	1.57	3.44	1.60
	CHARLESTON 01, SC	2.42	-0.01	1.06	1.42	0.73	1.16	0.96	1.52
	CHARLOTTESVILLE, VA*	2.71	1.73	2.18	-0.70	3.29	0.98	3.18	2.02
	CITY OF MONROE SD, LA	0.87	1.13	-0.36	-0.26	0.90	1.53	1.82	0.98
	CUSD 200, IL	2.22	1.92	2.47	2.08	0.96	1.95	1.88	1.29
	EDENTON-CHOWAN SCHOOLS, NC	0.59	0.15	-0.07	1.16	1.15	3.60	2.30	1.73
	HUNTINGTON UNION FREE SD, NY*	2.53	1.03	0.71	1.93	1.83	1.13	1.75	2.13
	MANHATTAN-OGDEN, KS	-0.39	0.00	0.97	1.43	1.99	2.33	1.86	1.20
	MCLEAN COUNTY USD 5, IL	0.82	0.43	1.17	0.53	0.81	0.95	1.54	1.39
	NEW HANOVER COUNTY SCHOOLS, NC	1.45	0.62	0.38	0.89	1.04	1.51	2.19	1.87
	OXFORD SD, MS*	1.74	1.12	1.86	1.35	1.55	2.01	1.83	2.27
	THOMASVILLE CITY, GA*	1.78	3.13	1.68	2.94	4.04	2.21	1.82	3.19
	WOODRIDGE LOCAL, OH	1.14	-0.44	-0.17	0.50	2.62	1.12	0.96	2.78
Panel B: Absolute Risk Differences (%B-%W)	ALACHUA, FL	2.21	-1.43	0.55	1.24	1.24	1.47	1.26	1.79
	AMORY SD, MS	-0.09	0.47	1.04	-0.18	0.91	0.62	2.09	0.44
	ASHEVILLE CITY SCHOOLS, NC*	3.11	0.92	0.87	1.54	1.90	2.25	3.63	1.87
	DECATUR CITY, GA	1.84	1.47	0.77	-0.58	1.26	0.64	1.78	2.05
	GENEVA CITY SD, NY	0.93	1.67	-0.20	1.35	2.44	2.79	2.02	1.32
	GLYNN COUNTY, GA	0.55	2.18	0.60	0.29	0.40	0.47	0.99	1.27
	HUNTINGTON UNION FREE SD, NY*	2.53	-0.39	-0.12	1.57	2.09	2.75	2.15	2.19
	LA GRANGE ISD, TX	0.61	2.13	-2.22	-1.34	3.02	0.61	3.19	1.84
	OXFORD SD, MS*	1.75	1.71	1.11	0.21	2.61	2.31	0.99	2.27
	SHAKER HEIGHTS CITY, OH*	3.03	-1.00	1.16	-0.17	0.72	1.36	2.56	2.87
	THOMASVILLE CITY, GA*	1.78	3.73	1.76	0.75	1.16	1.52	1.11	3.98
	TUPELO PUBLIC SD, MS	1.16	2.32	0.29	1.42	1.05	1.74	1.04	4.49

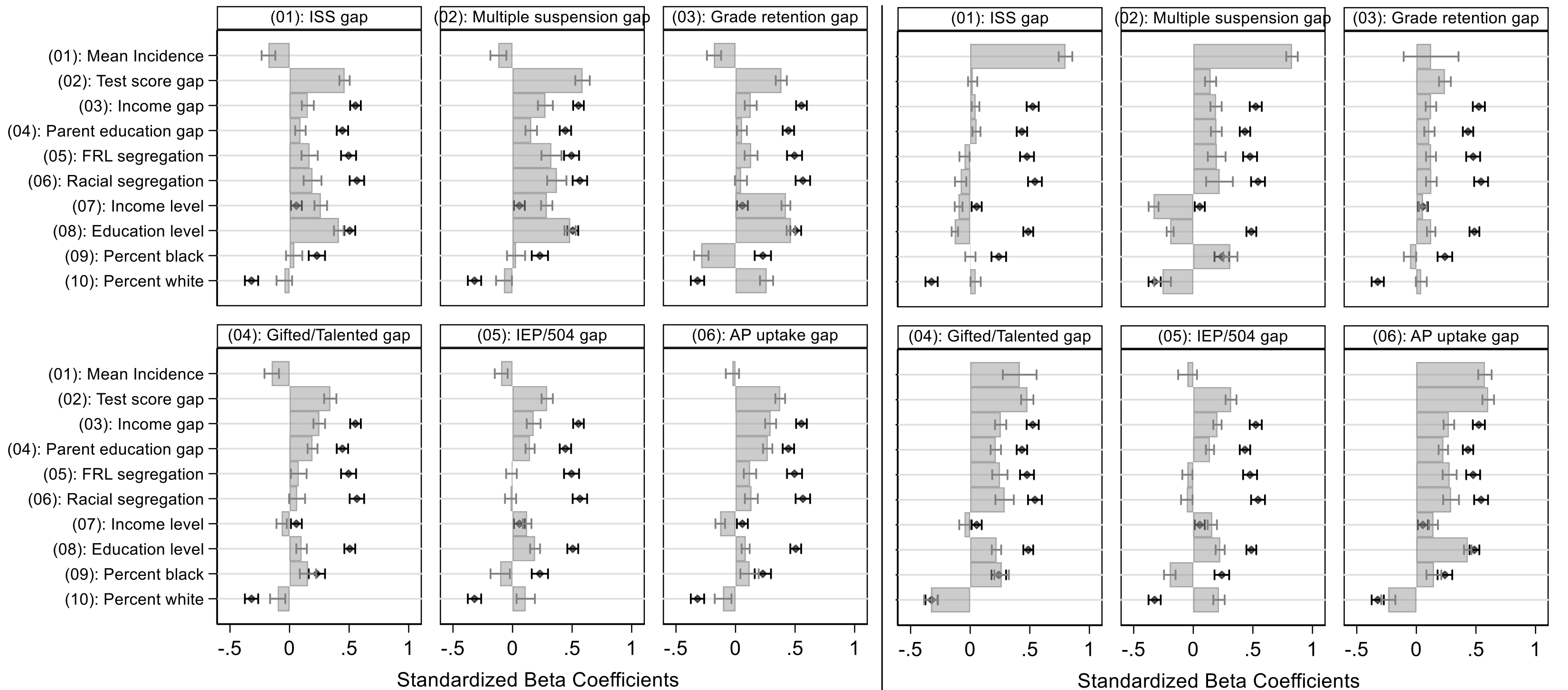
Note: Districts shown here have five unique categorical inequalities in the top quintile of the 2015/16 analytic samples. Districts in blue have five unique categorical inequalities in the top quintile and at least one categorical inequality in the top 10. Asterisks indicate that the district is repeated in both the relative and absolute risk metrics. The values shown in columns are standardized within the population sample to be mean zero and standard deviation one and are interpreted in standard deviation units. A value of 1 indicates a gap that is one standard deviation above the sample mean.

Table 9: Multivariate Regression Models and Adjusted R-Squared

	Test Score	ISS	Single Sus	Multiple Sus	Grade Retention	Gifted / Talented	IEP / 504	AP Up-take
Panel A: Relative Risk Ratios (Ln(%B/%W))								
M1: Racial Context Inequality	0.71	0.32	0.37	0.43	0.32	0.30	0.20	0.24
M2: M1+Racial Test Score Gap		0.38	0.44	0.48	0.36	0.37	0.24	0.29
M3: M1+Mean Incidence	0.72	0.32	0.38	0.43	0.33	0.33	0.20	0.24
M4: M3+Racial Test Score Gap		0.38	0.44	0.48	0.36	0.40	0.25	0.30
Panel B: Absolute Risk Differences (%B-%W)								
M1: Racial Context Inequality	0.69	0.26	0.17	0.27	0.10	0.35	0.26	0.37
M2: M1+Racial Test Score Gap		0.28	0.22	0.28	0.12	0.41	0.31	0.45
M3: M1+Mean Incidence	0.69	0.72	0.52	0.80	0.12	0.42	0.26	0.53
M4: M3+Racial Test Score Gap		0.73	0.57	0.81	0.13	0.47	0.31	0.60

Note: This table provides adjusted R-squared statistics from models of the form $Y = \boldsymbol{X\beta} + \delta_i + \lambda_y + \varepsilon$, where $\delta_i$and $\lambda_y$are state and year fixed effects, and $\boldsymbol{X}$ is a vector of predictor variables from Table 1. Model 1 (M1) includes variables related to racial socioeconomic inequality, racial composition and segregation (from Table 1). Model 2 (M2) includes M1 covariates plus racial test score inequality. Model 3 (M3) includes M1 covariates plus mean the mean incidence rate. Model 4 (M4) includes M1 covariates plus racial test score inequality and the mean incidence rate. All outcomes are taken from the 2011/12, 2013/14, and 2015/16 Office of Civil Rights Data Collection. The analytic samples in *Panels A* and *B* include 4,371 and 5,523 observations in the three years and 71% and 76% of Black students in the K-12 public school population, respectively.

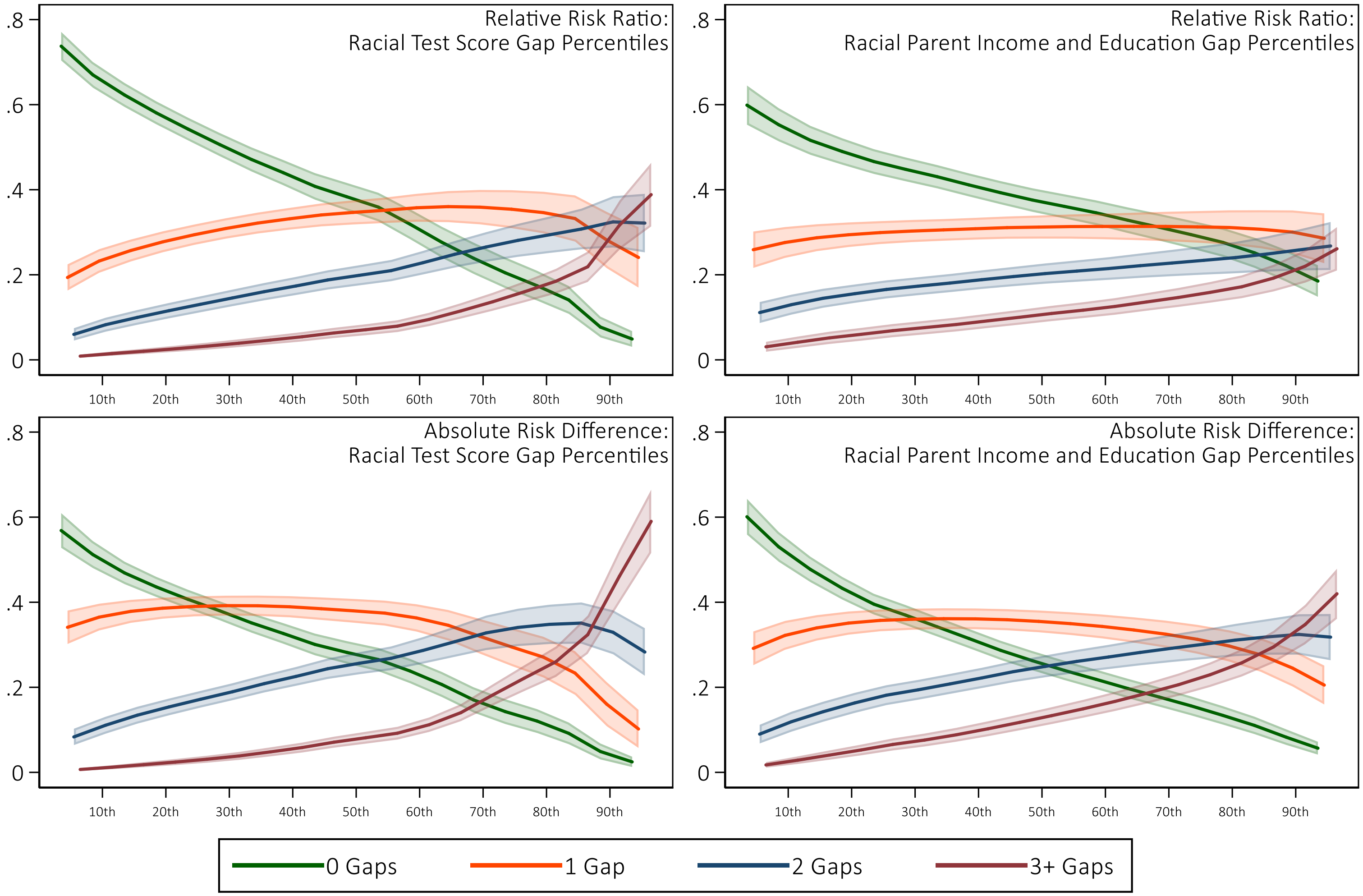
Figure 1: Standardized Bivariate Regression Coefficients, Black-White Gaps and Predictors



(A) Relative Risk Ratios (ln(%B/%W)) *(B) Absolute Risk Differences (%B-%W)*

Note: This figure presents standardized bivariate regression coefficients from models of the form $Y = \boldsymbol{X\beta} + \delta_i + \lambda_y + \varepsilon$, where δ_i and λ_y are state and year fixed effects, and $\boldsymbol{X}$ is a vector of predictor variables from Table 1. *Panel (A)* includes categorical inequalities in the metric of relative risk ratios, and *Panel (B)* is in the metric of absolute risk differences. All dependent and predictor variables are standardized so that coefficients are interpreted as a one standard deviation change in *X* is associated with a $\hat{\beta}$ (the estimated coefficient) standard deviation change in *Y*. Dependent variables are listed as title headers (1 through 6); independent variables are listed as rows (1 through 9). Range caps indicate 95% confidence interval. Gray bars correspond to $\hat{\beta}$ for the titled dependent variables. As reference, we include $\hat{\beta}$ using test score gaps as outcomes, shown as Black markers. The analytic samples in *Panels A* and *B* include 4,371 and 5,523 observations in the three years and 71% and 76% of Black students in the K-12 public school population, respectively.

Figure 2: Probabilities of Having *N* Gaps in the Top Quintile as a Function of Test Score and Parental SES Inequality



Note: This figure presents the average probability (marginal effects) of a district having 0, 1, 2 or 3 or more inequalities in the top quintile of the analytic samples (i.e., in the metrics of relative and absolute risk, 4,371 and 5,523 observations in the three years and 71% and 76% of Black students in the K-12 public school population, respectively) as a function of either (a) racial/ethnic test score inequality or (b) racial/ethnic income and parent education inequality. Estimates are generated from two separate multinomial logistic equations. Probabilities are calculated for the 5th to 95th percentiles of test score and socioeconomic inequality, respectively, in increments of five.

Data Appendix

The Stanford Education Data Archive (SEDA) provides Black-White test score gaps for academic years 2008-09 to 2012-13 in grades 3 through 8 for the subjects English/Language Arts (ELA) and math. The data provide achievement gaps for ELA and math for 2,803 public school districts in the United States, which includes 90% of the Black public-school population.[1] Achievement gaps are estimated based on "coarsened" proficiency data (i.e., percents of Black and White students with test scores in discrete performance categories, such as basic, proficient and advanced) from state accountability assessments and can be interpreted as effect sizes, or standardized mean differences (Reardon, et al., 2019; Reardon et al., 2017).

The Office of Civil Rights (OCR) Civil Rights Data Collection (CRDC) is a biennial survey required by the U.S. Department of Education's Office for Civil Rights. Data for the universe of schools in the United States are available for years 2011-12, 2013-14 and 2015-16. The CRDC collects data on a wide variety of outcomes, including demographics, math and science course-taking, advanced placement (AP) course-taking and passage rates, and school disciplinary outcomes. The data collected by the CRDC are intended for purposes of accountability related to issues of civil rights.[2]

From the CRDC data, we calculate the proportion of Black and White students who have been suspended, have been retained in any grade K-12, have taken one or more AP class, are enrolled in Gifted and Talented, or are classified as either having an IEP or 504 (special education) plan. The total counts of Black and White students for each of these variables are taken from the

[1] 2,853 districts have non-missing achievement gap data for ELA; 2,833 have non-missing data for math. 2,803 districts have non-missing data for both ELA and math.
[2] We requested data from the CRDC using the flat file request form, available here. http://ocrdata.ed.gov/RequestFlat-File

CRDC, and the total count of Black and White students is taken from the Common Core of Data (CCD). For variables corresponding to all grades K-12, the denominator includes K-12 enrollment. For AP course-taking, we use only grade 9-12 enrollment. We then have variables for the proportion White/Black for each of these variables for districts and years (2012, 2014 and 2016). Other outcomes are available from the CRDC but with much more missing data.

Gaps are constructed as either *relative risk ratios* or *absolute risk differences*. The former is calculated as the natural logarithm of the ratio of the percentage of Black or White students at the district level in which any of the preceding outcomes apply. For example, the Black-White gap in district *d* and year *t* for ISS is the log ratio of the percentage of Black students reported as receiving an in-school-suspension relative to the percentage of White students. We take the natural logarithm of the ratio so that ratio values lower than one are linearly proportional to values greater than one. The latter is calculated as the difference in the percentage of Black or White students at the district level in which any of the preceding outcomes apply.

Finally, all gaps are constructed so that higher values correspond to inequalities that disfavor Black students; thus, for example, with respect to achievement gaps, White achievement is the numerator, while with respect to discipline gaps, Black discipline is the numerator. In each case, gaps that are larger indicate Black students have worse relative outcomes.

District socioeconomic data are taken from the American Community Survey (ACS) Education Demographic and Geographic Estimates (EDGE) database. Data are based on multiple waves of the parent tabulation survey for years 2007-11, 2009-13, and 2011-15. The ACS uses a rolling survey design that encompasses multiple years. We merge the 2007-11 to the academic 2011/12 school year (i.e., merging the last survey year of the ACS to the Fall of the academic calendar), the 2009-13 to the academic 2013/14 school year, and the 2011-15 to the 2015/16 school year. The variables used here can be categorized into racial/ethnic family income differences and racial/ethnic parent education (having a Bachelor's degree or more) differences. We generate a measure of between school racial segregation (Herfindahl-Hirschman index) using data from the CCD. The proportion of students that are Black/White is also generated from the CCD.

Charter schools and schools administered by the state have a unique local education agency ID number (LEAID) but operate inside a geographic boundary assigned a different LEAID. For example, finance data for traditional public schools in New York City are assigned an LEAID of 3620580, but charter districts operating inside New York City are assigned unique LEAID numbers depending on the charter agency. Because Census and economic data are assigned to geographic areas and not to charter agencies, charter districts are reassigned the LEAID number that corresponds to the geographic boundary. These geographic boundaries are based on the latitude and longitude available in the school-universe file from the CCD. All charter districts are thereby subsumed into the geographic district; thus, the 1,825 school districts in the sample include both charter and traditional public schools. For additional discussion, see the SEDA Technical Documentation available here: https://cepa.stanford.edu/sites/default/files/SEDA%20Technical%20Documentation%20Version1_1.pdf.